Package 'CausalSpline'

March 25, 2026

Type Package

Title Nonlinear Causal Dose-Response Estimation via Splines

Version 0.1.0

Description Estimates nonlinear causal dose-response functions for continuous treatments using spline-based methods under standard causal assumptions (unconfoundedness / ignorability). Implements three identification strategies: Inverse Probability Weighting (IPW) via the generalised propensity score (GPS), G-computation (outcome regression), and a doubly-robust combination. Natural cubic splines and B-splines are supported for both the exposure-response curve f(T) and the propensity nuisance model. Pointwise confidence bands are obtained via the sandwich estimator or nonparametric bootstrap. Also provides fragility diagnostics including pointwise curvature-based fragility, uncertainty-normalised fragility, and regional integration over user-defined treatment intervals. Builds on the framework of Hirano and Imbens (2004) <doi:10.1111/j.1468-0262.2004.00481.x> for continuous treatments and extends it to fully nonparametric spline estimation.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.3

Depends R (>= 4.1.0)

Imports splines, stats, utils, ggplot2 (>= 3.4.0), sandwich, boot

Suggests testthat (>= 3.0.0), knitr, rmarkdown, patchwork, cobalt, dplyr

VignetteBuilder knitr

Config/testthat/edition 3

URL https://github.com/causalfragility-lab/CausalSpline

BugReports https://github.com/causalfragility-lab/CausalSpline/issues

NeedsCompilation no

Author Subir Hait [aut, cre] (ORCID: <https://orcid.org/0009-0004-9871-9677>)

Maintainer Subir Hait <haitsubi@msu.edu>

Repository CRAN

Date/Publication 2026-03-25 21:10:07 UTC

Contents

`CausalSpline-package` *CausalSpline: Nonlinear Causal Dose-Response Estimation via Splines*

Description

CausalSpline estimates causal dose-response functions $E[Y(t)]$ for continuous treatments under unconfoundedness, without imposing linearity on the treatment effect. The package fills a gap in the causal inference ecosystem: most tools assume a scalar $\beta_1 T$ treatment effect, whereas real policy or health applications frequently exhibit thresholds, diminishing returns, and non-monotone relationships.

Core functions

`causal_spline` Main fitting function. Supports IPW, G-computation, and doubly-robust estimation.

`gps_model` Fit the generalised propensity score model for continuous treatments.

`trim_weights` Winsorise extreme IPW weights.

`check_overlap` Diagnose positivity (ESS, weight plots).

`dose_response_curve` Extract the curve data frame.

`simulate_dose_response` Simulate nonlinear dose-response data for benchmarking.

`gradient_curve` Numerical first and second derivatives of the fitted dose-response curve.

`fragility_curve` Pointwise fragility with adaptive eps, interpretation zones, and uncertainty normalisation.

`region_fragility` Integrated fragility over a treatment interval.

Causal identification

Identification relies on two standard assumptions:

1. **Unconfoundedness** (strong ignorability): $Y(t) \perp T \mid X$ for all t.
2. **Positivity** (overlap): $f(t \mid X = x) > 0$ for all t in the support of T and all x in the support of X.

Methods

IPW The outcome is regressed on a spline of T using GPS-based inverse probability weights. Consistent if the GPS model is correctly specified.

G-computation The outcome model includes spline(T) + X. The curve is obtained by marginalising over the observed X distribution. Consistent if the outcome model is correctly specified.

Doubly Robust Combines IPW and g-computation. Consistent if at least one of the two models is correctly specified.

Author(s)

Maintainer: Subir Hait <haitsubi@msu.edu> (ORCID)

References

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. doi:10.1002/0470090456.ch7

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3), 706-710. doi:10.1093/biomet/87.3.706

See Also

Useful links:

- https://github.com/causalfragility-lab/CausalSpline
- Report bugs at https://github.com/causalfragility-lab/CausalSpline/issues

causal_spline *Nonlinear causal dose-response estimation via splines*

Description

Estimates the causal dose-response function $E[Y(t)]$ for a continuous treatment T under unconfoundedness, using either Inverse Probability Weighting (IPW) with the generalised propensity score (GPS) or G-computation (outcome regression). The exposure-response curve is modelled as a natural cubic spline or B-spline, allowing thresholds, diminishing returns, and other nonlinear patterns to be recovered without parametric assumptions on the functional form.

