

# Package ‘SemiPar’

September 23, 2014

**Version** 1.0-4.1

**Title** Semiparametric Regression

**Author** Matt Wand <mwand@uts.edu.au>

**Maintainer** Billy Aung Myint <r@use-r.com>

**Imports** MASS, cluster, nlme

**Suggests** lattice

**Description** Functions for semiparametric regression analysis, to complement the book: Ruppert, D., Wand, M.P. and Carroll, R.J. (2003). Semiparametric Regression. Cambridge University Press.

**License** GPL (>= 2)

**URL** <http://www.use-r.com/SPmanu.pdf>

**Repository** CRAN

**Date/Publication** 2014-09-23 17:33:48

**NeedsCompilation** no

## R topics documented:

age.income . . . . .	2
bpd . . . . .	3
calif.air.poll . . . . .	4
copper . . . . .	4
elec.temp . . . . .	5
ethanol . . . . .	6
fitted.spm . . . . .	7
fossil . . . . .	8
fuel.frame . . . . .	9
janka . . . . .	10
lidar . . . . .	10

lines.spm . . . . .	11
milan.mort . . . . .	12
monitor.mercury . . . . .	13
onions . . . . .	14
pig.weights . . . . .	15
plot.spm . . . . .	16
predict.spm . . . . .	17
print.spm . . . . .	18
ragweed . . . . .	19
residuals.spm . . . . .	20
retire.plan . . . . .	21
salinity . . . . .	22
sausage . . . . .	23
scallop . . . . .	24
sitka . . . . .	25
spm . . . . .	26
summary.spm . . . . .	27
term.structure . . . . .	29
trade.union . . . . .	30
ustemp . . . . .	31

<b>Index</b>	<b>32</b>
--------------	-----------

---

age.income	<i>Age/income data</i>
------------	------------------------

---

## Description

The age.income data frame has 205 pairs observations on Canadian workers from a 1971 Canadian Census Public Use Tape (Ullah, 1985).

## Usage

```
data(age.income)
```

## Format

This data frame contains the following columns:

**age** age in years.

**log.income** logarithm of income.

## Source

Ullah, A. (1985). Specification analysis of econometric models. *Journal of Quantitative Economics*, 2, 187-209.

## References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

## Examples

```
library(SemiPar)
data(age.income)
attach(age.income)
plot(age, log.income)
```

---

bpd

*Bronchopulmonary dysplasia data*

---

## Description

The bpd data frame has data on 223 human babies.

## Usage

```
data(bpd)
```

## Format

This data frame contains the following columns:

**birthweight** birthweight of baby (grammes).

**BPD** an indicator of presence of bronchopulmonary dysplasia (BPD): 0=absent, 1=present.

## Source

Pagano, M. and Gauvreau, K. (1993). *Principles of Biostatistics*. Duxbury Press.

## References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

## Examples

```
library(SemiPar)
data(bpd)
attach(bpd)
plot(birthweight, BPD)
boxplot(split(birthweight, BPD), col="green")
```

---

calif.air.poll      *California air pollution data*

---

### Description

The `calif.air.poll` data frame has 345 sets of observations ozone level and meteorological variables in Upland, California, U.S.A., in 1976.

### Usage

```
data(calif.air.poll)
```

### Format

This data frame contains the following columns:

**ozone.level** Ozone concentration (ppm) at Sandburg Air Force Base.

**daggett.pressure.gradient** Pressure gradient at Daggett, California.

**inversion.base.height** Inversion base height, feet.

**inversion.base.temp** Inversion base temperature, degrees Fahrenheit.

### Source

Brieman, L. and Friedman, J. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *Journal of the American Statistical Association*, **80**, 580–619.

### Examples

```
library(SemiPar)
data(calif.air.poll)
pairs(calif.air.poll)
```

---

copper      *Copper data*

---

### Description

The `copper` data frame has 442 sets of observations from a simulation based on a stockpile of mined material in the former Soviet Union. Boreholes have been drilled into the dump. The drill core is cut every 5 metres and assayed for copper and cobalt content in percentage by weight.

### Usage

```
data(copper)
```

**Format**

This data frame contains the following columns:

**sample.num** sample number.  
**id** sample identification number.  
**zone** zone code.  
**xcoord** x co-ordinate.  
**ycoord** y co-ordinate.  
**zcoord** z co-ordinate.  
**grade** grade measurement.  
**core.length** percentage of copper.

