Package 'SteinIV'

July 21, 2025

Version 0.1-1

Date 2016-01-01

Title Semi-Parametric Stein-Like Estimator with Instrumental Variables

Author Cedric E Ginestet <cedric.ginestet@kcl.ac.uk>

Maintainer Cedric E Ginestet <cedric.ginestet@kcl.ac.uk>

Depends R (>= 2.10.1)

Description Routines for computing different types of linear estimators, based on instrumental variables (IVs), including the semi-parametric Stein-like (SPS) estimator, originally introduced by Judge and Mittelhammer (2004) <DOI:10.1198/016214504000000430>.

License GPL (>= 2)

URL https://cran.r-project.org/package=SteinIV

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2016-01-26 15:19:33

Contents

`jive.est` *The Jackknife Instrumental Variable Estimator (JIVE).*

Description

Compute the JIVE for a multiple regression, as well as the set of standard errors for the individual vector entries, and the estimate of the asymptotic variance/covariance matrix.

Usage

```
jive.est(y,X,Z,SE=FALSE,n.bt=100)
```

Arguments

`y`	Numeric: A vector of observations, representing the outcome variable.
`X`	Numeric: A matrix of observations, whose number of columns corresponds to the number of predictors in the model, and the number of rows should be conformal with the number of entries in y. This matrix may contain both endogenous and exogenous variables.
`Z`	Numeric: A matrix of observations representing the intrumental variables (IVs) in the first-stage structural equation. The number of IVs should be at least as large as the number of endogenous variables in X.
`SE`	Logical: If TRUE, then the function also returns the standard errors of the individual JIVE estimators, and a bootstrap estimate of its asymptotic variance/covariance matrix.
`n.bt`	Numeric: The number of bootstrap samples performed for estimating the variance/covariance matrix. This automatically occurs, whenever the user selects the SE to be true.

Details

The JIVE was originally introduced by Angrist et al. (1995), in order to reduce the finite-sample bias of the TSLS estimator, when applied to a large number of instruments. Indeed, the TSLS estimator tends to behave poorly as the number of instruments increases. We briefly outline this method. See Angrist et al. (1999) for an exhaustive description.

The model is identical to the one used in the rest of this package. That is, the second-stage equation is modelled as $y = X\beta + \epsilon$, in which y is a vector of n observations representing the outcome variable, X is a matrix of order $n \times k$ denoting the predictors of the model, and comprised of both exogenous and endogenous variables, β is the k-dimensional vector of parameters of interest; whereas ϵ is an unknown vector of error terms. Moreover, the first-stage level of the model is given by a multivariate multiple regression. That is, this is a linear modle with a *multivariate* outcome variable, as well as *multiple* predictors. This first-stage model is represented in this manner, $X = Z\Gamma + \Delta$, where X is the matrix of predictors from the second-stage equation, Z is a matrix of instrumental variables (IVs) of order $n \times l$, Γ is a matrix of unknown parameters of order $l \times k$; whereas Δ denotes an unknown matrix of order $n \times k$ of error terms.

For computing the JIVE, we first consider the estimator of the regression parameter in the first-stage equation, which is denoted by

$$\hat{\Gamma} := (Z^T Z)^{-1}(Z^T X).$$

This matrix is of order $l \times k$. The matrix of predictors, X, projected onto the column space of the instruments is then given by $\hat{X} = Z\hat{\Gamma}$. The JIVE proceeds by estimating each row of $\hat{X}$ without using the corresponding data point. That is, the ith row in the jackknife matrix, $\hat{X}_J$, is estimated without using the ith row of X. This is conducted as follows. For every $i = 1, \ldots, n$, we first compute

$$\hat{\Gamma}_{(i)} := (Z_{(i)}^T Z_{(i)})^{-1}(Z_{(i)}^T X_{(i)}),$$

where $Z_{(i)}$ and $X_{(i)}$ denote matrices Z and X after removal of the ith row, such that these two matrices are of order $(n-1) \times l$ and $(n-1) \times k$, respectively. Then, the matrix $\hat{X}_J$ is constructed by stacking these jackknife estimates of $\hat{\Gamma}$, after they have been pre-multiplied by the corresponding rows of Z,