Usage

```
causal_spline(
  formula,
  data,
  method = c("ipw", "gcomp", "dr"),
  spline_type = c("ns", "bs"),
  df_exposure = 4L,
  knots_exposure = NULL,
  df_gps = 4L,
  gps_family = c("gaussian", "gamma", "beta"),
  trim_quantiles = c(0.01, 0.99),
  eval_grid = 100L,
  eval_points = NULL,
  se_method = c("sandwich", "bootstrap"),
  boot_reps = 500L,
  conf_level = 0.95,
  verbose = FALSE
)
```

Arguments

formula	A two-sided formula of the form `outcome ~ treatment | covariate1 + covariate2 + ...`. The vertical bar `|` separates the treatment variable from the confounders. Example: `Y ~ T | X1 + X2 + X3`.
data	A data frame containing all variables in `formula`.
method	Character string. Estimation method: `"ipw"` (inverse probability weighting via GPS), `"gcomp"` (g-computation / outcome regression), or `"dr"` (doubly-robust, combines both). Default `"ipw"`.
spline_type	Character string. Type of spline basis for $f(T)$: `"ns"` (natural cubic spline, default) or `"bs"` (B-spline).
df_exposure	Integer. Degrees of freedom for the treatment spline $f(T)$. Default `4`.
knots_exposure	Numeric vector of interior knot positions for the treatment spline. If `NULL` (default), knots are placed at quantiles of the treatment distribution.

df_gps	Integer. Degrees of freedom for the GPS (propensity) spline used in the "ipw" method. Default 4.
gps_family	Character string. Family for the GPS model: "gaussian" (default) for a linear GPS, "gamma", or "beta" for positive / bounded treatments.
trim_quantiles	Numeric vector of length 2 giving lower and upper quantiles for weight trimming. Default c(0.01, 0.99). Set to NULL to skip trimming.
eval_grid	Integer. Number of equally-spaced treatment values at which to evaluate $E[Y(t)]$. Default 100.
eval_points	Numeric vector of specific treatment values at which to evaluate the curve. Overrides eval_grid if supplied.
se_method	Character string. Method for standard errors: "sandwich" (default, fast) or "bootstrap".
boot_reps	Integer. Number of bootstrap replications when se_method = "bootstrap". Default 500.
conf_level	Numeric. Confidence level for intervals. Default 0.95.
verbose	Logical. Print progress messages. Default FALSE.

Value

An object of class "causal_spline" with elements:

curve A data frame with columns t (treatment grid), estimate (E[Y(t)]), se, lower, upper.

ate Estimated average treatment effect over the observed treatment range (scalar).

weights Numeric vector of final (trimmed, normalised) IPW weights (only for method = "ipw" or "dr").

gps_fit The fitted GPS model object.

outcome_fit The fitted outcome model object.

call The matched call.

method The estimation method used.

spline_type Spline type used.

df_exposure Degrees of freedom for the exposure spline.

data_summary Summary statistics of the treatment variable.

