Package ‘DirectEffects’

May 12, 2021

Version 0.2.1
Title Estimating Controlled Direct Effects for Explaining Causal Findings
Imports stats, sandwich, Formula, glue
Depends R (>= 3.5.0)
Suggests knitr, dplyr, ggplot2, reshape2, scales, testthat, rmarkdown
License GPL (>= 2)
URL https://www.mattblackwell.org/software/direct-effects/
BugReports https://github.com/mattblackwell/DirectEffects/issues
VignetteBuilder knitr
LazyData true
RoxygenNote 7.1.1
NeedsCompilation no
Author Matthew Blackwell [aut, cre],
Avidit Acharya [aut],
Maya Sen [aut],
Shiro Kuriwaki [aut],
Jacob Brown [aut]
Maintainer Matthew Blackwell <mblackwell@gov.harvard.edu>
Repository CRAN
Date/Publication 2021-05-12 20:02:11 UTC

R topics documented:

  boots_g ................................................................. 2
cdesens ................................................................. 3
boots_g

boots_g(seqg, boots = 1000)

Arguments

seqg A fitted sequential_g estimate, computed by sequential_g.
boots The number of bootstrap replicates. Defaults to 1000.

Value

An object of type seqgboots which is a matrix with boots rows and columns for each coefficient in the seqg model. Use summary to provide summary statistics, such as mean and quantiles.

Examples

data(ploughs)
form <- women_politics ~ plow +
  agricultural_suitability + tropical_climate + large_animals + rugged |
  years_civil_conflict + years_interstate_conflict + oil_pc +
  european_descent + communist_dummy + polity2_2000 |
  centered_ln_inc + centered_ln_incsq
s1 <- sequential_g(form, ploughs)
out.boots <- boots_g(s1)
summary(out.boots)
cdesens

Estimate sensitivity of ACDE estimates under varying levels of unobserved confounding

Description

Estimate how the Average Controlled Direct Effect varies by various levels of unobserved confounding. For each value of unmeasured confounding, summarized as a correlation between residuals, cdesens computes the ACDE. Standard errors are computed by a simple bootstrap.

Usage

```r
cdesens(
  seqg,  # Output from sequential_g. The function only supports specifications with one mediator variable.
  var,   # A character indicating the name of the variable for which the estimated ACDE is being evaluated.
  rho = seq(-0.9, 0.9, by = 0.05),  # A numerical vector of correlations between errors to test for. The original model assumes rho = 0
  bootstrap = c("none", "standard"),  # character of c("none", "standard"), indicating whether to include bootstrap standard errors. Default is "none".
  boots_n = 1000,  # Number of bootstrap replicates, defaults to 100.
  verbose = FALSE,  # Whether to show progress and messages, defaults to FALSE
  ...  # Other parameters to pass on to lm.fit() when refitting the model
)
```

Arguments

- **seqg**: Output from sequential_g. The function only supports specifications with one mediator variable.
- **var**: A character indicating the name of the variable for which the estimated ACDE is being evaluated.
- **rho**: A numerical vector of correlations between errors to test for. The original model assumes rho = 0
- **bootstrap**: character of c("none", "standard"), indicating whether to include bootstrap standard errors. Default is "none".
- **boots_n**: Number of bootstrap replicates, defaults to 100.
- **verbose**: Whether to show progress and messages, defaults to FALSE
- **...**: Other parameters to pass on to lm.fit() when refitting the model

Examples

```r
data(civilwar)

# main formula: Y - A + X | Z | M
form_main <- onset ~ ethfrac + lmtnest + ncontig + Oil | warl +
gdpenl + lpop + polity2l + relfrac | instab

# estimate CDE
```
direct <- sequential_g(form_main, data = civilwar)