$$\hat{X}_J := (z_1\hat{\Gamma}_{(1)}, \ldots, z_n\hat{\Gamma}_{(n)})^T,$$

where each z_i is an l-dimensional row vector. The JIVE estimator is then obtained by replacing $\hat{X}$ with $\hat{X}_J$ in the standard formula of the TSLS, such that

$$\hat{\beta}_J := (\hat{X}_J{}^T X)^{-1}(\hat{X}_J{}^T y).$$

In this package, we have additionally made use of the computational formula suggested by Angrist et al. (1999), in which each row of $\hat{X}_J$ is calculated using

$$z_i\hat{\Gamma}_{(i)} = \frac{z_i\hat{\Gamma} - h_i x_i}{1 - h_i},$$

where $z_i\hat{\Gamma}_{(i)}$, $z_i\hat{\Gamma}$ and x_i are k-dimensional row vectors; and with h_i denoting the leverage of the corresponding data point in the first-level equation of our model, such that each h_i is defined as $z_i(Z^T Z)^{-1} z_i^T$.

Value

`list` A list with one or three arguments, depending on whether the user has activated the SE flag. The first element (est) in the list is the TSLS estimate of the model in vector format. The second element (se) is the vector of standard errors; and the third element (var) is the sample estimate of the asymptotic variance/covariance matrix.

Author(s)

Cedric E. Ginestet <cedric.ginestet@kcl.ac.uk>

References

Angrist, J., Imbens, G., and Krueger, A.B. (1995). Jackknife instrumental variables esti- mation. Technical Working Paper 172, National Bureau of Economic Research.

Angrist, J.D., Imbens, G.W., and Krueger, A.B. (1999). Jackknife instrumental variables estimation. Journal of Applied Econometrics, 14(1), 57–67.

Examples

```
### Generate a simple example with synthetic data, and no intercept.
n <- 100; k <- 3; l <- 3;
Ga<- diag(rep(1,l)); be <- rep(1,k);
Z <- matrix(0,n,l); for(j in 1:l) Z[,j] <- rnorm(n);
X <- matrix(0,n,k); for(j in 1:k) X[,j] <- Z[,j]*Ga[j,j] + rnorm(n);
y <- X%*%be + rnorm(n);

### Compute JIVE estimator with SEs and variance/covariance matrix.
print(jive.est(y,X,Z))
print(jive.est(y,X,Z,SE=TRUE));
```

`jive.internal` *Internal function for the Jackknife Instrumental Variable Estimator (JIVE).*

Description

Compute the JIVE for a multiple regression

Usage

```
jive.internal(y,X,Z)
```

Arguments

y	Numeric: A vector of observations, representing the outcome variable.
X	Numeric: A matrix of observations, whose number of columns corresponds to the number of predictors in the model, and the number of rows should be conformal with the number of entries in y. This matrix may contain both endogenous and exogenous variables.
Z	Numeric: A matrix of observations representing the intrumental variables (IVs) in the first-stage structural equation. The number of IVs should be at least as large as the number of endogenous variables in X.

Details

See documentaion for the jive.est function. Users should use the jive.est function, instead.

Value

B	A vector of estimates for the coefficients of interest.

Author(s)

Cedric E. Ginestet <cedric.ginestet@kcl.ac.uk>

References

Angrist, J., Imbens, G., and Krueger, A.B. (1995). Jackknife instrumental variables esti- mation. Technical Working Paper 172, National Bureau of Economic Research.

Angrist, J.D., Imbens, G.W., and Krueger, A.B. (1999). Jackknife instrumental variables estimation. Journal of Applied Econometrics, 14(1), 57–67.

`ols.est` *The Ordinary Least Squares (OLS) Estimator.*

Description

Compute the OLS estimator of a multiple regression, as well as the set of standard errors for the individual vector entries, and the estimate of the asymptotic variance/covariance matrix.

Usage

```
ols.est(y,X,SE=FALSE)
```

Arguments

`y`	Numeric: A vector of observations, representing the outcome variable.
`X`	Numeric: A matrix of observations, whose number of columns corresponds to the number of predictors in the model, and the number of rows should be conformal with the number of entries in y. This matrix may contain both endogenous and exogenous variables.
`SE`	Logical: If TRUE, then the function also returns the standard errors of the individual TSLS estimator, and a sample estimate of its asymptotic variance/covariance matrix.