**Source**

Clark, I. and Harper, W.V. (2000). *Practical Geostatistics 2000*. Columbus, Ohio: Ecosse North America Llc.

**Examples**

```
library(SemiPar)
data(copper)
pairs(copper[,4:7])
```

---

elec.temp

*Electricity usage and temperature data*

---

**Description**

The elec.temp data frame has 55 observations on monthly electricity usage and average temperature for a house in Westchester County, New York, USA.

**Usage**

```
data(elec.temp)
```

**Format**

This data frame contains the following columns:

**usage** monthly electricity usage (kilowatt-hours) from a house in Westchester County, New York, USA.  
**temp** average temperature (degrees Fahrenheit) for the corresponding month.

**Source**

Chatterjee, S., Handcock, M. and Simonoff, J.S. (1995). *A Casebook for a First Course in Statistics and Data Analysis*, New York: John Wiley & Sons.

## References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

## Examples

```
library(SemiPar)
data(elec.temp)
attach(elec.temp)
plot(usage,temp)
```

---

ethanol

*Ethanol data*

---

## Description

The ethanol data frame contains 88 sets of measurements for variables from an experiment in which ethanol was burned in a single cylinder automobile test engine.

## Usage

```
data(ethanol)
```

## Format

This data frame contains the following columns:

**NOx** the concentration of nitric oxide (NO) and nitrogen dioxide (NO<sub>2</sub>) in engine exhaust, normalized by the work done by the engine.

**C** the compression ratio of the engine

**E** the equivalence ratio at which the engine was run – a measure of the richness of the air/ethanol mix.

## Source

Brinkman, N.D. (1981). Ethanol fuel – a single-cylinder engine study of efficiency and exhaust emissions. *SAE transactions* Vol. 90, No 810345, 1410–1424.

## References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(ethanol)
pairs(ethanol)
```

---

fitted.spm

*Fitted values for semiparametric regression.*

---

**Description**

Extracts fitted values from a semiparametric regression fit object.

**Usage**

```
## S3 method for class 'spm'
fitted(object,...)
```

**Arguments**

object            a fitted spm object as produced by spm().  
...                other possible arguments.

**Details**

Extracts fitted from a semiparametric regression fit object. The fitted are defined to be the set of values obtained when the predictor variable data are substituted into the fitted regression model.

**Value**

The vector of fitted.

**Author(s)**

M.P. Wand <mwand@uow.edu.au> (other contributors listed in SemiPar Users' Manual).

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

Ganguli, B. and Wand, M.P. (2005)  
*SemiPar 1.0 Users' Manual*.  
<http://www.uow.edu.au/~mwand/SPmanu.pdf>

**See Also**

[spm plot.spm](#) [lines.spm](#) [predict.spm](#) [summary.spm](#) [residuals.spm](#)

### Examples

```
library(SemiPar)
data(fossil)
attach(fossil)
fit <- spm(strontium.ratio~f(age))
plot(fit)
points(age,fitted(fit),col="red")
```

---

fossil

*Fossil data*

---

### Description

The fossil data frame has 106 observations on fossil shells.

### Usage

```
data(fossil)
```

### Format

This data frame contains the following columns:

**age** age in millions of years

**strontium.ratio** ratios of strontium isotopes

### Source

Bralower, T.J., Fullagar, P.D., Paull, C.K., Dwyer, G.S. and Leckie, R.M. (1997). Mid-cretaceous strontium-isotope stratigraphy of deep-sea sections. *Geological Society of America Bulletin*, **109**, 1421-1442.

### References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

### Examples

```
library(SemiPar)
data(fossil)
attach(fossil)
plot(age, strontium.ratio)
```



---

`fuel.frame`*Automobile data from consumer reports*

---

**Description**

The `fuel.frame` data frame contains data on 5 variables (columns) for 117 cars (rows).

**Usage**

```
data(fuel.frame)
```

**Format**

This data frame contains the following columns:

**car.name** character variable giving the name (make) of the car

**Weight** the weight of the car in pounds.

**Disp.** the engine displacement in litres.

**Mileage** gas mileage in miles/gallon.

**Fuel** a derived variable concerning fuel efficiency.

**Type** a factor giving the general type of car. The levels are: Small ,Sporty , Compact , Medium , Large , Van.

**Source**

Consumer Reports, April, 1990, pp. 235-288.

**References**

Chambers, J.M. and Hastie, T.J. (eds.) (1992)

*Statistical Models in S.*

Wadsworth and Brooks, Pacific Grove, California.

**Examples**

