Robust Inference for Non-Gaussian Linear Simultaneous Equations Models

Adam Lee & Geert Mesters (UPF & BSE)

May 10, 2022

Introduction

- Independence, heteroskedasticity, non-Gaussian distributions ... can all help to identify structural parameters
  - Simultaneous equations
    [Lanne and Lütkepohl, 2010, Moneta et al., 2013, Lanne et al., 2017, Maxand, 2018, Lanne and Luoto, 2019, Gouriéroux et al., 2017, Gouriéroux et al., 2019, Tank et al., 2019, Herwartz, 2019, Herwartz et al., 2019, Bekaert et al., 2019, Bekaert et al., 2020, Fiorentini and Sentana, 2020, Velasco, 2020, Guay, 2020, Moneta and Pallante, 2020, Drautzburg and Wright, 2021, Sims, 2021, Davis and Ng, 2022]
  - Measurement error models
    [Kendall and Stuart, 1970, Pal, 1980, Kapteyn and Wansbeek, 1983, Cardoso and Souloumiac, 1994, Dagenais and Dagenais, 1997, Ikeda and Toyama, 2000, Erickson and Whited, 2002, Bonhomme and Robin, 2009, Schennach, 2021]
  - Triangular systems
    [Klein and Vella, 2010, Lewbel, 2012, Lewbel, 2019, Lewbel et al., 2021]

Introduction

- Usage of such "statistical" identification methods is not without criticism [Olea et al., 2021]
- If assumptions (nearly) fail parameters become unidentified and standard inference methods fail to control size
- We would like to have inference methods that are uniformly valid
- This paper: robust inference for non-Gaussian linear simultaneous equations models

Simple example

Consider

$$Y_i = R'\epsilon_i \ , \qquad R = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} ,$$

where

- $\epsilon_i = (\epsilon_{i,1}, \epsilon_{i,2})'$ are independent with common density η
- $\mathbb{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = I$

Goal: conduct inference on α given $\{Y_1, \ldots, Y_n\}$

Identification problem

$$Y_i = R' \epsilon_i \ , \qquad \epsilon_{i,k} \sim \eta$$

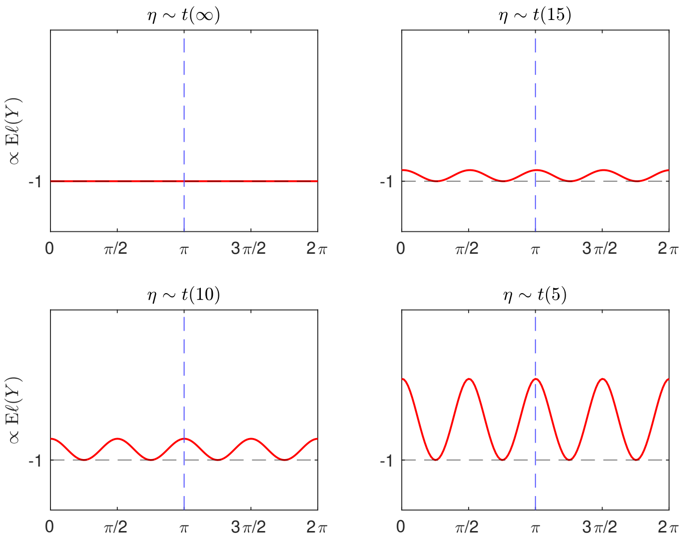
- Identification depends on η
  - if $\eta =$ Gaussian, expected log likelihood

    $$\mathbb{E}\ell(Y_i) \propto -\frac{1}{2}\mathbb{E}(RY_i)'(RY_i) = -1$$

    same value for all α
  - if $\eta \neq$ Gaussian, α is identified up to permutation and sign [Comon, 1994].

    $$\{R'PD : P \in \Pi(K), D \in \mathrm{S}(K)\}$$

    from Darmois-Skitovich theorem [Darmois, 1953,Skitivic, 1953]



$\eta \sim t(\infty)$
$\eta \sim t(15)$
$\eta \sim t(10)$
$\eta \sim t(5)$
$\propto \mathrm{E}\ell(Y)$
-1
0
$\pi/2$
π
$3\pi/2$
2π

