# Package 'OneSampleMR'

June 26, 2024

**Title** One Sample Mendelian Randomization and Instrumental Variable
Analyses

**Version** 0.1.5

**Description** Useful functions for one-sample (individual level data)
Mendelian randomization and instrumental variable analyses. The
package includes implementations of; the Sanderson and Windmeijer
(2016) <doi:10.1016/j.jeconom.2015.06.004> conditional F-statistic,
the multiplicative structural mean model Hernán and Robins (2006)
<doi:10.1097/01.ede.0000222409.00878.37>, and two-stage predictor
substitution and two-stage residual inclusion estimators explained by
Terza et al. (2008) <doi:10.1016/j.jhealeco.2007.09.009>.

**License** GPL (>= 3)

**URL** https://github.com/remlapmot/OneSampleMR,
https://remlapmot.github.io/OneSampleMR/

**BugReports** https://github.com/remlapmot/OneSampleMR/issues/

**Depends** R (>= 4.3)

**Imports** Formula, gmm, ivreg, lmtest, msm, rlang (>= 1.0.0)

**Suggests** haven, knitr, lfe, rmarkdown, testthat (>= 3.0.0)

**VignetteBuilder** knitr

**Config/testthat/edition** 3

**Encoding** UTF-8

**RoxygenNote** 7.3.1

**NeedsCompilation** no

**Author** Tom Palmer [aut, cre] (<https://orcid.org/0000-0003-4655-4511>),
Wes Spiller [aut] (<https://orcid.org/0000-0002-8169-5531>),
Eleanor Sanderson [aut] (<https://orcid.org/0000-0001-5188-5775>)

**Maintainer** Tom Palmer <remlapmot@hotmail.com>

**Repository** CRAN

**Date/Publication** 2024-06-26 18:50:02 UTC

# Contents

---

OneSampleMR-package        *OneSampleMR: One Sample Mendelian Randomization and Instrumental Variable Analyses*

---

### Description

Useful functions for one-sample (individual level data) Mendelian randomization and instrumental variable analyses. The package includes implementations of; the Sanderson and Windmeijer (2016) doi:10.1016/j.jeconom.2015.06.004 conditional F-statistic, the multiplicative structural mean model Hernán and Robins (2006) doi:10.1097/01.ede.0000222409.00878.37, and two-stage predictor substitution and two-stage residual inclusion estimators explained by Terza et al. (2008) doi:10.1016/j.jhealeco.2007.09.009.

### Author(s)

**Maintainer**: Tom Palmer <remlapmot@hotmail.com> (ORCID)

Authors:

- Wes Spiller <wes.spiller@bristol.ac.uk> (ORCID)

- Eleanor Sanderson <eleanor.sanderson@bristol.ac.uk> (ORCID)

### See Also

Useful links:

- https://github.com/remlapmot/OneSampleMR

- https://remlapmot.github.io/OneSampleMR/

- Report bugs at https://github.com/remlapmot/OneSampleMR/issues/

---

asmm | *Additive structural mean model*

---

## Description

`asmm` is not a function. This helpfile is to note that the additive structural mean model (ASMM) is simply fit with a linear IV estimator, such as available in `ivreg::ivreg()`.

## Details

For a binary outcome the ASMM estimates a causal risk difference.

## References

Clarke PS, Palmer TM, Windmeijer F. Estimating structural mean models with multiple instrumental variables using the Generalised Method of Moments. Statistical Science, 2015, 30, 1, 96-117. doi:10.1214/14STS503

Palmer TM, Sterne JAC, Harbord RM, Lawlor DA, Sheehan NA, Meng S, Granell R, Davey Smith G, Didelez V. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. American Journal of Epidemiology, 2011, 173, 12, 1392-1403. doi:10.1093/aje/kwr026

Robins JM. The analysis of randomised and nonrandomised AIDS treatment trials using a new approach to causal inference in longitudinal studies. In Health Service Research Methodology: A Focus on AIDS (L. Sechrest, H. Freeman and A. Mulley, eds.). 1989. 113–159. US Public Health Service, National Center for Health Services Research, Washington, DC.