```
library(SemiPar)
data(fuel.frame)
pairs(fuel.frame)
par(mfrow=c(2,2))
fuel.fit <- lm(Fuel ~ Weight + Disp.,fuel.frame)
plot(fuel.fit,ask=FALSE)
par(mfrow=c(1,1))
```

---

janka

*Janka hardness data*

---

### Description

The janka data frame has 36 observations on Australian timber samples.

### Usage

```
data(janka)
```

### Format

This data frame contains the following columns:

**dens** a measure of density of the timber.

**hardness** the Janka hardness (structural property) of the timber.

### Source

Williams, E.J. (1959) *Regression Analysis*, New York: John Wiley & Sons.

### References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

### Examples

```
library(SemiPar)
data(janka)
attach(janka)
plot(dens, hardness)
```

---

lidar

*LIDAR data*

---

### Description

The lidar data frame has 221 observations from a light detection and ranging (LIDAR) experiment.

### Usage

```
data(lidar)
```

**Format**

This data frame contains the following columns:

**range** distance travelled before the light is reflected back to its source.

**logratio** logarithm of the ratio of received light from two laser sources.

**Source**

Sigrist, M. (Ed.) (1994). *Air Monitoring by Spectroscopic Techniques* (Chemical Analysis Series, vol. 197). New York: Wiley.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(lidar)
attach(lidar)
plot(range,logratio)
```

---

lines.spm	<i>Add a curves to an existing plot.</i>
-----------	--

---

**Description**

Takes a fitted spm object produced by spm() and adds a curve. The function is only appropriate in the case of a single predictor.

**Usage**

```
## S3 method for class 'spm'
lines(x,...)
```

**Arguments**

x a fitted spm object as produced by spm().  
... other graphics parameters described in Appendix B of the SemiPar Users' Manual <http://uow.edu.au/~mwand/SPmanu.pdf>

**Details**

Takes a fitted spm object produced by spm() and adds a curve. The function is only appropriate in the case of a single predictor.

**Value**

The function adds a curve to a plot.

**Author(s)**

M.P. Wand <mwand@uow.edu.au> (other contributors listed in SemiPar Users' Manual).

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

Ganguli, B. and Wand, M.P. (2005)  
*SemiPar 1.0 Users' Manual*.  
<http://www.uow.edu.au/~mwand/SPmanu.pdf>

**See Also**

[spm plot](#). [spm predict](#). [spm summary](#). [spm residuals](#). [spm fitted](#). [spm](#)

**Examples**

```
library(SemiPar)
data(fossil)
attach(fossil)
fit <- spm(strontium.ratio~f(age))
plot(fossil,type="n")
lines(fit)
points(fossil)

# Now do several customisations

op <- par(bg="white")
par(bg="honeydew")
plot(fossil,type="n")
lines(fit,col="green",lwd=5,shade.col="mediumpurple1")
points(fossil,col="orange",pch=16)
par(op)
```

---

milan.mort

*Milan mortality data*

---

**Description**

The `milan.mort` data frame has data on 3652 consecutive days (10 consecutive years: 1st January, 1980 to 30th December, 1989) for the city of Milan, Italy.

**Usage**

```
data(milan.mort)
```

**Format**

This data frame contains the following columns:

**day.num** number of days since 31st December, 1979

**day.of.week** 1=Monday,2=Tuesday,3=Wednesday,4=Thursday, 5=Friday,6=Saturday,7=Sunday.

**holiday** indicator of public holiday: 1=public holiday, 0=otherwise.

**mean.temp** mean daily temperature in degrees Celcius.

**rel.humid** relative humidity.

**tot.mort** total number of deaths.

**resp.mort** total number of respiratory deaths.

**SO2** measure of sulphur dioxide level in ambient air.

**TSP** total suspended particles in ambient air.

**Source**

Vigotti, M.A., Rossi, G., Bisanti, L., Zanobetti, A. and Schwartz, J. (1996). Short term effect of urban air pollution on respiratory health in Milan, Italy, 1980-1989. *Journal of Epidemiology and Community Health*, **50**, S71-S75.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(milan.mort)
pairs(milan.mort,pch=".")
```

---

