Package ‘micsr’

February 17, 2024

Version 0.1-1

Date 2024-02-12

Title Microeconometrics with R

Depends R (>= 4.0.0)

Imports Formula, Rdpack, sandwich, generics, tibble, ggplot2, rlang, dplyr, tidyr, purrr, knitr, magrittr, tidyselect

Suggests rmarkdown, modelsummary, bookdown, AER, censReg, sampleSelection, mlogit, MASS, lmtest

Description Functions, data sets and examples for the book: Yves Croissant (2024) ``Microeconometrics with R'', Chapman and Hall/CRC The R Series. The package includes a set of estimators for models used in microeconometrics, especially for count data and limited dependent variables. Test functions include score test, Hausman test, Vuong test, Sargan test and conditional moment test. A small subset of the data set used in the book is also included.

Encoding UTF-8

License GPL (>= 2)

URL https://www.r-project.org

VignetteBuilder knitr

NeedsCompilation yes

RoxygenNote 7.2.3

LazyData true

RdMacros Rdpack

Author Yves Croissant [aut, cre] (<https://orcid.org/0000-0002-4857-7736>)

Maintainer Yves Croissant <yves.croissant@univ-reunion.fr>

Repository CRAN

Date/Publication 2024-02-17 20:50:02 UTC

R topics documented:

micsr-package **micsr** *: Microeconometrics with R*

Description

The micsr package is the companion package to the book "Microeconometrics with R" (Chapman and Hall/CRC The R Series). It includes function to estimate and to test models, miscellanous tools and data sets:

Details

- functions to estimate models:
  - `binomreg`: binomial regression models, Rivers and Vuong (1988),
  - `bivprobit`: bivariate probit model
  - `clm`: constrained linear models,
  - `escount`: endogenous switching and selection model for count data, Terza (1998),
  - `expreg`: exponential conditional mean models, Mullahy (1997),
  - `loglm`: log-linear models,
  - `ordreg`: ordered regression models,
  - `poisreg`: poisson models,
  - `pscore`: matching, Dehejia and Wahba (2002),
  - `tobit1`: tobit-1 model, Tobin (1958), Smith and Blundel (1986), Powel (1986).
- functions for statistical tests and diagnostic:
  - `cmtest`: conditional moment tests, Newey (1985), Tauchen (1985),
  - `ftest`: F statistic,
  - `hausman`: Hausman's test, Hausman (1978),
  - `ndvuong`: non-degenerate Vuong test, Vuong (1989), Shi (2015),
  - `rsq`: different flavors of R squared,
  - `sargan`: Sargan's test, Sargan (1958),
  - `scoretest`: score, or Lagrange multiplier test.
- miscellanous tools
  - `gaze`: print a short summary of an object,
  - `dummy`: generate a set of dummy variables from a factor,
  - `newton`: Newton-Raphson optimization method, using the analytical gradient and hessian,
  - `mills`: compute the inverse mills ratio and its first two derivatives,
  - `stder`: extract the standard errors of a fitted model,
  - `npar`: extract the number of parameters in a fitted model.
- data sets:
  - `apples`: Apple production, Ivaldi and al. (1996), constrained linear model,
  - `birthwt`: Cigarette smoking and birth weigth, Mullahy (1997), exponentional conditional mean regression model,

- `charitable`: Intergenerational transmission of charitable giving, Wilhem (2008), Tobit-1 model,
- `cigmales`: Cigarettes consumption and smoking habits, Mullahy (1997), exponentional conditional mean regression mdodel,
- `drinks`: Physician advice on alcohol consumption, Kenkel and Terza (2001), endogenous switching model for count data,
- `forediv`: Foreign exchange derivatives use by large US bank holding companies, Adkins (2012), instrumental variable probit model,
- `fin_reform`: Political economy of financial reforms, Abiad and Mody (2005), ordered regression model,
- `housprod`: Household production, Kerkhofs and Kooreman (2003), bivariate probit model,
- `mode_choice`: Choice between car and transit, Horowitz (1993), probit model,
- `trade_protection`: Lobying and trade protection, Atschke and Sherlund (2006), instrumental variable Tobit-1 model,
- `trips`: Determinants of household trip taking, Terza (1998), endogenous switching model for count data,
- `turnout`: Turnout in Texas liquor referenda, Coate and Conlin (2004), non-degenerate Vuong test,
- `twa`: Temporary help jobs and permanent employment, Ichino, Mealli and Nannicini (2008), matching.

- vignettes:
  - charitable: Estimating the Tobit-1 model with the charitable data set
  - escount: Endogenous switching or sample selection models for count data
  - expreg: Exponentional conditional mean models with endogeneity
  - ndvvuong: Implementation of Shi's non-degeranate Vuong test

We tried to keep the sets of package on which **micsr** depends on as small as possible. **micsr** depends on **Formula**, **generics**, **Rdpack**, **knitr**, **sandwich** and on a subset of the **tidyverse** metapackage (**ggplot2**, **dplyr**, **purrr**, **tidyselect**, **magrittr**, **tibble**, **rlang**). We borrowed the gaussian quadrature function from the **statmod** package (Smyth and al., 2023), and the distribution function of quadratic forms in normal variables from the **CompQuadForm** package (Duchesne and Lafaye, 2010).

References

Abiad A, Mody A (2005). "Financial Reform: What Shakes It? What Shapes It?" *American Economic Review*, **95**(1), 66-88.

Adkins LC (2012). "Testing parameter significance in instrumental variables probit estimators: some simulation." *Journal of Statistical Computation and Simulation*, **82**(10), 1415-1436.

Coate S, Conlin M (2004). "A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence." *American Economic Review*, **94**(5), 1476-1504.

Dehejia RH, Wahba S (2002). "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *The Review of Economics and Statistics*, **84**(1), 151-161. ISSN 0034-6535, doi:10.1162/003465302317331982.

Duchesne P, de Micheaux PL (2010). "Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods." *Computational Statistics and Data Analysis*, **54**, 858-862.

Hausman JA (1978). “Specification Tests in Econometrics.” *Econometrica*, **46**(6), 1251–1271.

Ichino A, Mealli F, Nannicini T (2008). “From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and Their Sensitivity?” *Journal of Applied Econometrics*, **23**(3), 305–327.

Ivaldi M, Ladoux N, Ossard H, Simioni M (1996). “Comparing Fourier and translog specifications of multiproduct technology: Evidence from an incomplete panel of French farmers.” *Journal of Applied Econometrics*, **11**(6), 649–667. doi:10.1002/(sici)10991255(199611)11:6<649::aid-jae416>3.0.co;24, `http://dx.doi.org/10.1002/(sici)1099-1255(199611)11:6<649::aid-jae416>3.0.co;2-4`.

Kenkel DS, Terza JV (2001). “The effect of physician advice on alcohol consumption: count regression with an endogenous treatment effect.” *Journal of Applied Econometrics*, **16**(2), 165-184.

Kerkhofs M, Kooreman P (2003). “Identification and Estimation of a Class of Household Production Models.” *Journal of Applied Econometrics*, **18**(3), 337–369.

Matschke X, Sherlund SM (2006). “Do Labor Issues Matter in the Determination of U.S. Trade Policy? An Empirical Reevaluation.” *American Economic Review*, **96**(1), 405-421.

Mullahy J (1997). “Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior.” *The Review of Economics and Statistics*, **79**(4), 586-593.

Newey WK (1985). “Maximum Likelihood Specification Testing and Conditional Moment Tests.” *Econometrica*, **53**(5), 1047–1070.

Powell J (1986). “Symmetrically trimed least squares estimators for tobit models.” *Econometrica*, **54**, 1435–1460.

Rivers D, Vuong QH (1988). “Limited information estimators and exogeneity tests for simultaneous probit models.” *Journal of Econometrics*, **39**(3), 347-366.