## Examples

```
# Single instrument example
# Data generation from the example in the ivtools ivglm() helpfile
set.seed(9)
n    <- 1000
psi0 <- 0.5
Z    <- rbinom(n, 1, 0.5)
X    <- rbinom(n, 1, 0.7*Z + 0.2*(1 - Z))
m0   <- plogis(1 + 0.8*X - 0.39*Z)
Y    <- rbinom(n, 1, plogis(psi0*X + log(m0/(1 - m0))))
dat1 <- data.frame(Z, X, Y)
fit1 <- ivreg::ivreg(Y ~ X | Z, data = dat1)
summary(fit1)

# Multiple instrument example
set.seed(123456)
n    <- 1000
psi0 <- 0.5
G1   <- rbinom(n, 2, 0.5)
G2   <- rbinom(n, 2, 0.3)
G3   <- rbinom(n, 2, 0.4)
```

```
U     <- runif(n)
pX    <- plogis(0.7*G1 + G2 - G3 + U)
X     <- rbinom(n, 1, pX)
pY    <- plogis(-2 + psi0*X + U)
Y     <- rbinom(n, 1, pY)
dat2 <- data.frame(G1, G2, G3, X, Y)
fit2 <- ivreg::ivreg(Y ~ X | G1 + G2 + G3, data = dat2)
summary(fit2)
```

---

fsw *Conditional F-statistic of Sanderson and Windmeijer (2016)*

---

## Description

fsw calculates the conditional F-statistic of Sanderson and Windmeijer (2016) for each endogenous variable in the model.

## Usage

```
fsw(object)

## S3 method for class 'ivreg'
fsw(object)
```

## Arguments

object        An object of class "ivreg" containing the results of an IV model fitted by
              ivreg::ivreg() for which to calculate the conditional F-statistics for each en-
              dogenous variable.

## Value

An object of class "fsw" with the following elements:

**fswres** matrix with columns for the conditional *F*-statistics, degrees of freedom, residual degrees of freedom, and p-value. 1 row per endogenous variable.

**namesendog** a character vector of the variable names of the endogenous variables.

**nendog** the number of endogenous variables.

**n** the sample size used for the fitted model.

## References

Sanderson E and Windmeijer F. A weak instrument *F*-test in linear IV models with multiple endogenous variables. Journal of Econometrics, 2016, 190, 2, 212-221, doi:10.1016/j.jeconom.2015.06.004.

## Examples

```
require(ivreg)
set.seed(12345)
n   <- 4000
z1  <- rnorm(n)
z2  <- rnorm(n)
w1  <- rnorm(n)
w2  <- rnorm(n)
u   <- rnorm(n)
x1  <- z1 + z2 + 0.2*u + 0.1*w1 + rnorm(n)
x2  <- z1 + 0.94*z2 - 0.3*u + 0.1*w2 + rnorm(n)
y   <- x1 + x2 + w1 + w2 + u
dat <- data.frame(w1, w2, x1, x2, y, z1, z2)
mod <- ivreg::ivreg(y ~ x1 + x2 + w1 + w2 | z1 + z2 + w1 + w2, data = dat)
fsw(mod)
```

---

msmm                    *Multiplicative structural mean model*

---

## Description

Function providing several methods to estimate the multiplicative structural mean model (MSMM) of Robins (1989).

## Usage

```
msmm(
  formula,
  instruments,
  data,
  subset,
  na.action,
  contrasts = NULL,
  estmethod = c("gmm", "gmmalt", "tsls", "tslsalt"),
  t0 = NULL,
  ...
)
```

## Arguments

formula, instruments

formula specification(s) of the regression relationship and the instruments. Either instruments is missing and formula has three parts as in y ~ x1 + x2 | z1 + z2 + z3 (recommended) or formula is y ~ x1 + x2 and instruments is a one-sided formula ~ z1 + z2 + z3 (only for backward compatibility).

data            an optional data frame containing the variables in the model. By default the variables are taken from the environment of the formula.

