Package 'semsfa'

July 23, 2025

Type Package

Title Semiparametric Estimation of Stochastic Frontier Models

Version 1.2

Date 2025-04-28

Description Semiparametric Estimation of Stochastic Frontier Models following a two step procedure: in the first step semiparametric or nonparametric regression techniques are used to relax parametric restrictions of the functional form representing technology and in the second step variance parameters are obtained by pseudolikelihood estimators or by method of moments.

Depends R (>= 3.1.2), mgcv, np, gamlss

Imports moments, doParallel, foreach, iterators

License GPL

Encoding UTF-8

RoxygenNote 7.3.2

NeedsCompilation no

Author Giancarlo Ferrara [aut, cre],
Francesco Vidoli [aut]

Maintainer Giancarlo Ferrara <giancarlo.ferrara@gmail.com>

Repository CRAN

Date/Publication 2025-04-29 18:50:07 UTC

Contents

`semsfa-package` *Semiparametric Stochastic Frontier Models*

Description

Semiparametric Estimation of Stochastic Frontier Models following the two step procedure originally proposed by Fan et al (1996) and further developed also by Vidoli and Ferrara (2014) and Ferrara and Vidoli (2017). In the first step semiparametric or nonparametric regression techniques are used to relax parametric restrictions regards the functional form representing technology and in the second step variance parameters are obtained by pseudolikelihood or method of moment estimators.

Author(s)

Giancarlo Ferrara, Francesco Vidoli
Maintainer: Giancarlo Ferrara <giancarlo.ferrara@gmail.com>

References

Aigner., D., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6:21-37

Fan, Y., Li, Q., Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. Journal of Business & Economic Statistics 14:460-468

Ferrara G., Vidoli F. (2017). Semiparametric stochastic frontier models: A generalized additive model approach. European Journal of Operational Research 258:761-777.

Hastie, T., Tibshirani, R., 1990. Generalized additive models. Chapman & Hall

Kumbhakar, S.C., Lovell, C.A.K, 2000. Stochastic Frontier Analysis. Cambridge University Press, U.K

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18:435-444

Vidoli, F., Ferrara, G., 2014. Analyzing Italian citrus sector by semi-nonparametric frontier efficiency models. Empirical Economics doi 10.1007/s00181-014-0879-6

`efficiencies.semsfa` *Prediction of the individual efficiency score*

Description

This function calculates and returns efficiency estimates from semiparametric stochastic frontier models estimated with `semsfa()`.

Usage

```
efficiencies.semsfa(semobj, log.output = TRUE, ...)
```

Arguments

`semobj`	a stochastic frontier model object returned by `semsfa()`
`log.output`	logical. Is the dependent variable logged?
`...`	further arguments to the summary method are currently ignored

Details

The estimation of the individual efficiency score for a particular point (x, y) on a production frontier might be obtained from the Jondrow et al. (1982) procedure. Defining:

$$\sigma^2 = \sigma_u^2 + \sigma_v^2, u_*(x) = -\sigma_u^2\varepsilon/\sigma^2, \sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$$

it can be shown that:

$$u|\varepsilon \sim N^+(\mu_*(x), \sigma_*^2(x)).$$

We can use this distribution to obtain point previsions of u trought the mean of the conditional distribution:

$$E(u|\varepsilon) = \mu_* + \sigma_* f(-\mu_*/\sigma_*)/(1 - F(\mu_*/\sigma_*))$$

where f and F represent the standard Normal density and cumulative distribution function, respectively; alternative formulas for cost frontier models are easy to get (please see Kumbhakar and Lovell, 2000).

If the response variable is measured in logs, a point estimate of the efficiency is then provided by $\exp(-u) \in (0, 1)$; otherwise, `(fitt-u)/fitt` where `fitt` is the estimated output evaluated at the frontier, given the inputs.

Value

An object of class `semsfa` containing the following additional results:

`u`	the prediction of the individual efficiency score
`efficiencies`	point estimate of the efficiency

Author(s)

Giancarlo Ferrara and Francesco Vidoli

References

Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P., 1982. On the estimation of technical inefficiency in stochastic frontier production models. Journal of Econometrics 19, 233-238.

Kumbhakar, S.C., Lovell, C.A.K., 2000. Stochastic Frontier Analysis. Cambridge University Press, New York.

See Also

`semsfa`, `summary.semsfa`, `plot.semsfa`.