References

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, 226164, 73-84. doi:10.1002/0470090456.ch7

Flores, C. A., Flores-Lagunes, A., Gonzalez, A., & Neumann, T. C. (2012). Estimating the effects of length of exposure to instruction in a training program: the case of job corps. *Review of Economics and Statistics*, 94(1), 153-171. doi:10.1162/REST_a_00177

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3), 706-710. doi:10.1093/biomet/87.3.706

Examples

```
# Simulate nonlinear dose-response data
set.seed(42)
dat <- simulate_dose_response(n = 200, dgp = "threshold")

# Fit with IPW
fit <- causal_spline(Y ~ T | X1 + X2 + X3, data = dat, method = "ipw",
                     df_exposure = 5)
summary(fit)
plot(fit)

# Fit with G-computation
fit_gc <- causal_spline(Y ~ T | X1 + X2 + X3, data = dat,
                        method = "gcomp", df_exposure = 5)
plot(fit_gc, truth = dat)
```

---

check_overlap *Diagnose positivity / overlap for a continuous treatment*

Description

Plots the distribution of the treatment variable conditional on covariate strata and returns effective sample size (ESS) and weight diagnostics to assess the positivity (overlap) assumption.

Usage

```
check_overlap(treatment, weights, plot = TRUE)
```

Arguments

`treatment`	Numeric vector of treatment values.
`weights`	Numeric vector of IPW weights (length must equal `length(treatment)`).
`plot`	Logical. If `TRUE`, returns a `ggplot2` diagnostic plot. Default `TRUE`.

Value

A list with:

`ess` Effective sample size: $(\sum w_i)^2 / \sum w_i^2$.

`weight_summary` Five-number summary of the weights.

`plot` ggplot2 object (if `plot = TRUE`).

Examples

```
dat <- simulate_dose_response(200)
X   <- as.matrix(dat[, c("X1", "X2", "X3")])
gps <- gps_model(dat$T, X)
w   <- trim_weights(abs(1 / stats::residuals(gps$model)), c(0.01, 0.99))
check_overlap(dat$T, w)
```

---

dose_response_curve *Extract the dose-response curve data frame*

Description

Convenience function to pull the estimated $E[Y(t)]$ curve from a fitted `causal_spline` object.

Usage

```
dose_response_curve(fit)
```

Arguments

`fit`	A `causal_spline` object.

Value

A data frame with columns `t`, `estimate`, `se`, `lower`, `upper`.

Examples

```
dat <- simulate_dose_response(200)
fit <- causal_spline(Y ~ T | X1 + X2 + X3, data = dat)
curve_df <- dose_response_curve(fit)
head(curve_df)
```

---

fragility_curve *Geometric fragility curve for a CausalSpline fit*

Description

Computes a pointwise shape-based fragility measure from the first and second derivatives of the fitted dose-response curve, with five enhancements:

1. **Adaptive eps**: stabilisation constant is $0.05 \times \text{median}(|f'(t)|)$ so interpretation is consistent across datasets.
2. **Interpretation zones**: fragility is classified into `"low"`, `"moderate"`, and `"high"` based on the 50th and 75th percentiles of the pointwise fragility distribution.
3. **Uncertainty-normalised fragility**: an additional column $F^*(t) = F(t)/\text{SE}(\hat{\mu}(t))$ combines shape instability with statistical uncertainty.
4. **Support density**: the kernel density of the treatment variable (if `t_obs` is supplied) is attached, flagging regions with sparse data.
5. **High-fragility flag**: logical column `high_fragility` marks points above the 75th percentile.

Usage

```
fragility_curve(
  object,
  grid = NULL,
  h = NULL,
  eps = NULL,
  type = c("curvature_ratio", "inverse_slope"),
  t_obs = NULL,
  ...
)
```

Arguments

`object`	A fitted `causal_spline` object.
`grid`	Optional numeric evaluation grid. If `NULL`, the fitted grid in `object$curve$t` is used.
`h`	Step size for finite differences. Default `NULL` (auto).
`eps`	Numeric or `NULL`. If `NULL` (default), set adaptively to $0.05 \times \text{median}(\|f'(t)\|)$. If numeric, used directly.
`type`	Character. Fragility definition: `"curvature_ratio"` (default) or `"inverse_slope"`.
`t_obs`	Optional numeric vector of observed treatment values (used to compute support density). If `NULL`, density column is omitted.
`...`	Ignored.

Value

A data frame of class `"fragility_curve"` with columns:

`t` Treatment grid.