Details

The OLS estimator is computed for a standard one-stage structural model. We here adopt the terminology commonly used in econometrics. See, for example, the references below for Cameron and Trivedi (2005), Davidson and MacKinnon (1993), as well as Wooldridge (2002). The second-stage equation is thus modelled as follows,

$$y = X\beta + \epsilon,$$

in which y is a vector of n observations representing the outcome variable, X is a matrix of order $n \times k$ denoting the predictors of the model, and comprised of both exogenous and endogenous variables, β is the k-dimensional vector of parameters of interest; whereas ϵ is an unknown vector of error terms. The formula for the OLS estimator is then obtained in the standard fashion by the following equation,

$$\hat{\beta}_{OLS} := (X^T X)^{-1}(X^T y),$$

with variance/covariance matrix given by

$$\hat{\Sigma}_{OLS} := \hat{\sigma}^2 (X^T X)^{-1},$$

in which the sample residual sum of squares is $\hat{\sigma}^2 := (y - X\hat{\beta}_{OLS})^T (y - X\hat{\beta}_{OLS})/(n-k)$.

Value

`list` A list with one or three arguments, depending on whether the user has activated the SE flag. The first element (est) in the list is the TSLS estimate of the model in vector format. The second element (se) is the vector of standard errors; and the third element (var) is the sample estimate of the asymptotic variance/covariance matrix.

Author(s)

Cedric E. Ginestet <cedric.ginestet@kcl.ac.uk>

References

Cameron, A. and Trivedi, P. (2005). Microeconometrics: Methods and Applications. Cam- bridge University press, Cambridge.

Davidson, R. and MacKinnon, J.G. (1993). Estimation and inference in econometrics. OUP Catalogue.

Wooldridge, J. (2002). Econometric analysis of cross-section and panel data. MIT press, London.

Examples

```
### Generate a simple example with synthetic data, and no intercept.
n <- 100; k <- 3; l <- 3;
Ga<- diag(rep(1,l)); be <- rep(1,k);
Z <- matrix(0,n,l); for(j in 1:l) Z[,j] <- rnorm(n);
X <- matrix(0,n,k); for(j in 1:k) X[,j] <- Z[,j]*Ga[j,j] + rnorm(n);
y <- X%*%be + rnorm(n);

### Compute OLS estimator with SEs and variance/covariance matrix.
print(ols.est(y,X))
print(ols.est(y,X,SE=TRUE))
```

---

sps.est *Semi-parametric Stein-like (SPS) estimator.*

Description

Computes the SPS estimator for a two-stage structural model, as well as the set of standard errors for each individual estimator, and the sample estimate of the asymptotic variance/covariance matrix.

Usage

```
sps.est(y,X,Z,SE=FALSE,ALPHA=TRUE,REF="TSLS",n.bt=100,n.btj=10)
```

Arguments

`y`	Numeric: A vector of observations, representing the outcome variable.
`X`	Numeric: A matrix of observations, whose number of columns corresponds to the number of predictors in the model, and the number of rows should be conformal with the number of entries in y. This matrix may contain both endogenous and exogenous variables.
`Z`	Numeric: A matrix of observations representing the intrumental variables (IVs) in the first-stage structural equation. The number of IVs should be at least as large as the number of endogenous variables in X.
`SE`	Logical: If TRUE, then the function also returns the standard errors of the individual SPS estimators, and a sample (or bootstrap, if JIVE is selected as a reference estimator) estimate of its asymptotic variance/covariance matrix.
`ALPHA`	Logical: If TRUE, the function returns the value of the sample estimate of the parameter controlling the respective contribution of the *reference* estimator (by default, this is the TSLS estimator), and the one of the *alternative* estimator (by default, this is the OLS estimator).
`REF`	Character: Controls the choice of the *reference* estimator in the SPS framework. This can accept two values: "TSLS" or "JIVE", with the former being the default option. The *alternative* estimator is always the OLS estimator.
`n.bt`	Numeric: The number of bootstrap samples performed, when the sample variance/covariance matrix is estimated using the bootstrap. This automatically occurs, whenever the user selects the JIVE as the reference estimator.
`n.btj`	Numeric: The number of boostrap iterations performed, when computing the SPS estimator, when using the JIVE as reference estimator. This option is only relevant, when JIVE has been selected as the reference estimator. These iterations are used to compute the various components entering in the calculation of the SPS estimator.