| subset | an optional vector specifying a subset of observations to be used in fitting the model. |
|---|---|
| na.action | a function that indicates what should happen when the data contain NAs. The default is set by the na.action option. |
| contrasts | an optional list. See the contrasts.arg of stats::model.matrix(). |
| estmethod | Estimation method, please use one of |

- "gmm" GMM estimation of the MSMM (the default).
- "gmmalt" GMM estimation of the alternative moment conditions for the MSMM as per Clarke et al. (2015). These are the same moment conditions fit by the user-written Stata command ivpois (Nichols, 2007) and by the official Stata command ivpoisson gmm ..., multiplicative (Stata-Corp., 2013).
- "tsls" the TSLS method of fitting the MSMM of Clarke et al. (2015). For binary $Y$ and $X$ this uses $Y * (1 - X)$ as the outcome and $Y * X$ as the exposure.
- "tslsalt" the alternative TSLS method of fitting the MSMM of Clarke et al. (2015). For binary $Y$ and $X$ this uses $Y * X$ as the outcome and $Y * (1 - X)$ as the exposure.

| t0 | A vector of starting values for the gmm optimizer. This should have length equal to the number of exposures plus 1. |
|---|---|
| ... | further arguments passed to or from other methods. |

## Details

Function providing several methods to estimate the multiplicative structural mean model (MSMM) of Robins (1989). These are the methods described in Clarke et al. (2015), most notably generalised method of moments (GMM) estimation of the MSMM.

An equivalent estimator to the MSMM was proposed in Econometrics by Mullahy (1997) and then discussed in several articles by Windmeijer (1997, 2002) and Cameron and Trivedi (2013). This was implemented in the user-written Stata command ivpois (Nichols, 2007) and then implemented in official Stata in the ivpoisson command (StataCorp., 2013).

## Value

An object of class "msmm". A list with the following items:

| fit | The object from either a gmm::gmm() or ivreg::ivreg() fit. |
|---|---|
| crrci | The causal risk ratio/s and it corresponding 95% confidence interval limits. |
| estmethod | The specified estmethod. |

If estmethod is "tsls", "gmm", or "gmmalt":

| ey0ci | The estimate of the treatment/exposure free potential outcome and its 95% confidence interval limits. |
|---|---|

If estmethod is "tsls" or "tslsalt":

| stage1 | An object containing the first stage regression from an stats::lm() fit. |
|---|---|

## References

Cameron AC, Trivedi PK. Regression analysis of count data. 2nd ed. 2013. New York, Cambridge University Press. ISBN:1107667275

Clarke PS, Palmer TM, Windmeijer F. Estimating structural mean models with multiple instrumental variables using the Generalised Method of Moments. Statistical Science, 2015, 30, 1, 96-117. doi:10.1214/14STS503

Hernán and Robins. Instruments for causal inference: An Epidemiologist's dream? Epidemiology, 2006, 17, 360-372. doi:10.1097/01.ede.0000222409.00878.37

Mullahy J. Instrumental-variable estimation of count data models: applications to models of cigarette smoking and behavior. The Review of Economics and Statistics. 1997, 79, 4, 586-593. doi:10.1162/003465397557169

Nichols A. ivpois: Stata module for IV/GMM Poisson regression. 2007. url

Palmer TM, Sterne JAC, Harbord RM, Lawlor DA, Sheehan NA, Meng S, Granell R, Davey Smith G, Didelez V. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. American Journal of Epidemiology, 2011, 173, 12, 1392-1403. doi:10.1093/aje/kwr026

Robins JM. The analysis of randomised and nonrandomised AIDS treatment trials using a new approach to causal inference in longitudinal studies. In Health Service Research Methodology: A Focus on AIDS (L. Sechrest, H. Freeman and A. Mulley, eds.). 1989. 113–159. US Public Health Service, National Center for Health Services Research, Washington, DC.