`estimate` Fitted $E[Y(t)]$.

`se` Standard error of fitted curve.

`derivative` First derivative.

`second_derivative` Second derivative.

`fragility` Pointwise fragility $F(t)$.

`fragility_norm` Uncertainty-normalised fragility $F^*(t) = F(t)/\text{SE}(\hat{\mu}(t))$.

`fragility_zone` Factor: `"low"`, `"moderate"`, `"high"`.

`high_fragility` Logical: `TRUE` if above 75th percentile.

`support_density` Kernel density of T at each grid point (only if `t_obs` supplied).

`fragility_type` Character. Type used.

Examples

```
dat <- simulate_dose_response(200, dgp = "threshold")
fit <- causal_spline(Y ~ T | X1 + X2 + X3, data = dat)
fc  <- fragility_curve(fit, t_obs = dat$T)
plot(fc)
```

`gps_model` *Fit the Generalised Propensity Score model*

Description

Fits a model for the conditional distribution of a continuous treatment $T|X$ (the generalised propensity score). Covariates are entered linearly by default; spline transformations of the treatment are not needed here since this models the treatment, not the outcome.

Usage

```
gps_model(
  treatment,
  covariates,
  spline_type = c("ns", "bs"),
  df = 4L,
  family = c("gaussian", "gamma", "beta"),
  verbose = FALSE
)
```

Arguments

`treatment`	Numeric vector of treatment values.
`covariates`	Numeric matrix of confounders.
`spline_type`	Character. `"ns"` or `"bs"` for spline transformations of continuous covariates. Default `"ns"`.

df	Integer. Degrees of freedom for covariate splines. Default 4. Set to NULL to use linear covariates only.
family	Character. Distribution for the GPS model. "gaussian" (default), "gamma", or "beta".
verbose	Logical. Default FALSE.

Value

A list with:

model Fitted model object (lm or glm).

family Family used.

r_squared R-squared of the GPS model (diagnostic).

Examples

```
dat <- simulate_dose_response(n = 200, dgp = "linear")
X   <- as.matrix(dat[, c("X1", "X2", "X3")])
gps <- gps_model(dat$T, X)
summary(gps$model)
```

---

gradient_curve *Numerical derivatives of a CausalSpline dose-response curve*

Description

Computes first and second numerical derivatives of the fitted dose-response curve $E[Y(t)]$ using central finite differences applied to predict.causal_spline. Useful for identifying regions of rapid change (high first derivative) or inflection / diminishing returns (second derivative changes sign).

Usage

```
gradient_curve(object, grid = NULL, h = NULL, eps = 1e-06, ...)
```

Arguments

object	A fitted causal_spline object.
grid	Optional numeric vector of treatment values at which to evaluate the derivatives. If NULL, the fitted evaluation grid stored in object$curve$t is used.
h	Numeric. Step size for finite differences. If NULL, chosen automatically as $(t_{max} - t_{min})/500$.
eps	Small positive constant used by fragility_curve to stabilise division. Default 1e-6.
...	Ignored.

Value

A data frame with columns:

`t` Treatment grid values.

`estimate` Fitted $E[Y(t)]$.

`se` Standard error of $E[Y(t)]$ from the fitted curve.

`derivative` First derivative $f'(t)$.

`second_derivative` Second derivative $f''(t)$.

Examples

```
dat <- simulate_dose_response(200, dgp = "threshold")
fit <- causal_spline(Y ~ T | X1 + X2 + X3, data = dat)
gd  <- gradient_curve(fit)
head(gd)
```

---

plot.causal_spline *Plot the estimated dose-response curve*

Description

Plots the estimated causal dose-response curve $E[Y(t)]$ against t with pointwise confidence bands and an optional rug for the observed treatment distribution.