# sensitivity
out_sens <- cdesens(direct, var = "ethfrac")

# plot sensitivity
plot(out_sens)

civilwar

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data on civil wars and internal conflict from 1945-1999.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>data(civilwar)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>A data frame with 6610 observations and 69 variables.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>• ccode. COW country id number</td>
</tr>
<tr>
<td>• country. country name</td>
</tr>
<tr>
<td>• cname. abbreviated country name</td>
</tr>
<tr>
<td>• cmark. 1 for first in each country series</td>
</tr>
<tr>
<td>• year. start year of war/conflict</td>
</tr>
<tr>
<td>• wars. number wars in progress in country year</td>
</tr>
<tr>
<td>• war. 1 if war ongoing in country year</td>
</tr>
<tr>
<td>• warl. lagged war, w/ 0 for start of country series</td>
</tr>
<tr>
<td>• onset. 1 for civil war onset</td>
</tr>
<tr>
<td>• ethonset. 1 if onset = 1 &amp; ethwar = 0</td>
</tr>
<tr>
<td>• durest. estimated war duration</td>
</tr>
<tr>
<td>• aim. 1 = rebels aim at center, 3 = aim at exit or autonomy, 2 = mixed or ambig.</td>
</tr>
<tr>
<td>• casename. Id for case, usually name of rebel group(s)</td>
</tr>
<tr>
<td>• ended. war ends = 1, 0 = ongoing</td>
</tr>
<tr>
<td>• ethwar. 0 = not ethnic, 1 = ambig/mixed, 2 = ethnic</td>
</tr>
<tr>
<td>• waryrs. war years for each onset</td>
</tr>
<tr>
<td>• pop. population, in 1000s</td>
</tr>
</tbody>
</table>
- lpop. log of pop
- polity2. revised polity score
- gdpen. gdp/pop based on pwt5.6, wdi2001,cow energy data
- gdptype. source/type of gdp/pop estimate
- gdpenl. lagged gdpenl, except for first in country series
- lgdpenl1. log of lagged gdpen
- lpop1. log population, lagged except for first in country series
- region. country’s region, based on MAR project
- western. Dummy for Western Democracies & Japan
- eeurop. Dummy for Eastern Europe
- lamerica. Dummy for Latin America
- ssafrica. Dummy for Sub-Saharan Africa
- asia. Dummy for Asia (not including Japan)
- nafrme. Dummy for North Africa/Middle East
- colbrit. Former British colony
- colfra. former French colony
- mtnest. Estimated percent mountainous terrain
- lmtnest. log of mtnest
- elevdiff. high - low elevation, in meters
- Oil. more than 1/3 export revenues from fuels
- ncontig. noncontiguous state
- ethfrac. ethnic frac. based on Soviet Atlas, plus estimates for missing in 1964
- plural. share of largest ethnic group (Fearon 2002 APSA)
- second. share of 2nd largest ethnic group (Fearon 2002 APSA)
- numlang. number languages in Ethnologue > min(1
- plurrel. size of largest confession
- minrelpc. size of second largest confession
- muslim. percent muslim
- nwstate. 1 in 1st 2 years of state’s existence
- polity2l. lagged polity2, except 1st in country series
- instab. > 2 change in Polity measure in last 3 yrs
- anocl. lagged anocracy (-6 < polity2l < 6)
- deml. lagged democracy (polity2l > 5)
- empehfrac. ethfrac coded for colonial empires
- empwarl. warl coded for data with empires
• emponset. onset coded for data with empires
• empgdpenl. gdpenl coded for empires data
• emplpopl. lpopl coded for empires data
• emplmntnest. lmntnest coded for empires data
• empncontig. ncontig coded for empires
• empolity2l. polity2l adjusted for empires (see fn38 in paper)
• sdwarsl. number Sambanis/Doyle civ wars in progress
• sdonset. onset of Sambanis/Doyle war
• colwars. number Collier/Hoeffler wars in progress
• colonset. onset of Collier/Hoeffler war
• cowwars. number COW civ wars in progress
• cowonset. onset of COW civ war
• cowwarl. 1 if COW war ongoing in last period
• sdwarl. 1 if S/D war ongoing in last period
• colwarl. 1 if C/H war ongoing in last period