StataCorp. Stata Base Reference Manual. Release 13. ivpoisson - Poisson model with continuous endogenous covariates. 2013. url

Windmeijer FAG, Santos Silva JMC. Endogeneity in Count Data Models: An Application to Demand for Health Care. Journal of Applied Econometrics. 1997, 12, 3, 281-294. doi:10/fdkh4n

Windmeijer, F. ExpEnd, A Gauss programme for non-linear GMM estimation of EXPonential models with ENDogenous regressors for cross section and panel data. CEMMAP working paper CWP14/02. 2002. url

## Examples

```
# Single instrument example
# Data generation from the example in the ivtools ivglm() helpfile
set.seed(9)
n    <- 1000
psi0 <- 0.5
Z    <- rbinom(n, 1, 0.5)
X    <- rbinom(n, 1, 0.7*Z + 0.2*(1 - Z))
m0   <- plogis(1 + 0.8*X - 0.39*Z)
Y    <- rbinom(n, 1, plogis(psi0*X + log(m0/(1 - m0))))
dat  <- data.frame(Z, X, Y)
fit  <- msmm(Y ~ X | Z, data = dat)
summary(fit)

# Multiple instrument example
set.seed(123456)
n    <- 1000
psi0 <- 0.5
```

```
G1   <- rbinom(n, 2, 0.5)
G2   <- rbinom(n, 2, 0.3)
G3   <- rbinom(n, 2, 0.4)
U    <- runif(n)
pX   <- plogis(0.7*G1 + G2 - G3 + U)
X    <- rbinom(n, 1, pX)
pY   <- plogis(-2 + psi0*X + U)
Y    <- rbinom(n, 1, pY)
dat2 <- data.frame(G1, G2, G3, X, Y)
fit2 <- msmm(Y ~ X | G1 + G2 + G3, data = dat2)
summary(fit2)
```

---

summary.msmm                     *Summarizing MSMM Fits*

---

### Description

Summarizing MSMM Fits

### Usage

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

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

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

### Arguments

| | |
|---|---|
| object | an object of class "msmm". |
| ... | further arguments passed to or from other methods. |
| | S3 summary and print methods for objects of class msmm and summary.msmm. |
| x | an object of class "summary.msmm". |
| digits | the number of significant digits to use when printing. |

### Value

summary.msmm() returns an object of class "summary.msmm". A list with the following elements:

| | |
|---|---|
| smry | An object from a call to either gmm::summary.gmm() or ivreg::summary.ivreg(). |
| object | The object of class msmm passed to the function. |

### Examples

```
# For examples see the examples at the bottom of help('msmm')
```

---

summary.tsps                    *Summarizing TSPS Fits*

---

### Description

S3 print and summary methods for objects of class `"tsps"` and print method for objects of class `"summary.tsps"`.

### Usage

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

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

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

### Arguments

| | |
|---|---|
| object | an object of class `"tsps"`. |
| ... | further arguments passed to or from other methods. |
| x | an object of class `"summary.tsps"`. |
| digits | the number of significant digits to use when printing. |

### Value

`summary.tsps()` returns an object of class `"summary.tsps"`. A list with the following elements:

| | |
|---|---|
| smry | An object from a call to [gmm::summary.gmm()](gmm::summary.gmm()) |
| object | The object of class `tsps` passed to the function. |

### Examples

```
# See the examples at the bottom of help('tsps')
```

---

summary.tsri                    *Summarizing TSRI Fits*

---

### Description

S3 print and summary methods for objects of class `"tsri"` and print method for objects of class `"summary.tsri"`.

### Usage

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

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

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

### Arguments

| | |
|---|---|
| object | an object of class `"tsri"`. |
| ... | further arguments passed to or from other methods. |
| x | an object of class `"summary.tsri"`. |
| digits | the number of significant digits to use when printing. |

### Value

`summary.tsri()` returns an object of class `"summary.tsri"`. A list with the following elements:

| | |
|---|---|
| smry | An object from a call to [gmm::summary.gmm()](gmm::summary.gmm()) |
| object | The object of class `tsps` passed to the function. |

### Examples

```
# See the examples at the bottom of help('tsri')
```

---

| tsps | *Two-stage predictor substitution (TSPS) estimators* |
|------|------------------------------------------------------|

---

## Description

Terza et al. (2008) give an excellent description of TSPS estimators. They proceed by fitting a first stage model of the exposure regressed upon the instruments (and possibly any measured confounders). From this the predicted values of the exposure are obtained. A second stage model is then fitted of the outcome regressed upon the predicted values of the exposure (and possibly measured confounders).