Usage

```
## S3 method for class 'causal_spline'
plot(
  x,
  rug = TRUE,
  truth = NULL,
  xlab = "Treatment (T)",
  ylab = "E[Y(t)]",
  title = NULL,
  ...
)
```

Arguments

`x`	A `causal_spline` object.
`rug`	Logical. Add a treatment distribution rug. Default `TRUE`.
`truth`	Optional data frame with columns `t` and `true_effect` for overlaying the true dose-response (useful in simulations).
`xlab`	Character. x-axis label. Default `"Treatment (T)"`.

ylab	Character. y-axis label. Default "E[Y(t)]".
title	Character. Plot title. Default NULL.
...	Ignored.

Value

A ggplot2 object.

Examples

```
dat <- simulate_dose_response(200, dgp = "threshold")
fit <- causal_spline(Y ~ T | X1 + X2 + X3, data = dat)
plot(fit)
```

---

plot.fragility_curve *Plot method for fragility_curve objects*

Description

Produces a dual-panel plot: the dose-response curve (top) and the fragility curve (bottom), with high-fragility regions shaded. If support_density is present in x (i.e. t_obs was supplied to fragility_curve), a scaled density ribbon is overlaid on the fragility panel to flag low-support regions.

Usage

```
## S3 method for class 'fragility_curve'
plot(x, ...)
```

Arguments

x	A fragility_curve data frame (output of fragility_curve).
...	Ignored.

Value

A combined patchwork plot if the **patchwork** package is installed, otherwise the two panels are printed separately and a list of two ggplot2 objects is returned invisibly.

Examples

```
dat <- simulate_dose_response(200, dgp = "threshold")
fit <- causal_spline(Y ~ T | X1 + X2 + X3, data = dat)
fc  <- fragility_curve(fit, t_obs = dat$T)
plot(fc)
```

predict.causal_spline *Predict E[Y(t)] at new treatment values*

Description

Predict E[Y(t)] at new treatment values

Usage

```
## S3 method for class 'causal_spline'
predict(object, newt, se_fit = FALSE, warn_extrap = TRUE, ...)
```

Arguments

object	A causal_spline object.
newt	Numeric vector of treatment values at which to predict.
se_fit	Logical. Return standard errors? Default FALSE.
warn_extrap	Logical. Warn if any values in newt fall outside the observed treatment range? Default TRUE.
...	Ignored.

Value

A data frame with columns t, estimate, extrapolated, and optionally se, lower, upper. The extrapolated column is TRUE for any newt value outside the observed treatment support.

Examples

```
dat <- simulate_dose_response(200)
fit <- causal_spline(Y ~ T | X1 + X2 + X3, data = dat)
predict(fit, newt = c(1, 2, 3, 4, 5))
```

print.causal_spline *Print method for causal_spline objects*

Description

Print method for causal_spline objects

Usage

```
## S3 method for class 'causal_spline'
print(x, ...)
```

Arguments

x	A causal_spline object.
...	Ignored.

Value

Invisibly returns the input causal_spline object x, unchanged. The function is called for its side effect of printing a compact summary to the console, showing the estimation method, spline type, degrees of freedom, sample size, treatment range, and number of evaluation points on the dose-response curve.

---

region_fragility *Regional fragility summary for a CausalSpline fit*

Description

Integrates the pointwise fragility curve over a treatment interval $[a, b]$ using the trapezoidal rule. Useful for comparing sensitivity across dose ranges (e.g., low vs. high dose) or summarising instability at a policy-relevant threshold.

Usage

```
region_fragility(
  object,
  a,
  b,
  grid = NULL,
  h = NULL,
  eps = NULL,
  type = c("curvature_ratio", "inverse_slope"),
  normalize = TRUE,
  t_obs = NULL,
  ...
)
```

Arguments

object	A fitted causal_spline object.
a	Numeric scalar. Lower bound of the integration interval.
b	Numeric scalar. Upper bound of the integration interval.
grid	Optional numeric evaluation grid within $[a, b]$. If NULL, a dense grid of 300 points is created automatically.
h	Step size for finite differences. Default NULL (auto).
eps	Adaptive eps passed to fragility_curve. Default NULL (auto).
type	Fragility definition: "curvature_ratio" (default) or "inverse_slope".