Source

doi: 10.1017/S0003055403000534

References


---

plot.cdesens

Plot output from cdesens

Description

Plot output from cdesens

Usage

```r
## S3 method for class 'cdesens'
plot(
  x,
  level = 0.95,
  xlim = NULL,
  ylim = NULL,
  xlab = NULL,
  ylab = "Estimated ACDE",
  bty = "n",
  
```
Arguments

- **x**: output from `cdesens`
- **level**: level of confidence interval to plot
- **xlim**: the x limits (x1, x2) of the plot for the sensitivity analysis parameter, rho. Default is to use the range of rho.
- **ylim**: the y limits of the plot for the estimated CDEs. Default is to show all of the confidence intervals.
- **xlab**: label for the x axis.
- **ylab**: label for the y axis.
- **bty**: a character string which determined the type of box which is drawn about plots. Defaults to not drawing a box. See `par` for more information.
- **col**: color for the line indicating the point estimates of the bias-adjusted ACDE.
- **lwd**: line width for the line indicating the point estimates of the bias-adjusted ACDE.
- **ci.col**: color for the polygon that shows the confidence intervals.
- **ref.lines**: a logical indicating whether horizontal and vertical lines at 0 should be plotted.
- **...**: Other parameters to pass on to `plot()`

---

**ploughs**

*Data on historical plough use and the socioeconomic status of women.*

Description

A dataset to replicate the analysis in Alesina, Giuliano, and Nunn (2013).

Usage

```r
data(ploughs)
```

Format

A data frame with 234 observations and 57 variables.
Details

- isocode. 3-letter code for the country.
- flfp2000. Female labor force participation in 2000
- female_ownership. Percent of firms with female ownership (in latest survey year)
- women_politics. Women in Politics in 2000, WDI
- agricultural_suitability. overall (millets, sorghum, wheat, barley, rye): share defined as suitable
- tropical_climate. Frac land: tropics and subtropics: using Ethnologue - pop weighted
- large_animals. presence of large animals
- political_hierarchies. Jurisdictional hierarchy beyond local community (v33): Using Ethnologue - pop weighted
- economic_complexity. Settlement patterns (v30)
- ln_income. ln (income)
- ln_income_squared. ln (income) ^2
- centered_ln_inc. de-meaned ln_inc
- centered_ln_incsq. de-meaned ln_inc squared
- country. country name
- communist_dummy. Communism indicator variable
- rugged. Ruggedness (Terrain Ruggedness Index, 100 m.)
- years_interstate_conflict. Years of interstate conflict, 1800-2007 - from COW
- polity2_2000. Polity 2 measure taken from the Polity IV dataset
- oil_pc. oil production/GDP
- ... other variables as annotated in the source.

Source

doi: 10.1093/qje/qjt005

References

sequential_g

Perform linear sequential g-estimation to estimate the controlled direct effect of a treatment net the effect of a mediator.

Description

Perform linear sequential g-estimation to estimate the controlled direct effect of a treatment net the effect of a mediator.

Usage

sequential_g(formula, data, subset, weights, na.action, offset, contrasts = NULL, verbose = TRUE, ...)

Arguments

- **formula**: formula specification of the first-stage, second-stage, and blip-down models. The right-hand side of the formula should have three components separated by the |, with the first component specifying the first-stage model with treatment and any baseline covariates, the second component specifying the intermediate covariates for the first-stage, and the third component specifying the blip-down model. See Details below for more information.

- **data**: A dataframe to apply formula on.

- **subset**: A vector of logicals indicating which rows of data to keep.

- **weights**: an optional vector of weights to be used in the fitting process. Should be NULL or a numeric vector. If non-NULL, weighted least squares is used with weights weights (that is, minimizing \( \text{sum}(w*e^2) \)); otherwise ordinary least squares is used. See also ‘Details’.

- **na.action**: a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The ‘factory-fresh’ default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.