## Usage

```
tsps(
  formula,
  instruments,
  data,
  subset,
  na.action,
  contrasts = NULL,
  t0 = NULL,
  link = "identity",
  ...
)
```

## Arguments

formula, instruments

> formula specification(s) of the regression relationship and the instruments. Either instruments is missing and formula has three parts as in y ~ x1 + x2 | z1 + z2 + z3 (recommended) or formula is y ~ x1 + x2 and instruments is a one-sided formula ~ z1 + z2 + z3 (only for backward compatibility).

data

> an optional data frame containing the variables in the model. By default the variables are taken from the environment of the formula.

subset

> an optional vector specifying a subset of observations to be used in fitting the model.

na.action

> a function that indicates what should happen when the data contain NAs. The default is set by the na.action option.

contrasts

> an optional list. See the contrasts.arg of [stats::model.matrix()](stats::model.matrix()).

t0

> A vector of starting values for the gmm optimizer. This should have length equal to the number of exposures plus 1.

link

> character; one of "identity" (the default), "logadd", "logmult", "logit". This is the link function for the second stage model. "identity" corresponds to linear regression; "logadd" is log-additive and corresponds to Poisson / log-binomial regression; "logmult" is log-multiplicative and corresponds to gamma regression; "logit" corresponds to logistic regression.

... further arguments passed to or from other methods.

## Details

`tsps()` performs GMM estimation to ensure appropriate standard errors on its estimates similar to the approach described in Clarke et al. (2015).

## Value

An object of class `"tsps"` with the following elements

**fit** the fitted object of class `"gmm"` from the call to `gmm::gmm()`.

**estci** a matrix of the estimates with their corresponding confidence interval limits.

**link** a character vector containing the specified link function.

## References

Burgess S, CRP CHD Genetics Collaboration. Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. Statistics in Medicine, 2013, 32, 27, 4726-4747. doi:10.1002/sim.5871

Clarke PS, Palmer TM, Windmeijer F. Estimating structural mean models with multiple instrumental variables using the Generalised Method of Moments. Statistical Science, 2015, 30, 1, 96-117. doi:10.1214/14STS503

Dukes O, Vansteelandt S. A note on G-estimation of causal risk ratios. American Journal of Epidemiology, 2018, 187, 5, 1079-1084. doi:10.1093/aje/kwx347

Palmer TM, Sterne JAC, Harbord RM, Lawlor DA, Sheehan NA, Meng S, Granell R, Davey Smith G, Didelez V. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. American Journal of Epidemiology, 2011, 173, 12, 1392-1403. doi:10.1093/aje/kwr026

Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. Journal of Health Economics, 2008, 27, 3, 531-543. doi:10.1016/j.jhealeco.2007.09.009

## Examples

```
# Two-stage predictor substitution estimator
# with second stage logistic regression
set.seed(9)
n            <- 1000
psi0         <- 0.5
Z            <- rbinom(n, 1, 0.5)
X            <- rbinom(n, 1, 0.7*Z + 0.2*(1 - Z))
m0           <- plogis(1 + 0.8*X - 0.39*Z)
Y            <- rbinom(n, 1, plogis(psi0*X + log(m0/(1 - m0))))
dat          <- data.frame(Z, X, Y)
tspslogitfit <- tsps(Y ~ X | Z , data = dat, link = "logit")
summary(tspslogitfit)
```

---

tsri *Two-stage residual inclusion (TSRI) estimators*

---

### Description

An excellent description of TSRI estimators is given by Terza et al. (2008). TSRI estimators proceed by fitting a first stage model of the exposure regressed upon the instruments (and possibly any measured confounders). From this the first stage residuals are estimated. A second stage model is then fitted of the outcome regressed upon the exposure and first stage residuals (and possibly measured confounders).