Examples

```
set.seed(0)
n<-200

#generate data
x<- runif(n, 1, 2)
fy<- 2+30*x-5*x^2
v<- rnorm(n, 0, 1)
u<- abs(rnorm(n,0,2.5))
#production frontier
y <- fy + v - u
dati<-data.frame(y,x)

#first-step: gam, second-step: fan (default)
o<-semsfa(y~s(x),dati,sem.method="gam")

#calculate efficiencies
a<-efficiencies.semsfa(o)
```

---

fan — *Pseudolikelihood estimator of the λ parameter*

Description

Pseudolikelihood estimator of the λ parameter

Usage

```
fan(lambda_fan, resp, Ey, ineffD)
```

Arguments

`lambda_fan`	the $\lambda = \sigma_u/\sigma_v$ parameter to be estimated
`resp`	the single response variable Y observed
`Ey`	the conditional expectation estimate obtained in the first step of the algorithm
`ineffD`	logical: TRUE for estimating a production function, FALSE for estimating a cost function; this is done for usage compatibility with frontier package

Value

Estimated λ parameter

Note

Internal usage only

Author(s)

Giancarlo Ferrara and Francesco Vidoli

References

Fan, Y., Li, Q., Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. Journal of Business & Economic Statistics 14:460-468

`plot.semsfa` *Default SEMSFA plotting*

Description

This function plots the semiparametric/nonparametric intermediate model object estimated in the first step of the algorithm and, if `efficiencies.semsfa()` is esecuted, individual point estimate of the efficiency.

Usage

```
## S3 method for class 'semsfa'
plot(x, g.type, mod, ...)
```

Arguments

`x`	a `semsfa` object as returned from `semsfa()` or `efficiencies.semsfa()`
`g.type`	a character string indicating the type of plot. Possible values are: "reg" to plot the semiparametric/nonparametric model object estimated in the first step from `semsfa()`, "eff" to draw point estimate of the efficiency obtained from `efficiencies.semsfa()`
`mod`	a character string indicating the plot style for g.type="eff": "hist" for histogram and "dens" for density plot
`...`	further arguments passed to plot.default.

Value

The function simply generates plots.

Author(s)

Giancarlo Ferrara and Francesco Vidoli

See Also

`semsfa`, `efficiencies.semsfa`.

Examples

```
set.seed(0)
n<-200

#generate data
x<- runif(n, 1, 2)
fy<- 2+30*x-5*x^2
v<- rnorm(n, 0, 1)
u<- abs(rnorm(n,0,2.5))
#production frontier
y <- fy + v - u
dati<-data.frame(y,x)

#first-step: gam, second-step: fan (default)
o<-semsfa(y~s(x),dati,sem.method="gam")
#the following plot will be like that generated by plot.gam
plot(o,g.type="reg")

#adding a covariate
z<- runif(n, 1, 2)
dati$z<-z

#first-step: kernel, second-step: fan (default)
o<-semsfa(y~x+z,dati,sem.method="kernel")
#the plot will be like that generated by a plot.npreg
## Not run: plot(o,g.type="reg")

#calculate efficiencies ...
a<-efficiencies.semsfa(o)
plot(a,g.type="eff",mod="dens")

#adding further parameters as for plot.default: col, main, xlim, ...
plot(a,g.type="eff",mod="dens",col=2,main="Density Efficiency",xlim=c(0,1),xlab="Efficiency")
```

---

semsfa *Semiparametric Estimation of Stochastic Frontier Models*

Description

Semiparametric Estimation of Stochastic Frontier Models following the two step procedure proposed by Fan et al (1996) and further developed by Ferrara and Vidoli (2017). In the first step semiparametric or nonparametric regression techniques are used to relax parametric restrictions regards the functional form of the frontier and in the second step variance parameters are obtained by pseudolikelihood or method of moments estimators.