`normalize`	Logical. Divide integral by interval length? Default `TRUE`.
`t_obs`	Optional numeric vector of observed treatment values for support density. Default `NULL`.
`...`	Ignored.

Value

A list with elements:

`interval` Named numeric vector `c(a, b)` after clamping to the observed support.

`type` Fragility type used.

`integral_fragility` Trapezoidal integral of fragility over $[a, b]$.

`average_fragility` Integral divided by interval length.

`normalized` Logical flag.

`curve` The full `fragility_curve` data frame.

Examples

```
dat <- simulate_dose_response(200, dgp = "threshold")
fit <- causal_spline(Y ~ T | X1 + X2 + X3, data = dat)
region_fragility(fit, a = 2, b = 5)
```

`simulate_dose_response`

Simulate nonlinear dose-response data

Description

Generates synthetic observational data with a continuous treatment, confounders, and a nonlinear dose-response outcome. Useful for testing, benchmarking, and illustrating the CausalSpline package.

Usage

```
simulate_dose_response(
  n = 500L,
  dgp = c("threshold", "diminishing", "nonmonotone", "linear", "sinusoidal"),
  confounding = 0.5,
  sigma_y = 1,
  seed = NULL
)
```

Arguments

n	Integer. Sample size. Default 500.
dgp	Character string. Data-generating process: "threshold" Flat below treatment = 3, steep rise above. "diminishing" Concave relationship with diminishing returns. "nonmonotone" Inverted-U relationship. "linear" Standard linear effect (useful as baseline). "sinusoidal" Oscillatory effect (difficult, high df needed).
confounding	Numeric scalar in [0, 1]. Strength of confounding (correlation between treatment and confounders). Default 0.5.
sigma_y	Numeric. Standard deviation of the outcome noise. Default 1.
seed	Integer or NULL. Random seed. Default NULL.

Value

A data frame with columns:

Y Observed outcome.

T Observed (confounded) treatment.

X1, X2, X3 Confounders.

Y0 Potential outcome at T = 0 (for evaluation).

true_effect f(T) at each observed T value.

Examples

```
dat <- simulate_dose_response(n = 200, dgp = "threshold", seed = 1)
plot(dat$T, dat$true_effect, type = "l",
     xlab = "Treatment", ylab = "True causal effect")

dat2 <- simulate_dose_response(n = 200, dgp = "nonmonotone",
                               confounding = 0.8, seed = 42)
hist(dat2$T, main = "Treatment distribution", xlab = "T")
```

---

summary.causal_spline *Summary method for causal_spline objects*

Description

Summary method for causal_spline objects

Usage

```
## S3 method for class 'causal_spline'
summary(object, ...)
```

Arguments

object	A causal_spline object.
...	Ignored.

Value

Invisibly returns the input causal_spline object object, unchanged. The function is called for its side effect of printing a detailed summary to the console, including the original call, estimation method, spline configuration, treatment variable statistics, IPW diagnostics (effective sample size and weight range, if applicable), and a table of the estimated dose-response curve at seven representative percentile points (treatment value, point estimate, standard error, and confidence interval bounds).

---

trim_weights *Trim extreme IPW weights*

Description

Winsorises (clips) weights at specified quantiles to reduce variance from extreme propensity scores. This is a standard stabilisation step when using inverse probability weighting for continuous treatments.

Usage

```
trim_weights(weights, quantiles = c(0.01, 0.99))
```

Arguments

weights	Numeric vector of (possibly unstabilised) IPW weights.
quantiles	Numeric vector of length 2: lower and upper quantile thresholds. Default c(0.01, 0.99).

Value

Numeric vector of trimmed weights (same length as input).

Examples

```
w <- rexp(200)
w_trimmed <- trim_weights(w, c(0.01, 0.99))
summary(w_trimmed)
```

Index