Sargan JD (1958). “The Estimation of Economic Relationships using Instrumental Variables.” *Econometrica*, **26**(3), 393–415.

Shi X (2015). “A nondegenerate Vuong test.” *Quantitative Economics*, 85-121.

Smith R, Blundell R (1986). “An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply.” *Econometrica*, **54**(3), 679-85.

Smyth G, Chen L, Hu Y, Dunn P, Phipson B, Chen Y (2023). *statmod: Statistical Modeling*. R package version 1.5.0, `https://CRAN.R-project.org/package=statmod`.

Tauchen G (1985). “Diagnostic testing and evaluation of maximum likelihood models.” *Journal of Econometrics*, **30**(1), 415-443.

Terza JV (1998). “Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects.” *Journal of Econometrics*, **84**(1), 129-154.

Tobin J (1958). “Estimation of Relationships for Limited Dependent Variables.” *Econometrica*, **26**(1), 24-36.

Vuong QH (1989). “Likelihood Ratio Tests for Selection and Non-Nested Hypotheses.” *Econometrica*, **57**(2), 397-333.

Wilhelm MO (2008). “Practical Considerations for Choosing Between Tobit and SCLS or CLAD Estimators for Censored Regression Models with an Application to Charitable Giving.” *Oxford Bulletin of Economics and Statistics*, **70**(4), 559-582.

`apples` *Apple production*

Description

yearly observations of 173 farms from 1984 to 1986

Format

a tibble containing:

- id: farm's id
- year: year
- capital: capital stock
- labor: quantity of labor
- materials: quantity of materials
- apples: production of apples
- otherprod: other productions
- pc: price of capital
- pl: price of labor
- pm: price of materials

Source

Journal of Applied Econometrics Data Archive : http://qed.econ.queensu.ca/jae/

References

Ivaldi M, Ladoux N, Ossard H, Simioni M (1996). "Comparing Fourier and translog specifications of multiproduct technology: Evidence from an incomplete panel of French farmers." *Journal of Applied Econometrics*, **11**(6), 649–667. doi:10.1002/(sici)10991255(199611)11:6<649::aid-jae416>3.0.co;24, http://dx.doi.org/10.1002/(sici)1099-1255(199611)11:6<649::aid-jae416>3.0.co;2-4.

binmeans — *Compute the mean of a variable for bins of another variable*

Description

Plot average values of the outcome for bins of the forcing variable, a common plot in regression discontinuity analysis

Usage

```
binmeans(x, y, width = NULL, center = NULL, g = NULL, ...)

## Default S3 method:
binmeans(x, y, width = NULL, center = NULL, g = NULL, ..., name_g = "colour")

## S3 method for class 'formula'
binmeans(x, y, width = NULL, center = NULL, ...)

StatBinmeans

geom_binmeans(
  mapping = NULL,
  data = NULL,
  stat = "binmeans",
  position = "identity",
  ...,
  center = NULL,
  width = NULL,
  na.rm = FALSE,
  show.legend = NA,
  inherit.aes = TRUE
)
```

Arguments

`x, y`	either two numeric vector for the default method, or a formula and a data frame for the formula method
`width`	the width of the bins
`center`	the cuting value of the forcing variable
`g`	a grouping variable
`...`	further arguments
`name_g`	internally used by the geom
`mapping`	Set of aesthetic mappings created by `aes()`. If specified and `inherit.aes = TRUE` (the default), it is combined with the default mapping at the top level of the plot. You must supply `mapping` if there is no plot mapping.

data	The data to be displayed in this layer. There are three options: If NULL, the default, the data is inherited from the plot data as specified in the call to ggplot(). A data.frame, or other object, will override the plot data. All objects will be fortified to produce a data frame. See fortify() for which variables will be created. A function will be called with a single argument, the plot data. The return value must be a data.frame, and will be used as the layer data. A function can be created from a formula (e.g. ~ head(.x, 10)).
stat	The statistical transformation to use on the data for this layer, either as a ggproto Geom subclass or as a string naming the stat stripped of the stat_ prefix (e.g. "count" rather than "stat_count")
position	Position adjustment, either as a string naming the adjustment (e.g. "jitter" to use position_jitter), or the result of a call to a position adjustment function. Use the latter if you need to change the settings of the adjustment.
na.rm	If FALSE, the default, missing values are removed with a warning. If TRUE, missing values are silently removed.
show.legend	logical. Should this layer be included in the legends? NA, the default, includes if any aesthetics are mapped. FALSE never includes, and TRUE always includes. It can also be a named logical vector to finely select the aesthetics to display.
inherit.aes	If FALSE, overrides the default aesthetics, rather than combining with them. This is most useful for helper functions that define both data and aesthetics and shouldn't inherit behaviour from the default plot specification, e.g. borders().

Format

An object of class StatBinmeans (inherits from Stat, ggproto, gg) of length 4.

---

binomreg *Binomial regression*

Description

A unified interface for binomial regression models, including linear probability, probit and logit models

Usage

```
binomreg(
  formula,
  data,
  weights,
  subset,
  na.action,
  offset,
```

```
  contrasts = NULL,
  link = c("identity", "probit", "logit"),
  method = c("ml", "twosteps", "minchisq", "test"),
  start = NULL,
  ...
)

## S3 method for class 'binomreg'
residuals(object, ..., type = c("deviance", "pearson", "response"))

## S3 method for class 'binomreg'
glance(x, ...)

## S3 method for class 'binomreg'
predict(object, ..., type = c("response", "link"), newdata = NULL)
```

Arguments

`formula`	a symbolic description of the model
`data`	a data frame,
`subset, weights, na.action, offset, contrasts`	see `stats::lm`,
`link`	one of `"identity"`, `"probit"` and `"logit"` to fit respectively the linear probability, the probit and the logit model
`method`	`"ml"` for maximum likelihood (the only relevant method for a regression without instrumental variables), `"twosteps"` for two-steps estimator, `"minchisq"` for minimum chi-squared estimator and `"test"` to get the exogeneity test
`start`	a vector of starting values
`...`	further arguments
`object, x, type`	a `binomreg` object and the type of residuals for the `residuals` method
`newdata`	a new data frame for the `predict` method

Value

an object of class `c("binomreg", "micsr")`, see `micsr::micsr` for further details

Examples

```
pbt <- binomreg(mode ~ cost + ivtime + ovtime, data = mode_choice, link = 'probit')
lpm <- binomreg(mode ~ cost + ivtime + ovtime, data = mode_choice, link = 'identity')
summary(pbt, vcov = "opg")
```

birthwt *Cigarette smoking and birth weight*

Description

a cross-section of 1388 individuals from 1988

Format

a tibble containing:

- birthwt: birth weight
- cigarettes: number of cigarettes smoked per day during pregnancy
- parity: birth order
- race: a factor with levels `"other"` and `"white"`
- sex: a factor with levels `"female"` and `"male"`
- edmother: number of years of education of the mother
- edfather: number of years of education of the father
- faminc: family income
- cigtax: per-pack state excise tax on cigarettes

Source

kindly provided by John Mullahy

References

Mullahy J (1997). "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior." *The Review of Economics and Statistics*, **79**(4), 586-593.

bivprobit *Bivariate probit*

Description

Estimation of bivariate probit models by maximum likelihood

Usage

```
bivprobit(
  formula,
  data,
  weights,
  subset,
  na.action,
  offset,
  method = c("newton", "bfgs"),
  ...
)

## S3 method for class 'bivprobit'
logLik(object, ..., type = c("model", "null"))
```

Arguments

`formula`	a symbolic description of the model, a two-part left and right hand side formula
`data`	a data frame,
`subset, weights, na.action, offset`	see `stats::lm`,
`method`	the optimization method, one of `"newton"` and `"bfgs"`
`...`	further arguments
`object`	a `bivprobit` object
`type`	for the `logLik` method

Value

an object of class `micsr`, see `micsr::micsr` for further details

Examples

```
bivprobit(mjob | fjob ~ meduc + ychild + owner | feduc + ychild + owner , housprod)
```

`charitable` *Intergenerational transmission of charitable giving*

Description

a cross-section of 2384 households from 2001

Format

a tibble containing:

- donation: the amount of charitable giving
- donparents: the amount of charitable giving of the parents
- education: the level of education of household's head, a factor with levels "less_high_school", "high_school", "some_college", "college", "post_college"
- religion: a factor with levels "none", "catholic", "protestant", "jewish" and "other"
- income: income
- married: a dummy for married couples
- south: a dummy for households living in the south

Source

kindly provided by Mark Ottoni Wilhelm.

