

Package ‘DirectEffects’

January 4, 2019

Version 0.2

Title Estimating Controlled Direct Effects for Explaining Causal Findings

Imports stats, sandwich, Formula, glue

Depends R (>= 3.0.0)

Suggests knitr, dplyr, ggplot2, reshape2, scales, testthat

Description A set of functions to estimate the controlled direct effect of treatment fixing a potential mediator to a specific value. Implements the sequential g-estimation estimator described in Vansteelandt (2009) <doi:10.1097/EDE.0b013e3181b6f4c9> and Acharya, Blackwell, and Sen (2016) <doi:10.1017/S0003055416000216>.

License GPL (>= 2)

URL <http://www.mattblackwell.org/software/direct-effects/>

BugReports <https://github.com/mattblackwell/DirectEffects/issues>

VignetteBuilder knitr

LazyData true

RoxygenNote 6.1.1

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 2019-01-04 10:00:08 UTC

R topics documented:

boots_g	2
cdesens	3

civilwar	4
plot.cdesens	6
ploughs	7
sequential_g	8
summary.seqg	11
summary.seqgboots	11

Index	12
--------------	-----------

boots_g	<i>Coefficient Estimates across Bootstrapped Samples</i>
---------	--

Description

Performs a simple bootstrap of a fitted DirectEffects model by re-estimating the model with bootstrap samples.

Usage

```
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	<i>Estimate sensitivity of ACDE estimates under varying levels of unobserved confounding</i>
---------	--

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

```
cdesens(seqg, var, rho = seq(-0.9, 0.9, by = 0.05),
        bootstrap = c("none", "standard"), boots_n = 1000, verbose = FALSE,
        ...)
```

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 <code>lm.fit()</code> when refitting the model

Examples

```
data(civilwar)

# main formula: Y ~ A + X | Z | M
form_main <- onset ~ ethfrac + lmtnest + ncontig + Oil | war1 +
  gdpnl + 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

Data on civil wars and internal conflict from 1945-1999.

Description

A dataset to replicate the analysis in Fearon and Laitin (2003).

Usage

```
data(civilwar)
```

Format

A data frame with 6610 observations and 69 variables.

Details

- ccode. COW country id number
- country. country name
- cname. abbreviated country name
- cmark. 1 for first in each country series
- year. start year of war/conflict
- wars. number wars in progress in country year
- war. 1 if war ongoing in country year
- warl. lagged war, w/ 0 for start of country series
- onset. 1 for civil war onset
- ethonset. 1 if onset = 1 & ethwar != 0
- durest. estimated war duration
- aim. 1 = rebels aim at center, 3 = aim at exit or autonomy, 2 = mixed or ambig.
- casename. Id for case, usually name of rebel group(s)
- ended. war ends = 1, 0 = ongoing
- ethwar. 0 = not ethnic, 1 = ambig/mixed, 2 = ethnic
- waryrs. war years for each onset
- pop. population, in 1000s
- 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

- lpopl1. 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
- ef. ethnic fractionalization based on Fearon 2002 APSA paper
- 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
- relfrac. religious fractionalization
- 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
- anoctl. lagged anocracy ($-6 < \text{polity2l} < 6$)
- deml. lagged democracy ($\text{polity2l} > 5$)
- empethfrac. 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
- emplmtnest. lmtnest coded for empires data
- empncontig. ncontig coded for empires
- empolity2l. polity2l adjusted for empires (see fn38 in paper)

- sdwars. 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

<https://doi.org/10.1017/S0003055403000534>

References

Fearon, James D., and David A. Laitin (2003). Ethnicity, Insurgency, and Civil War. *American Political Science Review*, 97(1), 75-90. doi:10.1017/S0003055403000534

plot.cdesens

Plot output from cdesens

Description

Plot output from cdesens

Usage

```
## S3 method for class 'cdesens'
plot(x, level = 0.95, xlim = NULL, ylim = NULL,
     xlab = NULL, ylab = "Estimated ACDE", bty = "n", col = "black",
     lwd = 2, ci.col = "grey70", ref.lines = TRUE, ...)
```

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 the 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 <code>plot()</code>

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

```
data(ploughs)
```

Format

A data frame with 234 observations and 57 variables.

Details

- `isocode`. 3-letter code for the country.
- `fifp2000`. 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
- `plow`. Animal plow cultivation variable (v39): Using Ethnologue - pop weighted
- `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`. $\ln(\text{income})$
- `ln_income_squared`. $\ln(\text{income})^2$
- `centered_ln_inc`. de-meanded \ln_{inc}
- `centered_ln_incsq`. de-meanded \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
- `serv_va_gdp2000`. Value Added in Service/GDP in 2000
- `polity2_2000`. Polity 2 measure taken from the Polity IV dataset
- `oil_pc`. oil production/GDP
- ... other variables as annotated in the source.

Source

<http://qje.oxfordjournals.org/content/128/2/469>

References

Alesina, A., Giuliano, P., & Nunn, N. (2013). On the Origins of Gender Roles: Women and the Plough. *The Quarterly Journal of Economics*, 128(2), 469-530.

sequential_g	<i>Perform linear sequential g-estimation to estimate the controlled direct effect of a treatment net the effect of a mediator.</i>
--------------	---

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 $\sum(w \cdot e^2)$); 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 <code>options</code> , and is <code>na.fail</code> if that is unset. The ‘factory-fresh’ default is <code>na.omit</code> . Another possible value is <code>NULL</code> , no action. Value <code>na.exclude</code> can be useful.
offset	this can be used to specify an <i>a priori</i> known component to be included in the linear predictor during fitting. This should be <code>NULL</code> or a numeric vector or matrix of extents matching those of the response. One or more <code>offset</code> terms can be included in the formula instead or as well, and if more than one are specified their sum is used. See <code>model.offset</code> .
contrasts	an optional list. See the <code>contrasts.arg</code> of <code>model.matrix.default</code> .
verbose	logical indicating whether to suppress progress bar. Default is <code>FALSE</code> .
...	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 `"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.
- `Ytilde`: 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 `lm.fit` or `lm.wfit`, whichever is used.

References

Vansteelandt, S. (2009). Estimating Direct Effects in Cohort and Case-Control Studies. *Epidemiology*, 20(6), 851-860.

Acharya, Avidit, Blackwell, Matthew, and Sen, Maya. (2016) "Explaining Causal Effects Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110:3 pp. 512-529

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	<i>Computes standard errors and p-values of DirectEffects estimates</i>
--------------	---

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 sequential_g .
...	additional arguments affecting the summary produced.

summary.seqgboots	<i>Summary of DirectEffect Bootstrap Estimates</i>
-------------------	--

Description

Summary of DirectEffect Bootstrap Estimates

Usage

```
## 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](#), [11](#)

`cdesens`, [3](#)

`civilwar`, [4](#)

`model.matrix.default`, [9](#)

`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](#), [8](#), [11](#)

`summary.seqg`, [11](#)

`summary.seqgboots`, [11](#)