Usage

```
semsfa(formula, data = list(), sem.method = "gam", var.method = "fan",
       ineffDecrease=TRUE, tol = 1e-05, n.boot=0,...)
```

Arguments

`formula`	an object of class "formula": a symbolic description of the model to be fitted. The details of model specification are given under 'Details'
`data`	a data frame containing the variables in the model
`sem.method`	a character string indicating the type of estimation method to be used in the first step for the semiparametric or nonparametric regression; possible values are "gam" (default), "kernel" or "loess"
`var.method`	the type of estimation method to be used in the second step for the variance components: "fan" (default) for Fan et al. (1996) approach and "mm" for method of moments
`ineffDecrease`	logical: TRUE (default) for estimating a production function, FALSE for estimating a cost function; this is done for usage compatibility with frontier package
`tol`	numeric. Convergence tolerance for pseudolikelihood estimators of variance parameters of the composed error term
`n.boot`	numeric. Number of bootstrap replicates to calculate standard error for the variance components, by default bootstrap standard errors will not be calculated (n.boot=0)
`...`	further arguments accepted by `mgcv::gam`, `np::npreg` or `loess`

Details

Parametric stochastic production frontier models, introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977), specify output in terms of a response function and a composite error term. The composite error term consists of a two-sided error representing random effects and a one-sided term representing technical inefficiency. The production stochastic frontier model can be written, in general terms, as:

$$y_i = f(x_i) + v_i - u_i, \qquad i = 1, ..., n,$$

where $Y_i \in R^+$ is the single output of unit i, $X_i \in R_p^+$ is the vector of inputs, $f(.)$ defines a production frontier relationship between inputs X and the single output Y. In following common practice, we assume that v and u are each identically independently distributed (iid) with $v \sim N(0, \sigma_v)$ and u distributed half-normally on the non-negative part of the real number line: $u \sim N^+(0, \sigma_u)$; furthermore, the probability density function of the composite disturbance can be rewritten in terms of $\lambda = \sigma_u/\sigma_v$ and $\sigma^2 = \sigma_v^2 + \sigma_u^2$ for the estimation algorithm. To overcome drawbacks due to the specification of a particular production function $f(\cdot)$ we consider the estimation of a Semiparametric Stochastic Production Frontier Models through a two step procedure originally proposed by Fan et al (1996): in the first step a semiparametric or nonparametric regression technique is used to estimate the conditional expectation, while in the second step λ and σ parameters are estimated by pseudolikelihood (via optimize) or by method of moments estimators (var.method argument). In the case of a cost function frontier (`ineffDecrease=FALSE`) the composite error term is $\varepsilon = v + u$. Vidoli and Ferrara (2014) and Ferrara and Vidoli(2017) suggest a Generalized Additive Model (GAM) framework in the first step even if any semiparametric or nonparametric tecnique may be used (Fan et al., 1996); for a recent review of SF analysis using R see Ferrara (2020).

The avalaible methods for the first step are:

- sem.method="gam" invokes gam() from mgcv;

- sem.method="gam.mono" invokes gamlss() from gamlss;
- sem.method="kernel" invokes npreg() from np;
- sem.method="loess" invokes loess() from stats.

Since in the first step different estimation procedure may be invoked from different packages, the formula argument has to be compatible with the corresponding function. The avalaible methods for the second step are:

- var.method="fan" pseudolikelihood;
- var.method="mm" Method of Moments.

Value

`semsfa()` returns an object of class `semsfa`. An `semsfa` object is a list containing the following components:

`formula`	the formula used
`y`	the response variable used as specified in `formula`
`data`	the data frame used
`call`	the matched call
`sem.method`	the type of semiparametric or nonparametric regression as given by sem.method ("gam", "gam.mono", "kernel", "loess")
`var.method`	the type of error component estimator ("fan", "mm")
`ineffDecrease`	logical, as given by ineffDecrease
`reg`	an object of class "gam", "gamlss"(gam.mono), "np"(kernel) or "loess" depending on sem.method
`reg.fitted`	fitted values on the "mean" frontier (semiparametric/non parametric regression)
`regkewness`	asymmetry index calculated on residuals obtained in the first step
`lambda`	λ estimate
`sigma`	σ estimate
`fitted`	fitted values on the frontier
`tol`	convergence tolerance for pseudolikelihood estimators used in optimize
`residual.df`	residual degree of freedom of the model
`bic`	'Bayesian Information Criterion' according to the formula `-2*log-likelihood+ log(n)*npar` where `npar` represents the number of parameters in the fitted model and `n` the number of observations
`n.boot`	number of bootstrap replicates used (default n.boot=0)
`boot.mat`	a matrix containing λ and σ values from each bootstrap replicate (if n.boot>0)
`b.se`	boostrapped standard errors for λ and σ (if n.boot>0)

Note

The function `summary` (i.e. `summary.semsfa`) can be used to obtain a summary of the results, `efficiencies.semsfa` to calculate efficiency scores and `plot` (i.e. `plot.semsfa`) to graph efficiency previsions and regression components (i.e. the first step).