References

Wilhelm MO (2008). "Practical Considerations for Choosing Between Tobit and SCLS or CLAD Estimators for Censored Regression Models with an Application to Charitable Giving." *Oxford Bulletin of Economics and Statistics*, **70**(4), 559-582.

`cigmales` *Cigarette smoking behaviour*

Description

a cross-section of 6160 individuals from 1979 to 1980

Format

a tibble containing:

- cigarettes: number of daily cigarettes smoked
- habit: smoking habit stock measure
- price: state-level average per-pack price of cigarettes in 1979
- restaurant: an indicator of whether the individual's state of residence had restrictions on smoking in restaurants in place in 1979
- income: family income in thousands
- age: age in years
- educ: schooling in years
- famsize: number of family members
- race: a factor with levels "other" and "white"
- reslgth: number of years the state's restaurant smoking restrictions had been in place in 1979
- lagprice: one-year lag of cigarette price

Source

kindly provided by John Mullahy

References

Mullahy J (1997). "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior." *The Review of Economics and Statistics*, **79**(4), 586-593.

---

clm — *Constrained least squares*

Description

Compute the least squares estimator using linear constrains on the coefficients.

Usage

```
clm(x, R, q = NULL)

## S3 method for class 'clm'
vcov(object, ...)

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

Arguments

`x`	a linear model fitted by `lm`,
`R`	a matrix of constrains (one line for each constrain, one column for each coefficient),
`q`	an optional vector of rhs values (by default a vector of 0)
`object`	a `clm` object for the `summary` and the `vcov` methods
`...`	further arguments

Value

an object of class `clm` which inherits from class `lm`

Examples

```
# Cobb-Douglas production function for the apple data set
# First compute the total production
apples <- apples %>% mutate(prod = apples + otherprod)
# unconstrained linear model
cd <- lm(log(prod) ~ log(capital) + log(labor) +
         log(materials), apples)
# constrained linear model imposing constant
```

```
# return to scales
crs <- clm(cd, R = matrix(c(0, 1, 1, 1), nrow = 1),
               q = 1)
```

`cmtest` *Conditional moments test*

Description

Conditional moments tests for maximum likelihood estimators, particularly convenient for the probit and the tobit model to test relevance of functional form, omitted variables, heteroscedasticity and normality.

Usage

```
cmtest(
  x,
  test = c("normality", "reset", "heterosc", "skewness", "kurtosis"),
  powers = 2:3,
  heter_cov = NULL,
  opg = FALSE
)

## S3 method for class 'tobit'
cmtest(
  x,
  test = c("normality", "reset", "heterosc", "skewness", "kurtosis"),
  powers = 2:3,
  heter_cov = NULL,
  opg = FALSE
)

## S3 method for class 'micsr'
cmtest(
  x,
  test = c("normality", "reset", "heterosc", "skewness", "kurtosis"),
  powers = 2:3,
  heter_cov = NULL,
  opg = FALSE
)

## S3 method for class 'censReg'
cmtest(
  x,
  test = c("normality", "reset", "heterosc", "skewness", "kurtosis"),
  powers = 2:3,
  heter_cov = NULL,
```

```
  opg = FALSE
)

## S3 method for class 'glm'
cmtest(
  x,
  test = c("normality", "reset", "heterosc", "skewness", "kurtosis"),
  powers = 2:3,
  heter_cov = NULL,
  opg = FALSE
)
```

Arguments

`x`	a fitted model, currently a tobit model either fitted by `AER::tobit`, `censReg::censReg` or `micsr::tobit1` or a probit model fitted by `glm` with `family = binomial(link = "probit")` or by `micsr::binomreg` with `link = "probit"`
`test`	the kind of test to be performed, either a normality test (or separately a test that the skewness or kurtosis are 0 and 3), a heteroscedasticity test or a reset test,
`powers`	the powers of the fitted values that should be used in the reset test,
`heter_cov`	a one side formula that indicates the covariates that should be used for the heteroscedasticity test (by default all the covariates used in the regression are used),
`opg`	a boolean, if `FALSE` (the default), the analytic derivatives are used, otherwise the outer product of the gradient formula is used

Value

an object of class `"htest"` containing the following components:

- data.mane: a character string describing the fitted model
- statistic: the value of the test statistic
- parameter: degrees of freedom
- p.value: the p.value of the test
- method: a character indicating what type of test is performed

Author(s)

Yves Croissant

References

Newey WK (1985). "Maximum Likelihood Specification Testing and Conditional Moment Tests." *Econometrica*, **53**(5), 1047–1070.

Pagan A, Vella F (1989). "Diagnostic Tests for Models Based on Individual Data: A Survey." *Journal of Applied Econometrics*, **4**, S29–S59.

Tauchen G (1985). "Diagnostic testing and evaluation of maximum likelihood models." *Journal of Econometrics*, **30**(1), 415-443.

Wells C (2003). "Retesting Fair's (1978) Model on Infidelity." *Journal of Applied Econometrics*, **18**(2), 237–239.

Examples

```
charitable$logdon <- with(charitable, log(donation) - log(25))
ml <- tobit1(logdon ~ log(donparents) + log(income) + education +
             religion + married + south, data = charitable)
cmtest(ml, test = "heterosc")
cmtest(ml, test = "normality", opg = TRUE)
```

`drinks` *Physician advice on alcohol consumption*

Description

a cross-section of 2467 individuals from 1990

Format

a tibble containing:

- drinks: number of drinks in the past 2 weeks
- advice: 1 if reveived a drining advice
- age: age in 10 years cathegories
- race: a factor with levels `"white"`, `"black"` and `"other"`
- marital: marital status, one of `"single"`, `"married"`, `"widow"`, `"separated"`
- region: one of `"west"`, `"northeast"`, `"midwest"` and `"south"`
- empstatus: one of `"other"`, `"emp"` and `"unemp"`
- limits: limits on daily activities, one of `"none"`, `"some"` and `"major"`
- income: monthly income ($1000)
- educ: education in years
- medicare: insurance through medicare
- medicaid: insurance through medicaid
- champus: military insurance
- hlthins: health insurance
- regmed: regoular source of care
- dri: see same doctor
- diabete: have diabetes
- hearthcond: have heart condition
- stroke: have stroke

Source

JAE data archive

References

Kenkel DS, Terza JV (2001). "The effect of physician advice on alcohol consumption: count regression with an endogenous treatment effect." *Journal of Applied Econometrics*, **16**(2), 165-184.

`dummy` *Transform a factor in a set of dummy variables*

Description

The normal way to store cathegorical variables in R is to use factors, each modality being a level of this factor. Sometimes however, is is more convenient to use a set of dummy variables.

Usage

```
dummy(x, ..., keep = FALSE, prefix = NULL, ref = FALSE)
```

Arguments

`x`	a data frame
`...`	series of the data frame, should be factors
`keep`	a boolean, if `TRUE`, the original series is kept in the data frame,
`prefix`	an optional prefix for the names of the computed dummies,
`ref`	a boolean, if `TRUE`, a dummy is created for all the levels, including the reference level

Value

a data frame

Examples

```
charitable %>% dummy(religion, education)
```

escount *Endogenous switching and sample selection models for count data*

Description

Heckman's like estimator for count data, using either maximum likelihood or a two-steps estimator

Usage