- **offset**: this can be used to specify an \textit{a priori} known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector or matrix of extents matching those of the response. One or more offset terms can be included in the formula instead or as well, and if more than one are specified their sum is used. See model.offset.
contrasts an optional list. See the contrasts.arg of model.matrix.default.
verbose logical indicating whether to suppress progress bar. Default is FALSE.
... additional arguments to be passed to the low level regression fitting functions (see below).

Details

The sequential_g function implements the linear sequential g-estimator developed by Vansteelandt (2009) with the consistent variance estimator developed by Acharya, Blackwell, and Sen (2016).

The formula specifies the full first-stage model including treatment, baseline confounders, intermediate confounders, and the mediators. The user places | bars to separate out these different components of the model. For example, the formula should have the form y ~ tr + x1 + x2 | z1 + z2 | m1 + m2. where tr is the name of the treatment variable, x1 and x2 are baseline covariates, z1 and z2 are intermediate covariates, and m1 and m2 are the names of the mediator variables. This last set of variables specify the 'blip-down' or 'demediation' function that is used to remove the average effect of the mediator (possibly interacted) from the outcome to create the blipped-down outcome. This blipped-down outcome is the passed to a standard linear model with the covariates as specified for the direct effects model.

See the references below for more details.

Value

Returns an object of class A "seqg". Similar to the output of a call to lm. Contains the following components:

- coefficients: a vector of named coefficients for the direct effects model.
- residuals: the residuals, that is the blipped-down outcome minus the fitted values.
- rank: the numeric rank of the fitted linear direct effects model.
- fitted.values: the fitted mean values of the direct effects model.
- weights: (only for weighted fits) the specified weights.
- df.residual: the residual degrees of freedom for the direct effects model.
- aliased: logical vector indicating if any of the terms were dropped or aliased due to perfect collinearity.
- terms: the list of terms object used. One for the baseline covariates and treatment (X) and one for the variables in the blip-down model (M).
- formula: the formula object used, possibly modified to drop a constant in the blip-down model.
- call: the matched call.
- na.action: (where relevant) information returned by model.frame of the special handling of NAs.
- xlevels: the levels of the factor variables.
- contrasts: the contrasts used for the factor variables.
- first_mod: the output from the first-stage regression model.
model: full model frame, including all variables.
• \( \widetilde{Y} \): the blipped-down response vector.
• \( X \): the model matrix for the second stage.
• \( M \): the model matrix for demediation/blip-down function.

In addition, non-null fits will have components assign, effects, and qr from the output of \texttt{lm.fit} or \texttt{lm.wfit}, whichever is used.

References

Examples
data(ploughs)
form_main <- women_politics ~ plow +
agricultural_suitability + tropical_climate + large_animals +
political_hierarchies + economic_complexity +
rugged | years_civil_conflict +
years_interstate_conflict + oil_pc +
european_descent + communist_dummy + polity2_2000 +
serv_va_gdp2000 | centered_ln_inc + centered_ln_incsq
direct <- sequential_g(form_main, ploughs)
summary(direct)

summary.seqg

Computes standard errors and p-values of DirectEffects estimates

Description
Computes standard errors and p-values of DirectEffects estimates

Usage
## S3 method for class 'seqg'
summary(object, ...)

Arguments
object An object of class seqg, computed by \texttt{sequential_g}.
... additional arguments affecting the summary produced.
Summary of DirectEffect Bootstrap Estimates

Usage

```r
## S3 method for class 'seqgboots'
summary(object, level = 0.95, ...)
```

Arguments

- `object` An output of class `seqg` estimated by `boots_g`.
- `level` level of intervals to estimate. Defaults to 0.95
- `...` additional arguments affecting the summary produced.
Index

boots_g, 2, 12

cdesens, 3
civilwar, 4

model.matrix.default, 10
model.offset, 9

na.exclude, 9
na.fail, 9
na.omit, 9

offset, 9
options, 9

par, 7
plot.cdesens, 6
ploughs, 7

sequential_g, 2, 9, 11
summary.seqg, 11
summary.seqgboots, 12