

Package ‘SIVMethod’

September 27, 2026

Type Package

Title Identification, Estimation and Inference Based on Structural Error Projection

Version 0.1.0

Description Estimation and inference for regression models with endogenous regressors using a semi-parametric projection approach. Instrumental variables are constructed internally from observed regressors by projecting out a space of basis functions used to represent the conditional mean of the structural error. A least absolute shrinkage and selection operator (LASSO) procedure selects basis functions for the projection. Tools are provided for simulation studies and empirical applications. The methods are based on Dong, Gao, Linton and Peng (2026) <[doi:10.48550/arXiv.2607.05699](https://doi.org/10.48550/arXiv.2607.05699)>.

License MIT + file LICENSE

Encoding UTF-8

URL <https://github.com/Greatknee/SIVMethod>

BugReports <https://github.com/Greatknee/SIVMethod/issues>

Depends R (>= 3.5)

Imports glmnet, stats, utils

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Config/roxygen2/version 8.0.0

Date 2026-09-14

NeedsCompilation no

Author Yuhuai Chen [aut, cre],
Chaohua Dong [aut],
Jiti Gao [aut],
Linton Oliver [aut],
Bin Peng [aut]

Maintainer Yuhuai Chen <yuhuai.chen0351@gmail.com>

Repository CRAN

Date/Publication 2026-09-27 16:00:09 UTC

Contents

Sim	2
Sim_test	3
SIV	3
Index	5

Sim	<i>Monte Carlo Simulation for Semiparametric Instrumental Variable Estimation</i>
-----	---

Description

Conducts Monte Carlo simulations for the designs in Example B.2.1. Compares ordinary least squares and the semiparametric projection method. Reports empirical bias, Monte Carlo standard deviations, and basis selection frequencies.

Usage

```
Sim(n, no_sim, case, k = NULL)
```

Arguments

- n Integer. Number of observations.
- no_sim Integer. Number of simulation replications.
- case Character. One of "A", "B", "C" or "D".
- k Integer. The truncation parameter.

Value

A list with:

- Output** Summary statistics matrix.
- Selection** Average selection frequencies.

Examples

```
Sim(100L, 200L, case = "A")
```

Sim_test

*Monte Carlo Simulation for Testing Theory***Description**

Evaluates the finite-sample size and power of the endogeneity test proposed in Section 3.2 using the simulation designs in Example B.2.3. Empirical rejection rates are reported across Monte Carlo replications.

Usage

```
Sim_test(n, no_sim, case = "A", c = 0.75, k = NULL)
```

Arguments

n	Integer. Number of observations.
no_sim	Integer. Number of simulation replications.
case	Character. One of "A" or "B".
c	Numeric scalar with $c \geq 0$. Used in $\epsilon_i = c\sqrt{k/nm}(x_i)$. If $c = 0$, do not need to specify case.
k	Integer vector. The range of truncation parameters to be considered via Generalized Cross-Validation.

Value

A list with:

Output Summary statistics matrix.

Rejection_rate Average test rejection rate.

Examples

```
Sim_test(100L, 200L, case = "A", c = 1)
```

SIV

*Semiparametric Instrumental Variable Estimation***Description**

Estimation and inference for regression models with endogenous regressors using a semiparametric projection approach. Instrumental variables are constructed internally from observed regressors by projecting out a space of basis functions used to represent the conditional mean of the structural error. A least absolute shrinkage and selection operator (LASSO) procedure selects basis functions for the projection. See further details in Dong, Gao, Linton and Peng (2026) <<https://arxiv.org/abs/2607.05699>>.

Usage

SIV(Y, X, k)

Arguments

Y	Response vector, n-by-1.
X	Regressor matrix, n-by-d.
k	Integer vector including the possible values of the truncation parameter, L_k-by-1.

Value

A list containing:

b_SIV_GCV, b_SIV_LSO Regression coefficient estimates using the full basis and post-LASSO basis, respectively.

gam_SIV_GCV, gam_LSO Corresponding series coefficient estimates.

sd_GCV, sd_LSO Standard errors of the regression estimates.

sd_gam Standard errors of gam_LSO.

e_GCV, e_LSO Residuals from the respective fits.

eps_LSO Estimated $m(X_i) + e_i$, assuming the first selected basis column is the constant term.

V_LSO Post-LASSO basis matrix, including a constant column.

Ind_LSO Logical selection indicators for the original basis.

m_all Four-column matrix containing estimated $m(x)$, lower and upper pointwise 95% wild bootstrap confidence limits, and evaluation points, respectively.

Index

Sim, [2](#)
Sim_test, [3](#)
SIV, [3](#)