Details

The SPS estimator is applied to a two-stage structural model. We here adopt the terminology commonly used in econometrics. See, for example, the references below for Cameron and Trivedi (2005), Davidson and MacKinnon (1993), as well as Wooldridge (2002). The second-stage equation is thus modelled as follows,

$$y = X\beta + \epsilon,$$

in which y is a vector of n observations representing the outcome variable, X is a matrix of order $n \times k$ denoting the predictors of the model, and comprised of both exogenous and endogenous variables, β is the k-dimensional vector of parameters of interest; whereas ϵ is an unknown vector of error terms. The first-stage level of the model is given by a multivariate multiple regression. That is, this is a linear modle with a *multivariate* outcome variable, as well as *multiple* predictors. This first-stage model is represented in this manner,

$$X = Z\Gamma + \Delta,$$

where X is the matrix of predictors from the second-stage equation, Z is a matrix of instrumental variables (IVs) of order $n \times l$, Γ is a matrix of unknown parameters of order $l \times k$; whereas Δ denotes an unknown matrix of order $n \times k$ of error terms.

As for the TSLS estimator, whenever certain variables in X are assumed to be exogenous, these variables should be incorporated into Z. That is, all the exogneous variables are their own instruments. Moreover, it is also assumed that the model contains at least as many instruments as predictors, in the sense that $l \geq k$, as commonly donein practice (Wooldridge, 2002). Also, the matrices, $X^T X$, $Z^T X$, and $Z^T Z$ are all assumed to be full rank. Finally, both X and Z should comprise a column of one's, representing the intercept in each structural equation.

The formula for the SPS estimator is then obtained as a weigthed combination of the OLS and TSLS estimators (using the default options), such that

$$\hat{\beta}_{SPS}(\alpha) := \alpha\hat{\beta}_{OLS} + (1 - \alpha)\hat{\beta}_{TSLS},$$

for every α. The *proportion parameter*, α, controls the respective contributions of the OLS and TSLS estimators. (Despite our choice of name, however, note that α needs not be bounded between 0 and 1.) This parameter is selected in order to minimize the trace of the theoretical MSE of the corresponding SPS estimator,

$$MSE(\hat{\beta}_{SPS}(\alpha)) = E[(\bar{\beta}(\alpha) - \beta)(\hat{\beta}(\alpha) - \beta)^T] = Var(\hat{\beta}(\alpha)) + Bias^2(\hat{\beta}(\alpha)),$$

where $\beta \in R^k$ is the true parameter of interest and the MSE is a $k \times k$ matrix. It is particularly appealing to combine these two estimators, because the asymptotic unbiasedness of the TSLS estimator guarantees that the resulting SPS is asymptotically unbiased. Thus, the MSE automatically strikes a trade-off between the unbiasedness of the TSLS estimator and the efficiency of the OLS estimator.

Value

`list` A list with one or four arguments, depending on whether the user has activated the SE flag, and the ALPHA flag. The first element (est) in the list is the SPS estimate of the model in vector format. The second element (se) is the vector of standard errors; the third element (var) is the sample estimate of the asymptotic variance/covariance matrix; the fourth element (alpha) is a real number representing the estimate of the contribution of the OLS to the combined SPS estimator.

Author(s)

Cedric E. Ginestet <cedric.ginestet@kcl.ac.uk>

References

Judge, G.G. and Mittelhammer, R.C. (2004). A semiparametric basis for combining esti- mation problems under quadratic loss. Journal of the American Statistical Association, 99(466), 479–487.

Judge, G.G. and Mittelhammer, R.C. (2012a). An information theoretic approach to econo- metrics. Cambridge University Press.

Judge, G. and Mittelhammer, R. (2012b). A risk superior semiparametric estimator for overidentified linear models. Advances in Econometrics, 237–255.

Judge, G. and Mittelhammer, R. (2013). A minimum mean squared error semiparametric combining estimator. Advances in Econometrics, 55–85.

Mittelhammer, R.C. and Judge, G.G. (2005). Combining estimators to improve structural model estimation and inference under quadratic loss. Journal of econometrics, 128(1), 1–29.