```
escount(
  formula,
  data,
  subset,
  weights,
  na.action,
  offset,
  start = NULL,
  R = 16,
  hessian = FALSE,
  method = c("twosteps", "ml"),
  model = c("es", "ss")
)
```

Arguments

`formula`	a `Formula` object which includes two responses (the count and the binomial variables) and two sets of covariates (for the count component and for the selection equation)
`data`	a data frame,
`subset, weights, na.action, offset`	see `stats::lm`
`start`	an optional vector of starting values,
`R`	the number of points for the Gauss-Hermite quadrature
`hessian`	if `TRUE`, the numerical hessian is computed, otherwise the covariance matrix of the coefficients is computed using the outer product of the gradient
`method`	one of `'ML'` for maximum likelihood estimation (the default) or `'twosteps'` for the two-steps NLS method
`model`	one of `'es'` for endogenous switching (the default) or `'ss'` for sample selection

Value

an object of class `c("escount,micsr)"`, see `micsr::micsr` for further details.

Author(s)

Yves Croissant

References

Terza JV (1998). “Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects.” *Journal of Econometrics*, **84**(1), 129-154.

Greene WH (2001). “Fiml Estimation of Sample Selection Models for Count Data.” In Negishi T, Ramachandran RV, Mino K (eds.), *Economic Theory, Dynamics and Markets: Essays in Honor of Ryuzo Sato*, chapter 6, 73–91. Springer US, Boston, MA.

Examples

```
trips_2s <- escount(trips | car ~ workschl + size + dist + smsa + fulltime + distnod +
realinc + weekend + car | . - car - weekend + adults, data = trips, method = "twosteps")
trips_ml <- update(trips_2s, method = "ml")
```

`expreg` *Instrumental variable estimation for exponential conditional mean models*

Description

Exponential conditional mean models are particularly useful for non-negative responses (including count data). Least squares and one or two steps IV estimators are available

Usage

```
expreg(
  formula,
  data,
  subset,
  weights,
  na.action,
  offset,
  method = c("iv", "gmm", "ls"),
  error = c("mult", "add"),
  ...
)
```

Arguments

`formula`	a two-part right hand side formula, the first part describing the covariates and the second part the instruments
`data`	a data frame,
`subset, weights, na.action, offset`	see `stats::lm`
`method`	one of `"gmm"` (the default), `"iv"` or `ls`.
`error`	one of `"mult"` (the default) or `"add"` in order to get a model with respectively a multiplicative or an additive error
`...`	further arguments

Value

an object of class "micsr", see micsr::micsr for further details.

Author(s)

Yves Croissant

References

Mullahy J (1997). "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior." *The Review of Economics and Statistics*, **79**(4), 586-593.

Examples

```
cigmales <- dplyr::mutate(cigmales,
                          age2 = age ^ 2, educ2 = educ ^ 2, educage = educ * age,
                          age3 = age ^ 3, educ3 = educ ^ 3)
expreg(cigarettes ~ habit + price + restaurant + income + age + age2 + educ + educ2 +
                 famsize + race | . - habit + reslgth + lagprice + age3 + educ3 + educage,
                    data = cigmales)
expreg(birthwt ~ cigarettes + parity + race + sex | parity + race + sex +
                 edmother + edfather + faminc + cigtax, data = birthwt)
```

---

federiv *Foreign exchange derivatives use by large US bank holding companies*

Description

a cross-section of 794 banks from 1996 to 2000

Format

a tibble containing:

- federiv: foreign exchange derivatives use, a dummy
- optval: option awards
- eqrat: leverage
- bonus: bonus
- ltass: logarithm of total assets
- linsown: logarithm of the percentage of the total shares outstanding that are owned by officers and directors
- linstown: logarithm of the percentage of the total shares outstanding that are owned by all institutional investors
- roe: return on equity
- mktbk: market to book ratio

- perfor: foreign to total interest income ratio
- dealdum: derivative dealer activity dummy
- div: dividends paid
- year: year, from 1996 to 2000
- no_emp: number of employees
- no_subs: number of subsidiaries
- no_off: number of offices
- ceo_age: CEO age
- gap: 12 month maturity mismatch
- cfa: ratio of cash flow to total assets

Source

Lee Adkin's home page https://learneconometrics.com/

References

Adkins LC (2012). "Testing parameter significance in instrumental variables probit estimators: some simulation." *Journal of Statistical Computation and Simulation*, **82**(10), 1415-1436.

Adkins LC, Carter DA, Simpson WG (2007). "Managerial Incentives And The Use Of Foreign-Exchange Derivatives By Banks." *Journal of Financial Research*, **30**(3), 399-413.

---

fin_reform *Political economy of financial reforms*

Description

a pseudo-panel of 35 countries from 1973 to 1996

Format

a tibble containing:

- country: the country id
- year: the year
- region: the region
- pol: political orientation of the government
- fli: degree of policy liberalization index (from 0 to 18)
- yofc: year of office
- gdpg: growth rate of the gdp
- infl: inflation rate
- bop: balance of payments crises

- bank: banking crises
- imf: IMF program dummy
- usint: international interest rates
- open: trade openess
- dindx: difference of the inflation rate
- indx: inflation rate divided by 18
- indxl: lag value of indx
- rhs1: indxl * (1 - indxl)
- max_indxl: maximumum value of indxl by year and region
- catchup: difference between max_indxl and indxl
- dum_bop: balance of paiement crisis in the first two previous years
- dum_bank: bank crises in the first two previous years
- dum_1yofc: dummy for first year of office
- recession: dummy for recessions
- hinfl: dummy for inflation rate greater than 50 percent

Source

AEA website

References

Abiad A, Mody A (2005). "Financial Reform: What Shakes It? What Shapes It?" *American Economic Review*, **95**(1), 66-88.

`ftest` *F statistic*

Description

Extract the F statistic that all the parameters except the intercept are zero. Currently implemented only for models fitted by `lm` or `ivreg::ivreg`.

Usage

```
ftest(x, ...)

## S3 method for class 'lm'
ftest(x, ...)

## S3 method for class 'ivreg'
ftest(x, ..., covariate = NULL)
```

Arguments

`x`	a fitted object
`...`	further arguments
`covariate`	the covariate for which the test should be performed for the `ivreg` method

Value

an object of class `"htest"`.

`gaze` — *Short print of the summary of an object*

Description

`print` and `print.summary` methods often returns long input, which is suitable for the console, but too verbal for a printed output like a book or an article written using quarto. `gaze` is a generic function which prints a short output

Usage

```
gaze(x, ...)

## S3 method for class 'lm'
gaze(
  x,
  ...,
  coef = NULL,
  digits = max(3L, getOption("digits") - 3L),
  signif.stars = FALSE
)

## S3 method for class 'micsr'
gaze(
  x,
  ...,
  coef = NULL,
  digits = max(3L, getOption("digits") - 3L),
  signif.stars = FALSE
)

## S3 method for class 'ivreg'
gaze(
  x,
  ...,
  coef,
  digits = max(3L, getOption("digits") - 3L),
```

```
  signif.stars = getOption("show.signif.stars")
)

## S3 method for class 'rdrobust'
gaze(x, ..., first_stage = FALSE)

## S3 method for class 'CJMrddensity'
gaze(x, ...)

## S3 method for class 'htest'
gaze(x, ..., digits = 3)

## S3 method for class 'anova'
gaze(x, ..., digits = 3)

## S3 method for class 'LMtestlist'
gaze(x, ..., digits = 3)
```

Arguments

`x`	an object,
`...`	further arguments for the different methods,
`coef`	the coefficients to be printed
`digits`	the number of digits for the `lm` and the `ivreg` methods
`signif.stars`	a boolean indicating whether the stars should be printed
`first_stage`	a boolean for the `rdrobust::rdrobust` method, if `TRUE` the results of the first stage estimation are printed

Value

returns invisibly its first argument

Examples

```
t.test(extra ~ group, sleep) %>% gaze
lm(dist ~ poly(speed, 2), cars) %>% gaze
lm(dist ~ poly(speed, 2), cars) %>% gaze(coef = "poly(speed, 2)2")
```

---

hausman *Hausman test*

Description

Hausman test; under the null both models are consistent but one of them is more efficient, under the alternative, only one model is consistent

Usage

```
hausman(x, y, omit = FALSE, ...)