### Usage

```
tsri(
  formula,
  instruments,
  data,
  subset,
  na.action,
  contrasts = NULL,
  t0 = NULL,
  link = "identity",
  ...
)
```

### Arguments

formula, instruments

formula specification(s) of the regression relationship and the instruments. Either instruments is missing and formula has three parts as in y ~ x1 + x2 | z1 + z2 + z3 (recommended) or formula is y ~ x1 + x2 and instruments is a one-sided formula ~ z1 + z2 + z3 (only for backward compatibility).

data             an optional data frame containing the variables in the model. By default the variables are taken from the environment of the formula.

subset           an optional vector specifying a subset of observations to be used in fitting the model.

na.action        a function that indicates what should happen when the data contain NAs. The default is set by the na.action option.

contrasts        an optional list. See the contrasts.arg of [stats::model.matrix()](stats::model.matrix()).

t0               A vector of starting values for the gmm optimizer. This should have length equal to the number of exposures plus 1.

link             character; one of "identity" (the default), "logadd", "logmult", "logit". This is the link function for the second stage model. "identity" corresponds to linear regression; "logadd" is log-additive and corresponds to Poisson / log-binomial regression; "logmult" is log-multiplicative and corresponds to gamma regression; "logit" corresponds to logistic regression.

`...`                                 further arguments passed to or from other methods.

## Details

TSRI estimators are sometimes described as a special case of control function estimators.

`tsri()` performs GMM estimation to ensure appropriate standard errors on its estimates similar to that described that described by Clarke et al. (2015). Terza (2017) described an alternative approach.

## Value

An object of class `"tsri"` with the following elements

**fit**  the fitted object of class `"gmm"` from the call to [gmm::gmm()](gmm::gmm()).

**estci**  a matrix of the estimates with their corresponding confidence interval limits.

**link**  a character vector containing the specified link function.

## References

Bowden J, Vansteelandt S. Mendelian randomization analysis of case-control data using structural mean models. Statistics in Medicine, 2011, 30, 6, 678-694. doi:10.1002/sim.4138

Clarke PS, Palmer TM, Windmeijer F. Estimating structural mean models with multiple instrumental variables using the Generalised Method of Moments. Statistical Science, 2015, 30, 1, 96-117. doi:10.1214/14STS503

Dukes O, Vansteelandt S. A note on G-estimation of causal risk ratios. American Journal of Epidemiology, 2018, 187, 5, 1079-1084. doi:10.1093/aje/kwx347

Palmer T, Thompson JR, Tobin MD, Sheehan NA, Burton PR. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. International Journal of Epidemiology, 2008, 37, 5, 1161-1168. doi:10.1093/ije/dyn080

Palmer TM, Sterne JAC, Harbord RM, Lawlor DA, Sheehan NA, Meng S, Granell R, Davey Smith G, Didelez V. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. American Journal of Epidemiology, 2011, 173, 12, 1392-1403. doi:10.1093/aje/kwr026

Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. Journal of Health Economics, 2008, 27, 3, 531-543. doi:10.1016/j.jhealeco.2007.09.009

Terza JV. Two-stage residual inclusion estimation: A practitioners guide to Stata implementation. The Stata Journal, 2017, 17, 4, 916-938. doi:10.1177/1536867X1801700409

## Examples

```
# Two-stage residual inclusion estimator
# with second stage logistic regression
set.seed(9)
n           <- 1000
psi0        <- 0.5
Z           <- rbinom(n, 1, 0.5)
```

```
X            <- rbinom(n, 1, 0.7*Z + 0.2*(1 - Z))
m0           <- plogis(1 + 0.8*X - 0.39*Z)
Y            <- rbinom(n, 1, plogis(psi0*X + log(m0/(1 - m0))))
dat          <- data.frame(Z, X, Y)
tsrilogitfit <- tsri(Y ~ X | Z , data = dat, link = "logit")
summary(tsrilogitfit)
```

# Index