Examples

```
### Generate a simple example with synthetic data, and no intercept.
n <- 100; k <- 3; l <- 3;
Ga<- diag(rep(1,l)); be <- rep(1,k);
Z <- matrix(0,n,l); for(j in 1:l) Z[,j] <- rnorm(n);
X <- matrix(0,n,k); for(j in 1:k) X[,j] <- Z[,j]*Ga[j,j] + rnorm(n);
y <- X%*%be + rnorm(n);

### Compute SPS estimator with SEs and variance/covariance matrix.
print(sps.est(y,X,Z))
print(sps.est(y,X,Z,SE=TRUE));
```

`sps.internal` *Internal function for Semi-parametric Stein-like (SPS) estimator.*

Description

Computes the SPS estimator for a two-stage structural model, as well as a sample estimate of the alpha parameter controlling the degree of combination between the OLS and TSLS estimators.

Usage

```
sps.internal(y,X,Z,REF="TSLS",ALPHA=FALSE,n.btj=10)
```

Arguments

`y`	Numeric: A vector of observations, representing the outcome variable.
`X`	Numeric: A matrix of observations, whose number of columns corresponds to the number of predictors in the model, and the number of rows should be conformal with the number of entries in y. This matrix may contain both endogenous and exogenous variables.
`Z`	Numeric: A matrix of observations representing the intrumental variables (IVs) in the first-stage structural equation. The number of IVs should be at least as large as the number of endogenous variables in X.
`REF`	Character: Controls the choice of the *reference* estimator in the SPS framework. This can accept two values: "TSLS" or "JIVE", with the former being the default option. The *alternative* estimator is always the OLS estimator.
`ALPHA`	Logical: If TRUE, the function returns the value of the sample estimate of the parameter controlling the respective contribution of the *reference* estimator (by default, this is the TSLS estimator), and the one of the *alternative* estimator (by default, this is the OLS estimator).
`n.btj`	Numeric: The number of boostrap iterations performed, when computing the SPS estimator, when using the JIVE as reference estimator. This option is only relevant, when JIVE has been selected as the reference estimator. These iterations are used to compute the various components entering in the calculation of the SPS estimator.

Details

See documentaion for the sps.est function. Users should use the sps.est function, instead.

Value

`list` The first term (est) is a vector of estimates for the coefficients of interest, and the second term (alpha) representing the estimate of the contribution of the OLS to the combined SPS estimator.

Author(s)

Cedric E. Ginestet <cedric.ginestet@kcl.ac.uk>

References

Judge, G.G. and Mittelhammer, R.C. (2004). A semiparametric basis for combining esti- mation problems under quadratic loss. Journal of the American Statistical Association, 99(466), 479–487.

Judge, G.G. and Mittelhammer, R.C. (2012a). An information theoretic approach to econo- metrics. Cambridge University Press.

Judge, G. and Mittelhammer, R. (2012b). A risk superior semiparametric estimator for over-identified linear models. Advances in Econometrics, 237–255.

Judge, G. and Mittelhammer, R. (2013). A minimum mean squared error semiparametric combining estimator. Advances in Econometrics, 55–85.

Mittelhammer, R.C. and Judge, G.G. (2005). Combining estimators to improve structural model estimation and inference under quadratic loss. Journal of econometrics, 128(1), 1–29.

`tr` *Trace of a matrix.*

Description

Compute the trace of a square matrix.

Usage

```
tr(X)
```

Arguments

`X` Numeric: A square matrix.

Details

This computes the sum of the diagonal elements of a square matrix.

Value

`numeric` A real number.

Author(s)

Cedric E. Ginestet <cedric.ginestet@kcl.ac.uk>

`tsls.est` *The Two-Stage Least Squares (TSLS) estimator.*

Description

Computes the TSLS estimator for a two-stage structural model, as well as the set of standard errors for each individual estimator, and the sample estimate of the asymptotic variance/covariance matrix.

Usage

```
tsls.est(y,X,Z,SE=FALSE)
```

Arguments

`y`	Numeric: A vector of observations, representing the outcome variable.
`X`	Numeric: A matrix of observations, whose number of columns corresponds to the number of predictors in the model, and the number of rows should be conformal with the number of entries in y. This matrix may contain both endogenous and exogenous variables.
`Z`	Numeric: A matrix of observations representing the intrumental variables (IVs) in the first-stage structural equation. The number of IVs should be at least as large as the number of endogenous variables in X.
`SE`	Logical: If TRUE, then the function also returns the standard errors of the individual TSLS estimator, and a sample estimate of its asymptotic variance/covariance matrix.