## S3 method for class 'ivreg'
hausman(x, y, omit = FALSE, ...)

## S3 method for class 'micsr'
hausman(x, y, omit = NULL, ...)
```

Arguments

x	the first model,
y	the second model
omit	a character containing the effects that are removed from the test
...	further arguments

Value

an object of class "htest".

Author(s)

Yves Croissant

References

Hausman JA (1978). “Specification Tests in Econometrics.” *Econometrica*, **46**(6), 1251–1271.

---

housprod *Household Production*

Description

a cross-section of 819 households from 1984

Format

a tibble containing:

- mjob: dummy, 1 if male has paid job
- fjob: dummy, 1 if female has paid job
- mtime: home production time male (minutes per day)
- ftime: home production time female (minutes per day)
- mwage: net hourly wage rate male (estimate imputed if mjob=0)
- fwage: net hourly wage rate female (estimate imputed if fjob=0)

- mage: age male
- meduc: years of schooling male
- fage: age female
- feduc: years of schooling female
- owner: dummy, 1 if houseownwers
- fsize: family size
- ychild: number of children younger than 7 years old in the household
- cars: number of cars in the household
- nonlabinc: non-labour income (in units of 1000 Swedish Kronor)

Source

JAE data archive

References

Kerkhofs M, Kooreman P (2003). "Identification and Estimation of a Class of Household Production Models." *Journal of Applied Econometrics*, **18**(3), 337–369.

`ivldv` *Instrumental variable estimators for limited dependent variable*

Description

Estimation of simultaneous-equation models when the response is binomial or censored

Usage

```
ivldv(
  formula,
  data,
  subset = NULL,
  weights = NULL,
  na.action,
  offset,
  method = c("twosteps", "minchisq", "ml", "test"),
  model = c("probit", "tobit"),
  robust = TRUE,
  left = 0,
  right = Inf,
  trace = 0,
  ...
)

endogtest(x, ...)
```

```
## S3 method for class 'formula'
endogtest(x, ..., data, model = c("probit", "tobit"))

## S3 method for class 'ivldv'
endogtest(x, ...)
```

Arguments

`formula`	a symbolic description of the model,
`data`	a data frame,
`subset, weights, na.action, offset`	see `lm`,
`method`	one of `"ml"` for maximum likelihood, "twosteps"and"minchisq"‘
`model`	one of `"probit"` or `"tobit"`,
`robust`	a boolean, if `TRUE`, a consistent estimation of the covariance of the coefficients is used for the 2-steps method,
`left, right`	left and right limits of the dependent variable. The default is respectively 0 and +Inf which corresponds to the most classic (left-zero truncated) tobit model,
`trace`	a boolean (the default if `FALSE`) if `TRUE` some information about the optimization process is printed,
`...`	further arguments
`x`	on object returned by `ivldv`

Value

An object of class `c('ivldv', 'lm')`

Author(s)

Yves Croissant

References

Smith R, Blundell R (1986). “An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply.” *Econometrica*, **54**(3), 679-85.

Rivers D, Vuong QH (1988). “Limited information estimators and exogeneity tests for simultaneous probit models.” *Journal of Econometrics*, **39**(3), 347-366.

Examples

```
inst <- ~ sic3 + k_serv + inv + engsci + whitecol + skill + semskill + cropland +
    pasture + forest + coal + petro + minerals + scrconc + bcrconc + scrcomp +
    bcrcomp + meps + kstock + puni + geog2 + tenure + klratio + bunion
trade_protection <- dplyr::mutate(micsr::trade_protection,
                                  y = ntb / (1 + ntb),
                                  x1 = vshipped / imports / elast,
```

```
                                 x2 = cap * x1,
                                 x3 = labvar)
GH <- ivldv(Formula::as.Formula(y  ~  x1 + x2, inst), trade_protection,
            method = "twosteps", model = "tobit")
Full <- ivldv(Formula::as.Formula(y ~ x1 + x2 + labvar, inst), trade_protection,
              method = "twosteps", model = "tobit")
Short <- ivldv(Formula::as.Formula(y ~ x1 + I(x2 + labvar), inst),
                 trade_protection, method = "twosteps", model = "tobit")
bank_msq <- ivldv(federiv ~ eqrat + optval + bonus + ltass + linsown + linstown +
                  roe + mktbk + perfor + dealdum + div + year | . - eqrat - bonus -
                  optval + no_emp + no_subs + no_off + ceo_age + gap + cfa,
                  data = federiv, method = "minchisq")
bank_ml <- update(bank_msq, method = "ml")
bank_2st <- update(bank_msq, method = "twosteps")
```

`loglm` *Log-linear model*

Description

Estimation of log-linear model; the estimation is done by `lm`, but the correct log-likelihood related quantities are returned

Usage

```
loglm(formula, data)
```

Arguments

`formula, data`	see `lm`

Value

An object of class `"micsr"`, see `micsr::micsr` for further details.

Author(s)

Yves Croissant

Examples

```
lm_model <- lm(log(dist) ~ log(speed), cars)
log_model <- loglm(dist ~ log(speed), cars)
coef(lm_model)
coef(log_model)
# same coefficients, supplementary sigma coefficient for `loglm`
logLik(lm_model)
logLik(log_model)
# log_model returns the correct value for the log-likelihood
```

micsr	micsr *class*

Description

The micsr class is intend to deal with a lot of different models that are estimated in the micsr package. More specifically, some models may be estimated using different estimation methods, like maximum likelihood, GMM or two-steps estimators. Objects of class micsr have an est_method item which is used by the different methods in order to have a relevent behaviour for the different methods.

Usage

```
llobs(x, ...)

## S3 method for class 'micsr'
coef(object, ..., subset = NA)

## S3 method for class 'micsr'
vcov(object, ..., vcov = c("info", "hessian", "opg"), subset = NA)

## S3 method for class 'micsr'
summary(object, ..., vcov = c("hessian", "info", "opg"), subset = NA)

## S3 method for class 'summary.micsr'
coef(object, ...)

## S3 method for class 'micsr'
print(x, digits = max(3L, getOption("digits") - 3L), ...)

## S3 method for class 'summary.micsr'
print(
  x,
  digits = max(3, getOption("digits") - 2),
  width = getOption("width"),
  ...
)

## S3 method for class 'micsr'
logLik(object, ..., type = c("model", "null", "saturated"))

## S3 method for class 'micsr'
BIC(object, ..., type = c("model", "null"))

## S3 method for class 'micsr'
AIC(object, ..., k = 2, type = c("model", "null"))
```

```
## S3 method for class 'micsr'
deviance(object, ..., type = c("model", "null"))