```
monitor.mercury
```

```
Mercury biomonitoning data
```

---

**Description**

The monitor.mercury data frame has 22 observations from sampling locations around a solid waste incinerator in Warren County, New Jersey, USA

**Usage**

```
data(monitor.mercury)
```

**Format**

This data frame contains the following columns:

**UTM.North** longitude of sampling location.

**UTM.East** latitude of sampling location.

**mercury.concentration** mercury concentration in dry sphagnum moss grown at the sampling location.

**Source**

Opsomer, J.D., Agras, J., Carpi, A. and Rodrigues, G. (1995), An application of locally weighted regression to airborne mercury deposition around an incinerator site, *Environmetrics*, **6**, 205-221.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(monitor.mercury)
pairs(monitor.mercury)
```

---

onions

*Onions data*

---

**Description**

The onions data frame contains 84 sets of observations from an experiment involving the production of white Spanish onions in two South Australian locations.

**Usage**

```
data(onions)
```

**Format**

This data frame contains the following columns:

**dens** areal density of plants (plants per square metre)

**yield** onion yield (grammes per plant).

**location** indicator of location: 0=Purnong Landing, 1=Virginia.

**Source**

Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach*. New York: Marcel Dekker.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(onions)
attach(onions)
points.cols <- c("red", "blue")
plot(dens, yield, col=points.cols[location+1], pch=16)
legend(100, 250, c("Purnong Landing", "Virginia"), col=points.cols, pch=rep(16, 2))
```

---

pig.weights

*Pig weight data*

---

**Description**

The pig.weights data frame has 9 repeated weight measures on 48 pigs.

**Usage**

```
data(pig.weights)
```

**Format**

This data frame contains the following columns:

**id.num** identification number of pig.

**num.weeks** number of weeks since measurements commenced.

**weight** bodyweight of pig "id.num" after "num.weeks" weeks.

**Source**

Diggle, P.J., Heagerty, P., Liang, K.-Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data, Second Edition*, Oxford: Oxford University Press.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

## Examples

```
library(SemiPar)
data(pig.weights)
library(lattice)
xyplot(weight~num.weeks,data=pig.weights,groups=id.num,type="b")
```

---

plot.spm

*Semiparametric regression plotting*

---

## Description

Takes a fitted spm object produced by `spm()` and plots the component smooth functions that make it up, on the scale of the linear predictor.

## Usage

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

## Arguments

`x` a fitted spm object as produced by `spm()`.  
`...` other graphics parameters described in Appendix B of the SemiPar Users' Manual <http://www.uow.edu.au/~mwand/SPmanu.pdf>

## Details

Produces plots with each panel corresponding to a component of the semiparametric regression model.

## Value

The function generates plots.

## Author(s)

M.P. Wand <[mwand@uow.edu.au](mailto:mwand@uow.edu.au)> (other contributors listed in SemiPar Users' Manual).

## References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>  
Ganguli, B. and Wand, M.P. (2005)  
*SemiPar 1.0 Users' Manual*.  
<http://www.uow.edu.au/~mwand/SPmanu.pdf>



**See Also**

[spm lines.spm predict.spm summary.spm](#)

**Examples**

```
library(SemiPar)
data(fossil)
attach(fossil)
fit <- spm(strontium.ratio~f(age))
plot(fit)

# Now do several customisations

op <- par(bg="white")
par(bg="honeydew")
plot(fit,ylim=range(strontium.ratio),col="green",
     lwd=5,shade.col="mediumpurple1",rug.col="blue")
points(age,strontium.ratio,col="orange",pch=16)
par(op)
```

---

predict.spm

*Semiparametric regression prediction.*

---

**Description**

Takes a fitted spm object produced by spm() and obtains predictions at new data values.

**Usage**

```
## S3 method for class 'spm'
predict(object,newdata,se,...)
```

**Arguments**

object	a fitted spm object as produced by spm().
newdata	a data frame containing the values of the predictors at which predictions are required. The columns should have the same name as the predictors.
se	when this is TRUE standard error estimates are returned for each prediction. The default is FALSE.
...	other arguments.

**Details**

Takes a fitted spm object produced by spm() and obtains predictions at new data values as specified by the 'newdata' argument. If 'se=TRUE' then standard error estimates are also obtained.

**Value**

If `se=FALSE` then a vector of predictions at 'newdata' is returned. If `se=TRUE` then a list with components named 'fit' and 'se' is returned. The 'fit' component contains the predictions. The 'se' component contains standard error estimates.

**Author(s)**

M.P. Wand <[mwand@uow.edu.au](mailto:mwand@uow.edu.au)> (other contributors listed in SemiPar Users' Manual).

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

Ganguli, B. and Wand, M.P. (2005)  
*SemiPar 1.0 Users' Manual*.  
<http://www.uow.edu.au/~mwand/SPmanu.pdf>

**See Also**

[spm](#) [lines.spm](#) [plot.spm](#) [summary.spm](#)

**Examples**