Details

The TSLS estimator is applied to a two-stage structural model. We here adopt the terminology commonly used in econometrics. See, for example, the references below for Cameron and Trivedi (2005), Davidson and MacKinnon (1993), as well as Wooldridge (2002). The second-stage equation is thus modelled as follows,

$$y = X\beta + \epsilon,$$

in which y is a vector of n observations representing the outcome variable, X is a matrix of order $n \times k$ denoting the predictors of the model, and comprised of both exogenous and endogenous variables, β is the k-dimensional vector of parameters of interest; whereas ϵ is an unknown vector of error terms.

The first-stage level of the model is given by a multivariate multiple regression. That is, this is a linear modle with a *multivariate* outcome variable, as well as *multiple* predictors. This first-stage model is represented in this manner,

$$X = Z\Gamma + \Delta$$

, where X is the matrix of predictors from the second-stage equation, Z is a matrix of instrumental variables (IVs) of order $n \times l$, Γ is a matrix of unknown parameters of order $l \times k$; whereas Δ denotes an unknown matrix of order $n \times k$ of error terms.

Whenever certain variables in X are assumed to be exogenous, these variables should be incorporated into Z. That is, all the exogneous variables are their own instruments. Moreover, it is also assumed that the model contains at least as many instruments as predictors, in the sense that $l \geq k$, as commonly donein practice (Wooldridge, 2002). Also, the matrices, $X^T X$, $Z^T X$, and $Z^T Z$ are all assumed to be full rank. Finally, both X and Z should comprise a column of one's, representing the intercept in each structural equation.

The formula for the TSLS estimator is then obtained in the standard fashion by the following equation,

$$\hat{\beta}_{TSLS} := (\hat{X}^T \hat{X})^{-1}(\hat{X}^T y),$$

where $\hat{X} := H_z X$, is the orthogonal projection of the matrix X, onto the vector space spanned by the columns of Z; and $H_z := Z(Z^T Z)^{-1} Z^T$ is the hat matrix of the first-stage multivariate regression.

When requested by the user, the standard errors of each entry in $\hat{\beta}_{TSLS}$ are also provided, as a vector. These are computed by taking the squareroot of the diagonal entries of the sample asymptotic variance/covariance matrix, which is given by the following equation,

$$\hat{\Sigma}_{TSLS} := \hat{\sigma}^2 (\hat{X}^T \hat{X})^{-1},$$

in which the sample residual sum of squares is $\hat{\sigma}^2 := (y - X\hat{\beta}_{TSLS})^T (y - X\hat{\beta}_{TSLS})/(n-k)$.

Value

`list` A list with one or three arguments, depending on whether the user has activated the SE flag. The first element (est) in the list is the TSLS estimate of the model in vector format. The second element (se) is the vector of standard errors; and the third element (var) is the sample estimate of the asymptotic variance/covariance matrix.

Author(s)

Cedric E. Ginestet <cedric.ginestet@kcl.ac.uk>

References

Cameron, A. and Trivedi, P. (2005). Microeconometrics: Methods and Applications. Cam- bridge University press, Cambridge.

Davidson, R. and MacKinnon, J.G. (1993). Estimation and inference in econometrics. OUP Catalogue.

Wooldridge, J. (2002). Econometric analysis of cross-section and panel data. MIT press, London.

Examples

```
### Generate a simple example with synthetic data, and no intercept.
n <- 100; k <- 3; l <- 3;
Ga<- diag(rep(1,l)); be <- rep(1,k);
Z <- matrix(0,n,l); for(j in 1:l) Z[,j] <- rnorm(n);
X <- matrix(0,n,k); for(j in 1:k) X[,j] <- Z[,j]*Ga[j,j] + rnorm(n);
y <- X%*%be + rnorm(n);

### Compute TSLS estimator with SEs and variance/covariance matrix.
print(tsls.est(y,X,Z));
print(tsls.est(y,X,Z,SE=TRUE));
```

Index