## S3 method for class 'micsr'
predict(object, ..., newdata = NULL)

## S3 method for class 'micsr'
model.part(object, ..., lhs = 1)

## S3 method for class 'micsr'
model.matrix(object, formula = NULL, ..., rhs = 1)

## S3 method for class 'micsr'
estfun(x, ...)

## S3 method for class 'micsr'
vcovHC(x, type, omega = NULL, sandwich = TRUE, ...)

## S3 method for class 'micsr'
bread(x, ...)

## S3 method for class 'micsr'
nobs(object, ...)

## S3 method for class 'micsr'
llobs(x, ...)

## S3 method for class 'mlogit'
llobs(x, ...)

## S3 method for class 'micsr'
tidy(x, conf.int = FALSE, conf.level = 0.95, ...)

## S3 method for class 'micsr'
glance(x, ...)
```

Arguments

`x, object`	an object which inherits the `micsr` class
`...`	further arguments
`subset`	a character which indicates which subset of coefficients should be extracted: one of `noinst` (all the coefficients except those corresponding to instrumental variables), `all`, `covar` (only the coefficients of the covariates), `inst` (only the coefficients of the instrumental variables) and `misc` (ony the "miscelanous" coefficients, typicaly a standard deviation or a coefficient of correlation)
`vcov`	the method used to compute the covariance matrix of the estimators (only for the ML estimator), one of `hessian` (the opposite of the inverse of the hessian),

	info (the inverse of the opposite of the expected value of the hessian), opg (the outer product of the gradient)
digits, width	see print
type, omega, sandwich	see sandwich::sandwich
k	see AIC
newdata	a new data frame to compute the predictions
lhs, rhs	see Formula::model.frame.Formula
formula	a formula
conf.int, conf.level	see broom:tidy.lm

Value

Objects of class micsr share a lot of common elements with lm: coefficients, residuals, fitted.values, model, terms, df.residual, xlevels, na.action, and call. npar is a named vector containing the index of subset of coefficients, it is used to print a subset of the results. It also has a est_method element and, depending of its value, contains further elements. In particular, for model fitted by maximum likelihood, value contains the individual contribution to the log-likelihood function, gradient the individual contribution to the gradient, hessian the hessian and information the information matrix. logLik contains the log-likelihood values of the proposed, null and saturated models. tests contains the values of the test that all the coefficients of the covariates are 0, using the three classical tests.

The llobs function is provided as a generic to extract the individual contributions to the log-likelihood

Specific methods have been writen for micsr objects: nobs, generics::tidy, generics::glance, sandwich::meat, sandwich::estfun, predict, model.matrix, Formula::model.part.

logLik, BIC, AIC and deviance methods have a type argument to select theproposed, null or saturated model.

vcov and summary methods have a vcov argument to select the estimator of the covariance matrix, which can be either based on the hessian, the gradient or the information.

vcov, summary and coef have a subset argument to select only a subset of the coefficients

---

mills *Compute the inverse Mills ratio and its first two derivatives*

Description

The inverse Mills ratio is used in several econometric models, especially different flavours of tobit model.

Usage

```
mills(x, deriv = 0)
```

Arguments

`x`	a numeric
`deriv`	one of 0 (the default, returns the inverse Mills ratio), 1 (the first derivative) and 2 (the second derivative)

Value

a numeric.

`mode_choice` *Choice between car and transit*

Description

a cross-section of 842 individuals

Format

a tibble containing:

- mode: 1 for car, 0 for transit
- cost: transit fare minus automobile travel cost in US$
- ivtime: transit in-vehicule travel time minus in-vehicule travel time (minutes)
- ovtime: transit out-of vehicule time minus out-of vehicule travel time (minutes)
- cars: number of cars owned by the traveler's household

Source

GAMS's website https://www.gams.com/latest/gamslib_ml/libhtml/gamslib_mws.html

References

Horowitz JL (1993). "Semiparametric estimation of a work-trip mode choice model." *Journal of econometrics*, **58**(1-2), 49-70.

ndvuong *Non-degenerate Vuong test*

Description

An unhanced version of the Vuong test with a small-sample bias correction

Usage

```
ndvuong(
  x,
  y,
  size = 0.05,
  pval = TRUE,
  nested = FALSE,
  vartest = FALSE,
  ndraws = 10000,
  diffnorm = 0.1,
  seed = 1,
  numbers = NULL,
  nd = TRUE,
  print.level = 0
)
```

Arguments

`x`	a first fitted model
`y`	a second fitted model
`size`	the size of the test
`pval`	should the p-value be computed ?
`nested`	a boolean, `TRUE` for nested models
`vartest`	a boolean, if `TRUE`, the variance test is computed
`ndraws`	the number of draws for the simulations
`diffnorm`	a creuser
`seed`	the seed
`numbers`	a user provided matrix of random numbers
`nd`	a boolean, if `TRUE` (the default) the non-degenarate Vuong test is computed
`print.level`	the level of details to be printed

Value

an object of class `"htest"`.

References

Vuong QH (1989). "Likelihood Ratio Tests for Selection and Non-Nested Hypotheses." *Econometrica*, **57**(2), 397-333.

Shi X (2015). "A nondegenerate Vuong test." *Quantitative Economics*, 85-121.

See Also

the classical Vuong test is implemented in `pscl::vuong` and `nonnest2::vuongtest`.

`newton` *Newton-Raphson method for numerical optimization*

Description

The Newton-Raphson method use the gradient and the hessian of a function. For well behaved functions, it is extremely accurate.

Usage

```
newton(
  fun,
  coefs,
  trace = 0,
  direction = c("min", "max"),
  tol = sqrt(.Machine$double.eps),
  maxit = 500,
  ...
)
```

Arguments

`fun`	the function to optimize
`coefs`	a vector of starting values
`trace`	if positive or true, some information about the computation is printed
`direction`	either `"min"` or `"max"`
`tol`	the tolerance
`maxit`	maximum number of iterations
`...`	further arguments, passed to fun

Value

a numeric vector, the parameters at the optimum of the function.

npar — *Number of parameters of a fitted model*

Description

The number of observation of a fitted model is typically obtained using the `nobs` method. There is no such generics to extract the same information about the number of parameters. `npar` is such a generic and has a special method for `micsr` objects with a `subset` argument that enables to compute the number of parameters for a subset of coefficients. The default method returns the length of the vector of coefficients extracted using the `coef` function.

Usage

```
npar(x, subset = NULL)

## Default S3 method:
npar(x, subset = NULL)

## S3 method for class 'micsr'
npar(x, subset = NULL)
```

Arguments

`x`	a fitted model
`subset`	a character indicating the subset of coefficients (only relevant for `micsr` models).

Value

an integer.

Author(s)

Yves Croissant

ordreg — *Ordered regression*

Description

Maximum-likelihood estimation of a model for which the response is ordinal

Usage

```
ordreg(
  formula,
  data,
  weights,
  subset,
  na.action,
  offset,
  link = c("probit", "logit"),
  start = NULL,
  ...
)
```

Arguments

`formula`	a symbolic description of the model
`data`	a data frame
`subset, weights, na.action, offset`	see `lm`
`link`	one of `probit` and `logit`
`start`	a vector of starting values, in this case, no estimation
`...`	further arguments

Value

an object of class `micsr`, see `micsr::micsr` for further details.

Examples

```
mod1 <- ordreg(factor(dindx) ~ rhs1 + catchup, fin_reform, link = "logit")
```

---

poisreg *Poisson regression*

Description

A unified interface to perform Poisson, Negbin and log-normal Poisson models

Usage

```
poisreg(
  formula,
  data,
  weights,
  subset,
  na.action,
```

```
    offset,
    start = NULL,
    mixing = c("none", "gamma", "lognorm"),
    method = c("bfgs", "newton"),
    vlink = c("nb1", "nb2"),
    ...
)
```

Arguments

`formula`	a symbolic description of the model, (for the count component and for the selection equation)
`data`	a data frame
`subset, weights, na.action, offset`	see `stats::lm`,
`start`	a vector of starting values
`mixing`	the mixing distribution, one of `"none"`, `"gamma"` and `"lognorm"`
`method`	the optimization method, one of `"newton"` and `"bfgs"`
`vlink`	one of `"nb1"` and `"nb2"`
`...`	further arguments

Value

an object of class `c("poisreg", "micsr")`, see `micsr::micsr` for further details.

Examples

```
nb1 <- poisreg(trips ~ workschl + size + dist + smsa + fulltime + distnod +
              realinc + weekend + car, trips, mixing = "gamma", vlink = "nb1")
```

`pscore` *Propensity scores*

Description

Propensity scores estimation, using an algorithm that checks the balancing hypothesis using strata and enable the estimation of the treatment effect using stratification methods

Usage

```
pscore(formula, data, maxiter = 4, tol = 0.005, link = c("logit", "probit"))

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

## S3 method for class 'pscore'
```

```
print(
  x,
  ...,
  digits = getOption("digits"),
  var_equal = c("none", "strata", "group", "both")
)

## S3 method for class 'summary.pscore'
print(
  x,
  ...,
  digits = getOption("digits"),
  format = c("rst", "pipe", "simple"),
  step = c("all", "strata", "covariates", "atet")
)

## S3 method for class 'pscore'
nobs(object, ..., smpl = c("total", "cs"))

## S3 method for class 'summary.pscore'
nobs(object, ..., smpl = c("total", "cs"))

rg(object, ...)