You must take the natural logarithm of the response variable before fitting a stochastic frontier production or cost model.

Author(s)

Giancarlo Ferrara and Francesco Vidoli

References

Aigner., D., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6:21-37

Fan, Y., Li, Q., Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. Journal of Business & Economic Statistics 14:460-468

Ferrara G. (2020). Stochastic frontier models using R. In: Handbook of Statistics 42:299-326. Financial, Macro and Micro Econometrics Using R, Edited by Vinod and Rao. Elsevier

Ferrara G., Vidoli F. (2017). Semiparametric stochastic frontier models: A generalized additive model approach. European Journal of Operational Research 258:761-777.

Hastie, T., Tibshirani, R., 1990. Generalized additive models. Chapman & Hall

Kumbhakar, S.C., Lovell, C.A.K, 2000. Stochastic Frontier Analysis. Cambridge University Press, U.K

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from cobb-douglas production functions with composed error. International Economic Review 18:435-444

Vidoli, F., Ferrara, G., 2014. Analyzing Italian citrus sector by semi-nonparametric frontier efficiency models. Empirical Economics doi 10.1007/s00181-014-0879-6

See Also

`summary.semsfa`, `efficiencies.semsfa`, `plot.semsfa`.

Examples

```
set.seed(0)
n<-200

x<- runif(n, 1, 2)
v<- rnorm(n, 0, 1)
u<- abs(rnorm(n,0,2.5))

#cost frontier
fy<- 2+30*x+5*x^2
y <- fy + v + u

dati<-data.frame(y,x)
```

```
#first-step: gam, second-step: fan
o<-semsfa(y~s(x),dati,sem.method="gam",ineffDecrease=FALSE)
#first-step: gam, second-step: mm
## Not run: o<-semsfa(y~s(x),dati,sem.method="gam",ineffDecrease=FALSE,var.method="mm")
plot(x,y)
curve(2+30*x+5*x^2,add=TRUE)
points(sort(x),o$fitted[order(x)],col=3,type="l")

#production frontier
fy<- 2+30*x-5*x^2
y <- fy + v - u

dati<-data.frame(y,x)
#first-step: gam, second-step: fan
o<-semsfa(y~s(x),dati,sem.method="gam",ineffDecrease=TRUE)

plot(x,y)
curve(2+30*x-5*x^2,add=TRUE)
points(sort(x),o$fitted[order(x)],col=3,type="l")
```

---

summary.semsfa *Summary for* semsfa *object*

Description

Create and print summary results of a stochastic frontier model object returned by semsfa() with regard to the "CONDITIONAL EXPECTATION ESTIMATE" of the first step and to the "VARIANCE COMPONENTS ESTIMATE" of the compound error.

Usage

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

Arguments

object	an semsfa object returned by semsfa()
...	further arguments to the summary method are currently ignored

Details

Please note that if bootstrap is carried out the t-statistic is not reliable for testing the statistical significance of σ and λ, because these parameters are censored and cannot follow a t-distribution. We suggest to compare the BIC of the semiparametric estimated model with the base model.

Value

summary.semsfa returns a list of class summary.semsfa that is identical to an object returned by semsfa() with few modifications if bootstrap is carried out:

b.t t-statistic given the bootstrapped standard errors for λ and σ (b.se)

b.pv p-values of the t-statistic

Note

summary returns the same result if applied to an object created with semsfa or efficiencies.semsfa

Author(s)

Giancarlo Ferrara and Francesco Vidoli

See Also

semsfa, efficiencies.semsfa

Examples

```
#generate data
set.seed(0)
n<-200

x<- runif(n, 1, 2)
fy<- 2+30*x-5*x^2
v<- rnorm(n, 0, 1)
u<- abs(rnorm(n,0,2.5))
#production frontier
y <- fy + v - u

dati<-data.frame(y,x)

#first-step: gam, second-step: fan (default)
#without bootstrap
o<-semsfa(y~s(x),dati,sem.method="gam")
summary(o)

# ... with bootstrap
o<-semsfa(y~s(x),dati,sem.method="gam",n.boot=100)
summary(o)
```

Index