```
library(SemiPar)
data(fossil)
attach(fossil)
fit <- spm(strontium.ratio~f(age))
newdata.age <- data.frame(age=c(90,100,110,120,130))
preds <- predict(fit,newdata=newdata.age,se=TRUE)
print(preds)

plot(fit,xlim=c(90,130))
points(unlist(newdata.age),preds$fit,col="red")
points(unlist(newdata.age),preds$fit+2*preds$se,col="blue")
points(unlist(newdata.age),preds$fit-2*preds$se,col="green")
```

---

```
print.spm
```

*Prints semiparametric regression fit object.*

---

**Description**

Prints a brief description of a semiparametric regression fit object to the screen.

**Usage**

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

**Arguments**

x                    a fitted spm object as produced by spm().  
...                  other possible arguments.

**Details**

Prints a brief description of a semiparametric regression fit object to the screen.

**Value**

The function prints to the screen.

**Author(s)**

M.P. Wand <mwand@uow.edu.au> (other contributors listed in SemiPar Users' Manual).

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

Ganguli, B. and Wand, M.P. (2005)  
*SemiPar 1.0 Users' Manual*.  
<http://www.uow.edu.au/~mwand/SPmanu.pdf>

**See Also**

[spm plot](#). [spm lines](#). [spm predict](#). [spm summary](#). [spm residuals](#). [spm fitted](#). [spm](#)

**Examples**

```
library(SemiPar)
data(fossil)
attach(fossil)
fit <- spm(strontium.ratio~f(age))
print(fit)
```

---

ragweed

*Ragweed data*

---

**Description**

The ragweed data frame has data on ragweed levels and meteorological variables for 335 days in Kalamazoo, Michigan, U.S.A.

**Usage**

```
data(ragweed)
```

**Format**

This data frame contains the following columns:

**ragweed** ragweed level (grains per cubic metre).

**year** one of 1991, 1992, 1993 or 1994.

**day.in.seas** day number in the current ragweed pollen season.

**temperature** temperature of following day (degrees Fahrenheit).

**rain** indicator of significant rain the following day: 1=at least 3 hours of steady or brief but intense rain, 0=otherwise.

**wind.speed** wind speed forecast for following day (knots).

**Source**

Stark, P. C., Ryan, L. M., McDonald, J. L. and Burge, H. A. (1997). Using meteorologic data to model and predict daily ragweed pollen levels. *Aerobiologia*, **13**, 177-184.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(ragweed)
pairs(ragweed,pch=".")
```

---

residuals.spm

*Residuals for semiparametric regression.*


---

**Description**

Extracts residuals from a semiparametric regression fit object.

**Usage**

```
## S3 method for class 'spm'
residuals(object,...)
```

**Arguments**

**object** a fitted spm object as produced by spm().  
**...** other possible arguments.

**Details**

Extracts residuals from a semiparametric regression fit object. The residuals are defined to be the difference between the response variable and the fitted values.

**Value**

The vector of residuals.

**Author(s)**

M.P. Wand <mwand@uow.edu.au> (other contributors listed in SemiPar Users' Manual).

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

Ganguli, B. and Wand, M.P. (2005)  
*SemiPar 1.0 Users' Manual*.  
<http://www.uow.edu.au/~mwand/SPmanu.pdf>

**See Also**

[spm](#) [plot.spm](#) [lines.spm](#) [predict.spm](#) [summary.spm](#) [fitted.spm](#)

**Examples**

```
library(SemiPar)
data(fossil)
attach(fossil)
fit <- spm(strontium.ratio~f(age))
plot(age,residuals(fit))
abline(0,0)
```

---

retire.plan

*Retirement plan data*

---

**Description**

The retire.plan data frame has data on "401(k)" retirement plans for employees of 92 firms managed by a company code-named Best Retirement Inc. (BRI).