Standard inference approach

$$Y_i = R' \epsilon_i \ , \qquad \epsilon_{i,k} \sim \eta$$

(i) **Assume** $\eta \neq$ Gaussian

(ii) Estimate α

- parametric or non-parametric
- likelihood-based or moment-based

(iii) Conduct hypothesis tests/ make confidence intervals

Finite sample size distortions

Table: Empirical Rejection Frequencies $a = 0.05$

η	$t(\infty)$		$t(15)$		$t(10)$		$t(5)$	
n	200	500	200	500	200	500	200	500
W	0.378	0.352	0.314	0.329	0.207	0.177	0.076	0.042
LR	0.000	0.000	0.024	0.037	0.065	0.059	0.060	0.051
LM	0.042	0.051	0.043	0.048	0.051	0.046	0.045	0.043

- η is fixed at the true value
- α estimated by maximum likelihood
- Test $H_0 : \alpha = \alpha_0$ vs $H_1 : \alpha \neq \alpha_0$
- Wald (W), Likelihood ratio (LR), Lagrange multiplier (LM)

Handling η

- LM test works well

  Intuition, LM fixes $\alpha = \alpha_0$ under H_0; no need for identification

- In practice, η is unknown

  (i) Parametric approach: $\eta = \eta(\beta)$ with β finite dimensional nuisance parameter; LM becomes Neyman's $C(\alpha)$ details

  (ii) Semi-parametric approach: η infinite dimensional nuisance parameter $\Rightarrow$ **This paper**

Intuition for semi-parametric score test

$$Y_i = R'\epsilon_i \ , \qquad \epsilon_{i,k} \sim \eta$$

- Orthogonalize scores for α with respect to scores for η

  $\Rightarrow$ efficient score function $\kappa(Y_i)$ and information matrix $\mathcal{I} = \mathbb{E}\kappa(Y_i)\kappa(Y_i)'$, see [Bickel et al., 1998, van der Vaart, 2002]

- Under $H_0 : \alpha = \alpha_0$ we have $\mathbb{E}\kappa(Y_i) = 0$, i.e. moment conditions

- Estimate $\hat{\kappa}(Y_i)$ and $\hat{\mathcal{I}}$, and construct

$$\hat{S} = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\kappa}(Y_i) \right)' \hat{\mathcal{I}}^{t,\dagger} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\kappa}(Y_i) \right)$$

  under mild regularity conditions $\hat{S} \xrightarrow{d} \chi^2$

Comments

- Two difficulties are handled "simultaneously"
  - α is fixed $\Rightarrow$ avoids need for identification
  - Orthogonalizing wrt η allows for regularized estimation of η
- At points of no-identification $\mathcal{I}$ can be singular, need eigenvalue truncated estimate

  [Lütkepohl and Burda, 1997, Andrews and Guggenberger, 2019]
- Computationally trivial, only B-spline regressions to estimate $\kappa(Y_i)$

This paper

- Consider
$$Y_i = BX_i + A(\alpha, \sigma)^{-1}\epsilon_i$$
as **semi-parametric** model
  - Parametric: α, $\beta = (\sigma, b)$, with $b = \text{vec}(B)$
  - Non-parametric: η_k density $\epsilon_{i,k}$

- Test $H_0 : \alpha = \alpha_0$ using identification/singularity robust **semi-parametric score statistic**
  - Yields correct size/coverage regardless of **distance to Gaussianity**
  - Under non-singularity – asymptotically uniformly most powerful invariant (AUMPI) test

Extensions

- Same test applies for general semi-parametric models P_θ with $\theta = (\alpha, \beta, \eta)$
  - Robust to identification problems, boundary problems, regularization
  - Examples: single-index models, duration models, shape restricted models, (non-)parametric IV, etc
  - Paper provides a general theory, fully expounded in [Lee, 2022] (uniformity and optimality results)
- Linear dynamic models (e.g. SVAR): identical implementation, substantially more cumbersome proofs

  [Hoesch et al., 2020] nearly finished:-)

What about GMM?