## S3 method for class 'pscore'
rg(object, ..., smpl = c("total", "cs"))

## S3 method for class 'summary.pscore'
rg(object, ..., smpl = c("total", "cs"))

stdev(object, ...)

## S3 method for class 'pscore'
mean(x, ..., var_equal = c("none", "strat", "group", "both"))

## S3 method for class 'summary.pscore'
mean(x, ...)

## S3 method for class 'pscore'
stdev(object, ..., var_equal = c("none", "strata", "group", "both"))

## S3 method for class 'summary.pscore'
stdev(object, ..., var_equal = c("none", "strata", "group", "both"))
```

Arguments

`formula` a Formula object; the left-hand side should contain two variables (`x1 + x2`), where x1 is the group variable and x2 the outcome. The group variable can be either a dummy for treated individuals or a factor with levels `"treated"` and

	"control"
data	a data frame
maxiter	the maximum number of iterations
tol	stratas are cut in halves as long as the hypothesis of equal means is rejected at the tol level,
link	the link for the binomial glm estimation, either "logit" or "probit"
...	further arguments
x, object	a "pscore" or a "summary.pscore" object
digits	number of digits for the print methods
var_equal	to compute the variance of the ATET, variances can be computed at the class/group level (var_equal = "none"), at the class level (var_equal = "group"), at the group level (var_equal = "strata") or globally (var_equal = "both")
format	one of "rst" "pipe" and "simple", this argument is passed to knitr::kable
step	for the print.summary method, the step of the test to be printed: one of "all" (the default), strata, covariates and atet
smpl	the sample to use, either the whole sample (smpl = "total") or the sample with common support (smpl = "cs")

Value

an object of class "pscore", with the following elements:

- strata: a tibble containing the stratas, the frequencies, the means and the variances of the propensity scores for treated and controled observations
- cov_balance: a tibble containing the results of the balancing tests for every covariate; the results for the class with the lowest p-value is reported
- unchecked_cov: a character vector containing the names of the covariates for which the balancing test could be computed
- model: a tibble containing the original data, with supplementary columns: gp_ for the groups, resp_ for the outcome and cls_ for the stratas
- pscore: the glm model fitted to compute the propensity scores

References

Dehejia RH, Wahba S (2002). “Propensity Score-Matching Methods for Nonexperimental Causal Studies.” *The Review of Economics and Statistics*, **84**(1), 151-161. ISSN 0034-6535, doi:10.1162/003465302317331982.

Becker SO, Ichino A (2002). “Estimation of average treatment effects based on propensity scores.” *Stata Journal*, **2**(4), 358-377(20).

Examples

```
data_tuscany <- dplyr::mutate(dplyr::filter(twa, region == "Tuscany"),
  dist2 = dist ^ 2, livselfemp = I((city == "livorno") * (occup == "selfemp")),
  perm = ifelse(outcome == "perm", 1, 0))
formula_tuscany <- group | perm ~ city + sex + marital + age +
   loc + children + educ + pvoto + training +
   empstat + occup + sector + wage + hour + feduc + femp + fbluecol +
   dist + dist2 + livselfemp
pscore(formula_tuscany, data_tuscany)
```

---

rsq *Coefficient of determination*

Description

A generic function to compute different flavors of coefficients of determination

Usage

```
rsq(x, type)

## S3 method for class 'lm'
rsq(x, type = c("raw", "adj"))

## S3 method for class 'micsr'
rsq(
  x,
 type = c("mcfadden", "cox_snell", "cragg_uhler", "aldrich_nelson", "veall_zimm",
    "estrella", "cor", "ess", "rss", "tjur", "mckel_zavo", "w", "lm", "lr")
)
```

Arguments

x	fitted model
type	the type of coefficient of determination

Value

a numeric scalar.

Examples

```
pbt <- binomreg(mode ~ cost + ivtime + ovtime, data = mode_choice, link = 'probit')
rsq(pbt)
rsq(pbt, "estrella")
rsq(pbt, "veall_zimm")
```

sargan *Sargan test for GMM models*

Description

When a IV model is over-identified, the set of all the empirical moment conditions can't be exactly 0. The test of the validity of the instruments is based on a quadratic form of the vector of the empirical moments

Usage

```
sargan(object, ...)

## S3 method for class 'ivreg'
sargan(object, ...)

## S3 method for class 'micsr'
sargan(object, ...)
```

Arguments

`object`	a model fitted by GMM
`...`	further arguments

Value

an object of class "htest".

Examples

```
cigmales <- cigmales %>%
       mutate(age2 = age ^ 2, educ2 = educ ^ 2,
              age3 = age ^ 3, educ3 = educ ^ 3,
              educage = educ * age)
gmm_cig <- expreg(cigarettes ~ habit + price + restaurant + income + age + age2 +
                 educ + educ2 + famsize + race | . - habit + age3 + educ3 +
                 educage + lagprice + reslgth, data = cigmales,
                 twosteps = FALSE)
sargan(gmm_cig)
```

`scoretest` *Score test*

Description

Score test, also knowned as Lagrange multiplier tests

Usage

```
scoretest(x, y, ...)

## S3 method for class 'micsr'
scoretest(x, y, ..., vcov = NULL)
```

Arguments

x	the first model,
y	the second model
...	further arguments
vcov	omit a character containing the effects that are removed from the test

Value

an object of class "htest".

Author(s)

Yves Croissant

Examples

```
mode_choice <- mode_choice %>%
   mutate(cost = cost * 8.42,
          gcost = (ivtime + ovtime) * 8 + cost)
pbt_unconst <- binomreg(mode ~ cost + ivtime + ovtime, data = mode_choice, link = "probit")
pbt_const <- binomreg(mode ~ gcost, data = mode_choice, link = "logit")
scoretest(pbt_const , . ~ . + ivtime + ovtime)
```

stder *Extract the standard errors of estimated coefficients*

Description

The standard errors are a key element while presenting the results of a model. They are the second column of the table of coefficient and are used to compute the t/z-value. `stderr` enables to retrieve easily the vector of standard errors, either from a fitted model or from a matrix of covariance

Usage

```
stder(x, .vcov, ...)