**Usage**

```
data(retire.plan)
```

**Format**

This data frame contains the following columns:

**contrib** contribution to retirement plan at end of first year

**group** 1=client has group life of group health insurance policy, 0=otherwise.

**turnover** employee turnover rate.

**eligible** number of employees eligible to participate in 401(k) plans.

**vest** 1=plan has immediate vesting of employer contributions, 0=otherwise.

**failsafe** 1=plan has a fail-safe provision, 0=otherwise.

**match** percentage of contributions matched by the employer.

**salary** average annual employee salary in dollars.

**estimate** underwriter's estimate of end-of-year contributions in dollars.

**susan** 1=plan was sold by a sales representative who has been specifically trained to deal exclusively with 401(k) plans (code-named Susan Shepard).

**Source**

Bryant, P.G. and Smith, M.A. (1995). *Practical data analysis: case studies in business statistics*. Chicago: Irwin.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(retire.plan)
pairs(retire.plan)
```

---

salinity

*Salinity data*

---

**Description**

The salinity data frame has 28 observations on hydrological measurements from Pamlico Sound, North Carolina, USA.

**Usage**

```
data(salinity)
```

**Format**

This data frame contains the following columns:

**salinity** salinity in Pamlico Sound.

**lagged.salinity** salinity in Pamlico Sound during the previous six weeks.

**trend** trend=1 if the data is the first six-week period of the spring, and so forth. Used to detect possible effects of the seasonal warming trend.

**discharge** discharge of fresh water from rivers into the sound.

**Source**

Ruppert, D, and Carroll, R.J. (1980), Trimmed least squares estimation in the linear model, *Journal of the American Statistical Association*, **75**, 828-838.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(salinity)
pairs(salinity)
```

---

sausage

*Sausage data*

---

**Description**

The sausage data frame has data on 54 'hot dog' sausages.

**Usage**

```
data(sausage)
```

**Format**

This data frame contains the following columns:

**type** type of meat.

**calories** number of calories.

**sodium** measure of sodium content.

**Source**

Moore, D.S. and McCabe, G.P. (2003). *Introduction to the Practice of Statistics, Fourth Edition*, W.H. Freeman and Company.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(sausage)
attach(sausage)
points.cols <- c("red", "blue", "green")
plot(sodium, calories, col=points.cols[type], pch=16)
legend(200, 180, c("beef", "pork", "poultry"), col=points.cols, pch=rep(16, 3))
```

---

scallop

*Scallop abundance data*

---

**Description**

The scallop data frame has 148 triplets concerning scallop abundance; based on a 1990 survey cruise in the Atlantic continental shelf off Long Island, New York, U.S.A.

**Usage**

```
data(scallop)
```

**Format**

This data frame contains the following columns:

**latitude** degrees latitude (north of the Equator).

**longitude** degrees longitude (west of Greenwich).

**tot.catch** size of scallop catch at location specified by "latitude" and "longitude".

**Source**

Ecker, M.D. and Heltshe, J.F. (1994). Geostatistical estimates of scallop abundance. In *Case Studies in Biometry*. Lange, N., Ryan, L., Billard, L., Brillinger, D., Conquest, L. and Greenhouse, J. (eds.) New York: John Wiley & Sons, 107-124.



## References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

## Examples

```
library(SemiPar)
data(scallop)
pairs(scallop)
```

---

sitka	<i>Sitka spruce data</i>
-------	--------------------------

---

## Description

The `sitka` data frame contains measurements of log-size for 79 Sitka spruce trees grown in normal or ozone-enriched environments. Within each year, the data are organised in four blocks, corresponding to four controlled environment chambers. The first two chambers, containing 27 trees each, have an ozone-enriched atmosphere, the remaining two, containing 12 and 13 trees respectively, have a normal (control) atmosphere.

## Usage

```
data(sitka)
```

## Format

This data frame contains the following columns:

**id.num** identification number of tree.  
**order** time order ranking within each tree.  
**days** time in days since 1st January, 1988.  
**log.size** tree size measured on a logarithmic scale.  
**ozone** indicator ozone treatment: 0=control,1=ozone.

## Source

Diggle, P.J., Heagerty, P., Liang, K.-Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data, Second Edition*, Oxford: Oxford University Press.

## References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(sitka)
attach(sitka)
library(lattice)
ozone.char <- rep("control",nrow(sitka))
ozone.char[ozone==1] <- "ozone"
xyplot(log.size~days|ozone.char,data=sitka,groups=id.num,type="b")
```

spm

*Fit a SemiParametric regression Model***Description**

spm is used to fit semiparametric regression models using the mixed model representation of penalized splines (per Ruppert, Wand and Carroll, 2003).

**Usage**

```
spm(form,random=NULL,group=NULL,family="gaussian",
     spar.method="REML",omit.missing=NULL)
```

**Arguments**

form	a formula describing the model to be fit. Note, that an intercept is always included, whether given in the formula or not.
random	"random=~1" specifies inclusion of a random intercept according to the groups specified by the "group" argument.
group	a vector of labels for specifying groups.
family	for specification of the type of likelihood model assumed in the fitting. May be "gaussian","binomial" or "poisson"
spar.method	method for automatic smoothing parameter selection. May be "REML" (restricted maximum likelihood) or "ML" (maximum likelihood).
omit.missing	a logical value indicating whether fields with missing values are to be omitted.

**Details**

See the SemiPar Users' Manual for details and examples.

**Value**

An list object of class "spm" containing the fitted model. The components are:

fit	mimics fit object of lme() for family="gaussian" and glmmPQL() for family="binomial" or family="poisson".
info	information about the inputs.
aux	auxiliary information such as variability estimates.

**Author(s)**

M.P. Wand <mwand@uow.edu.au> (other contributors listed in SemiPar Users' Manual).

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

Ganguli, B. and Wand, M.P. (2005)  
*SemiPar 1.0 Users' Manual*.  
<http://www.uow.edu.au/~mwand/SPmanu.pdf>

**See Also**

[gam](#) (in package 'mgcv') [lme](#) (in package 'nlme') [glmPQL](#) (in package 'MASS') [plot.spm](#) [summary.spm](#)

**Examples**

```
library(SemiPar)
data(fossil)
attach(fossil)
fit <- spm(strontium.ratio~f(age))
plot(fit)
summary(fit)

data(calif.air.poll)
attach(calif.air.poll)
fit <- spm(ozone.level ~ f(daggett.pressure.gradient)+
          f(inversion.base.height) +
          f(inversion.base.temp))

summary(fit)
par(mfrow=c(2,2))
plot(fit)

# The SemiPar User Manual contains several other examples
# and details of plotting parameters.
#
# The current version of the manual is posted on the web-site:
#
#   www.uow.edu.au/~mwand/SPmanu.pdf
```

**Description**

Takes a fitted `spm` object produced by `spm()` and summarises the fit.

**Usage**

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

**Arguments**

object	a fitted spm object as produced by spm().
...	other arguments.

**Details**

Produces tables for the linear (parametric) and non-linear (nonparametric) components. The linear table provides coefficient estimates, standard errors and p-values. The non-linear table provides degrees of freedom values and other information.

**Value**

The function generates summary tables.

**Author(s)**

M.P. Wand <mwand@uow.edu.au> (other contributors listed in SemiPar Users' Manual).

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

Ganguli, B. and Wand, M.P. (2005)  
*SemiPar 1.0 Users' Manual*.  
<http://www.uow.edu.au/~mwand/SPmanu.pdf>

**See Also**

[spm plot.spm predict.spm](#)

**Examples**

```
library(SemiPar)  
data(onions)  
attach(onions)  
log.yield <- log(yield)  
fit <- spm(log.yield~location+f(dens))  
summary(fit)
```

---

term.structure	<i>Term structure data</i>
----------------	----------------------------

---

### Description

The term.structure data frame has 117 observations on the prices of U.S. STRIPS (Separate Trading on Registered Interest and Principal of Securities) on December 31, 1995.

### Usage

```
data(term.structure)
```

### Format

This data frame contains the following columns:

**time.to.maturity** time in years between 31st December, 1995, and the date on which the STRIPS matures.

**price** price of the STRIPS as a percent of par.

### Source

University of Houston Fixed Income Database.

### References

Jarrow, R., Ruppert, D., and Yu, Y. (2004). Estimating the term structure of corporate debt with a semiparametric penalized spline model, *Journal of the American Statistical Association*, **99**, 57-66.

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

### Examples

```
library(SemiPar)
data(term.structure)
attach(term.structure)
plot(time.to.maturity,price)
```

---

`trade.union`*Trade union data*

---

**Description**

The `trade.union` data frame has data on 534 U.S. workers.