## Default S3 method:
stder(x, .vcov = NULL, ...)
```

Arguments

`x`	a fitted model or a matrix of covariance
`.vcov`	a function that computes a covariance matrix, or a character
`...`	further arguments

Value

a numeric vector

tobit1 *Truncated response model*

Description

Estimation of models for which the response is truncated, either on censored or truncated samples using OLS, NLS, maximum likelihood, two-steps estimators or trimmed estimators

Usage

```
tobit1(
  formula,
  data,
  subset = NULL,
  weights = NULL,
  start = NULL,
  left = 0,
  right = Inf,
  scedas = NULL,
```

```
  sample = c("censored", "truncated"),
  method = c("ml", "lm", "twosteps", "trimmed", "nls", "minchisq", "test"),
  trace = FALSE,
  ...
)
```

Arguments

`formula`	a symbolic description of the model; if two right hand sides are provided, the second one described the set of instruments if `scedas` is `NULL`, which is the default. Otherwise, the second part indicates the set of covariates for the variance function
`data, subset, weights`	see `lm`
`start`	an optional vector of starting values
`left, right`	left and right truncation points for the response The default is respectively 0 and +Inf which corresponds to the most classic (left-zero truncated) tobit model
`scedas`	the functional form used to specify the conditional variance, either `"exp"` or `"pnorm"`
`sample`	either `"censored"` (the default) to estimate the censored (tobit) regression model or `"truncated"` to estimated the truncated regression model
`method`	one of `"ml"` for maximum likelihood, `"lm"` for (biased) least squares estimators, `"twosteps"` for two-steps consistent estimators, `"trimmed"` for symetrically censored estimator, `"minchisq"` and `"test"`. The last two are only relevant for instrumental variable estimation (when the formula is a two-parts formula and `scedas` is `NULL`)
`trace`	a boolean (the default if `FALSE`) if `TRUE` some information about the optimization process is printed
`...`	further arguments

Value

An object of class `c("tobit1", "micsr")`, see `micsr::micsr` for further details.

Author(s)

Yves Croissant

References

Powell J (1986). "Symmetrically trimed least squares estimators for tobit models." *Econometrica*, **54**, 1435–1460.

Examples

```
charitable$logdon <- with(charitable, log(donation) - log(25))
ml <- tobit1(logdon ~ log(donparents) + log(income) + education +
             religion + married + south, data = charitable)
scls <- update(ml, method = "trimmed")
tr <- update(ml, sample = "truncated")
nls <- update(tr, method = "nls")
```

`trade_protection` *Lobying from Capitalists and Unions and Trade Protection*

Description

a cross-section of 194 United States

Format

a tibble containing:

- ntb: nontariff barrier coverage ratio
- vshipped: value of shipments
- imports: importations
- elast: demand elasticity
- cap: lobying
- labvar: labor market covariate
- sic3: 3-digit SIC industry classification
- k_serv: physical capital, factor share
- inv: Inventories, factor share
- engsci: engineers and scientists, factor share
- whitecol: white collar, factor share
- skill: skilled, factor share
- semskill: semi-skilled, factor share
- cropland: cropland, factor shaer
- pasture: pasture, factor share
- forest: forest, factor share
- coal: coal, factor share
- petro: petroleum, factor share
- minerals: minerals, factor share
- scrconc: seller concentration
- bcrconc: buyer concentration
- scrcomp: seller number of firms

- bcrcomp: buyer number of firms
- meps: scale
- kstock: capital stock
- puni: proportion of workers union
- geog2: geographic concentration
- tenure: average worker tenure, years
- klratio: capital-labor ratio
- bunion:

Source

American Economic Association Data Archive : https://www.aeaweb.org/aer/

References

Matschke X, Sherlund SM (2006). "Do Labor Issues Matter in the Determination of U.S. Trade Policy? An Empirical Reevaluation." *American Economic Review*, **96**(1), 405-421.

`trips` *Determinants of household trip taking*

Description

a cross-section of 577 households from 1978

Format

a tibble containing:

- trips: number of trips taken by a member of a household the day prior the survey interview
- car: 1 if household owns at least one motorized vehicule
- workschl: share of trips for work or school vs personal business or pleasure
- size: number of individuals in the household
- dist: distance to central business district in kilometers
- smsa: a factor with levels `"small"` (less than 2.5 million population) and `"large"` (more than 2.5 million population)
- fulltime: number of fulltime workers in household
- adults: number of adults in household
- distnod: distace from home to nearest transit node, in blocks
- realinc: household income divided by median income of census tract in which household resides
- weekend: 1 if the survey period is either saturday or sunday

Source

kindly provided by Joseph Terza

References

Terza JV (1998). "Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects." *Journal of Econometrics*, **84**(1), 129-154.

Terza JV, Wilson PW (1990). "Analyzing Frequencies of Several Types of Events: A Mixed Multinomial-Poisson Approach." *The Review of Economics and Statistics*, **72**(1), 108-115.

`turnout` *Turnout*

Description

these three models are replication in R of stata's code available on the web site of the American Economic Association. The estimation is complicated by the fact that some linear constraints are imposed.

Format

a list of three fitted models:

- group: the group-rule-utilitarian model
- intens: the intensity model
- sur: the reduced form SUR model

Details

Turnout in Texas liquor referenda

Source

American Economic Association data archive.

References

Coate S, Conlin M (2004). "A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence." *American Economic Review*, **94**(5), 1476-1504.

Examples

```
ndvuong(turnout$group, turnout$intens)
ndvuong(turnout$group, turnout$sur)
ndvuong(turnout$intens, turnout$sur)
```

twa *Temporary help jobs and permanent employment*

Description

a cross-section of 2030 individuals

Format

a tibble containing:

- id: identification code
- age: age
- sex: a factor with levels "female" and "male"
- marital: marital status, "married" or "single"
- children: number of children
- feduc: father's education
- fbluecol: father blue-color
- femp: father employed at time 1
- educ: years of education
- pvoto: mark in last degree as fraction of max mark
- training: received professional training before treatment
- dist: distance from nearest agency
- nyu: fraction of school-to-work without employment
- hour: weekly hours of work
- wage: monthly wage
- hwage: hourly wage at time 1
- contact: contacted a temporary work agency
- region: one of "Tuscany" and "Sicily"
- city: the city
- group: one of "control" and "treated"
- sector: the sector
- occup: occupation, one of "nojob", "selfemp", "bluecol" and "whitecol"
- empstat: employment status, one of "empl", "unemp" and "olf" (out of labor force)
- contract: job contract, one of "nojob", "atyp" (atypical) and "perm" (permanent)
- loc: localisation, one of "nord", "centro", "sud" and "estero"
- outcome: one of "none", "other", "fterm" and "perm"

Source

Journal of Applied Econometrics Data Archive : http://qed.econ.queensu.ca/jae/

References

Ichino A, Mealli F, Nannicini T (2008). "From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and Their Sensitivity?" *Journal of Applied Econometrics*, **23**(3), 305–327.

`vuong_sim` *Simulated pdfs for the Vuong statistics using linear models*

Description

This function can be used to reproduce the examples given by Shi (2015) which illustrate the fact that the distribution of the Vuong statistic may be very different from a standard normal

Usage

```
vuong_sim(N = 1000, R = 1000, Kf = 15, Kg = 1, a = 0.125)
```

Arguments

N	sample size
R	the number of replications
Kf	the number of covariates for the first model
Kg	the number of covariates for the second model
a	the share of the variance of y explained by the two competing models

Value

a numeric of length N containing the values of the Vuong statistic

References

Shi X (2015). "A nondegenerate Vuong test." *Quantitative Economics*, 85-121.

Examples

```
vuong_sim(N = 100, R = 10, Kf = 10, Kg = 2, a = 0.5)
```

zellner_revankar *Generalized production function*

Description

Log-likelihood function for the generalized production function of Zellner and Revankar (1969)

Usage

```
zellner_revankar(
  theta,
  y,
  Z,
  sum = FALSE,
  gradient = TRUE,
  hessian = TRUE,
  repar = TRUE
)
```

Arguments

theta	the vector of parameters
y	the vector of response
Z	the matrix of covariates
sum	if FALSE, a vector of individual contributions to the likelihood and the matrix of individual contributions to the gradient are returned, if TRUE a log-likelihood scalar and a gradient vector are returned
gradient	if TRUE, the gradient is returned as an attribute
hessian	if TRUE, the hessian is returned as an attrubute
repar	if TRUE, the likelihood is parametrized such that the constant return to scale hypothesis implies that two coefficients are 0

Value

a function.

Author(s)

Yves Croissant

References

Zellner A, Revankar NS (1969). "Generalized Production Functions." *Review of Economic Studies*, **36**(2), 241-250.

Index