**Usage**

```
data(trade.union)
```

**Format**

This data frame contains the following columns:

**years.educ** number of years of education.

**south** indicator of living in southern region of U.S.A.

**female** gender indicator: 0=male,1=female.

**years.experience** number of years of work experience

**union.member** indicator of trade union membership: 0=non-member, 1=member.

**wage** wages in dollars per hour.

**age** age in years.

**race** 1=black, 2=Hispanic, 3=white.

**occupation** 1=management, 2=sales, 3=clerical, 4=service, 5=professional, 6=other.

**sector** 0=other, 1=manufacturing, 2=construction.

**married** indicator of being married: 0=unmarried, 1=married.

**Source**

Berndt, E.R. (1991) *The Practice of Econometrics*. New York: Addison-Wesley.

**References**

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

**Examples**

```
library(SemiPar)
data(trade.union)
pairs(trade.union,pch=".")
```

---

ustemp	<i>U.S. temperature data</i>
--------	------------------------------

---

### Description

The ustemp data frame has 56 observations on the temperature and location of 56 U.S. cities.

### Usage

```
data(ustemp)
```

### Format

This data frame contains the following columns:

**city** character string giving name of city and state (two-letter abbreviation).

**min.temp** average minimum January temperature.

**latitude** degrees latitude (north of Equator).

**longitude** degrees longitude (west of Greenwich).

### Source

Peixoto, J.L. (1990). A property of well-formulated polynomial regression models. *American Statistician*, **44**, 26-30.

### References

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003)  
*Semiparametric Regression* Cambridge University Press.  
<http://stat.tamu.edu/~carroll/semiregbook/>

### Examples

```
library(SemiPar)
data(ustemp)
attach(ustemp)
grey.levs <- min.temp+20
col.vec <- paste("grey", as.character(grey.levs), sep="")
plot(-longitude, latitude, col=col.vec, pch=16, cex=3, xlim=c(-130, -60))
text(-longitude, latitude, as.character(city))
```

# Index

## \*Topic **datasets**

- age.income, 2
- bpd, 3
- calif.air.poll, 4
- copper, 4
- elec.temp, 5
- ethanol, 6
- fossil, 8
- fuel.frame, 9
- janka, 10
- lidar, 10
- milan.mort, 12
- monitor.mercury, 13
- onions, 14
- pig.weights, 15
- ragweed, 19
- retire.plan, 21
- salinity, 22
- sausage, 23
- scallop, 24
- sitka, 25
- term.structure, 29
- trade.union, 30
- ustemp, 31

## \*Topic **models**

- fitted.spm, 7
- lines.spm, 11
- plot.spm, 16
- predict.spm, 17
- print.spm, 18
- residuals.spm, 20
- spm, 26
- summary.spm, 27

## \*Topic **nonlinear**

- spm, 26

## \*Topic **regression**

- fitted.spm, 7
- lines.spm, 11
- plot.spm, 16

- predict.spm, 17
- print.spm, 18
- residuals.spm, 20
- spm, 26
- summary.spm, 27

## \*Topic **smooth**

- fitted.spm, 7
- lines.spm, 11
- plot.spm, 16
- predict.spm, 17
- print.spm, 18
- residuals.spm, 20
- spm, 26
- summary.spm, 27

- age.income, 2

- bpd, 3

- calif.air.poll, 4
- copper, 4

- elec.temp, 5
- ethanol, 6

- fitted.spm, 7, 12, 19, 21
- fossil, 8
- fuel.frame, 9

- gam, 27
- glmPQL, 27

- janka, 10

- lidar, 10
- lines.spm, 7, 11, 17–19, 21
- lme, 27

- milan.mort, 12
- monitor.mercury, 13

- onions, 14



pig.weights, 15  
plot.spm, 7, 12, 16, 18, 19, 21, 27, 28  
predict.spm, 7, 12, 17, 17, 19, 21, 28  
print.spm, 18  
  
ragweed, 19  
residuals.spm, 7, 12, 19, 20  
retire.plan, 21  
  
salinity, 22  
sausage, 23  
scallop, 24  
sitka, 25  
spm, 7, 12, 17–19, 21, 26, 28  
summary.spm, 7, 12, 17–19, 21, 27, 27  
  
term.structure, 29  
trade.union, 30  
  
ustemp, 31