- Higher order moment restrictions can also identify α, e.g. $\mathbb{E}\epsilon_{i,1}^2\epsilon_{i,2} = \mathbb{E}\epsilon_{i,2}^2\epsilon_{i,1} = 0$ are sufficient in simple example.
  - Combined with identification robust GMM statistic yields size correct inference [Drautzburg and Wright, 2021]
  - Benefit: do not need to assume full independence
  - Downsides: typically requires (a) many moments and (b) existence higher order moments

Road map

(i) Model & Implementation

(ii) Some theory

(iii) Simulation results

(i) Model & Implementation

General LSEM model

$$Y_i = BX_i + A(\alpha, \sigma)^{-1}\epsilon_i \ , \qquad i = 1, \ldots, n \ ,$$

where

- Y_i is $K \times 1$ and X_i is $d \times 1$
- ϵ_i is $K \times 1$ independent components $\mathbb{E}\epsilon_i = 0$ and $\mathrm{Var}(\epsilon_i) = I_K$
- Parameters

$$\theta = (\gamma, \eta) \qquad \gamma = (\alpha, \beta) \qquad \beta = (\sigma, b) \qquad b = \mathrm{vec}(B)$$

  - α possibly unidentified
  - β identified finite dimensional
  - η identified infinite dimensional

- Test

$$H_0 : \alpha = \alpha_0 \qquad \text{against} \qquad H_1 : \alpha \neq \alpha_0 \ .$$

Implementing efficient score test (i)

Efficient score function estimates (orthogonalized wrt η)

$$\hat{\ell}_{\gamma}(V_i) = \begin{bmatrix} \hat{\ell}_{\gamma,\alpha}(V_i) \\ \hat{\ell}_{\gamma,\beta}(V_i) \end{bmatrix} \quad \text{with} \quad \hat{\ell}_{\beta}(V_i) = \begin{bmatrix} \hat{\ell}_{\gamma,\sigma}(V_i) \\ \hat{\ell}_{\gamma,b}(V_i) \end{bmatrix}$$

where $V_i = Y_i - BX_i$ with components

$$\hat{\ell}_{\gamma,\alpha_l}(V_i) = \sum_{j,k=1,j\neq k}^{K} \zeta^{\alpha}_{l,k,j} \hat{\phi}_k(A_{k\bullet}V_i)A_{j\bullet}V_i + \sum_{k=1}^{K} \zeta^{\alpha}_{l,k,k}\left[\hat{\tau}_{k,1}A_{k\bullet}V_i + \hat{\tau}_{k,2}\kappa(A_{k\bullet}V_i)\right]$$

$$\hat{\ell}_{\gamma,\sigma_l}(V_i) = \sum_{j,k=1,j\neq k}^{K} \zeta^{\sigma}_{l,k,j} \hat{\phi}_k(A_{k\bullet}V_i)A_{j\bullet}V_i + \sum_{k=1}^{K} \zeta^{\sigma}_{l,k,k}\left[\hat{\tau}_{k,1}A_{k\bullet}V_i + \hat{\tau}_{k,2}\kappa(A_{k\bullet}V_i)\right]$$

$$\hat{\ell}_{\gamma,b_l}(V_i) = \sum_{k=1}^{K}[-A_{k\bullet}D_{b,l}][(X_i - \bar{X}_n)\hat{\phi}_k(A_{k\bullet}V_i) - \bar{X}_n(\hat{\varsigma}_{k,1}A_{k\bullet}V_i + \hat{\varsigma}_{k,2}\kappa(A_{k\bullet}V_i))]$$

Implementing efficient score test (i)

Components:

- *Known functions*: $\zeta^{\alpha}_{l,k,j} := [D_{\alpha,l}]_{k\bullet} A^{-1}_{\bullet j}$, $\zeta^{\sigma}_{l,k,j} := [D_{\sigma,l}]_{k\bullet} A^{-1}_{\bullet j}$, $D_{\alpha,l} = \partial A(\alpha,\sigma)/\partial\alpha_l$, $D_{\sigma,l} = \partial A(\alpha,\sigma)/\partial\sigma_l$, $D_{b_l} = \partial B/\partial b_l$ and $\kappa(z) = 1 - z^2$
- *Simple coefficients*: $\hat{\tau}_k = (\hat{\tau}_{k,1}, \hat{\tau}_{k,2})'$ and $\hat{\varsigma}_k = (\hat{\varsigma}_{k,1}, \hat{\varsigma}_{k,2})'$

$$\hat{\tau}_k = \hat{M}_k^{-1} \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \hat{\varsigma}_k = \hat{M}_k^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\hat{M}_k = \begin{pmatrix} 1 & \frac{1}{n}\sum_{i=1}^n (A_{k\bullet} V_i)^3 \\ \frac{1}{n}\sum_{i=1}^n (A_{k\bullet} V_i)^3 & \frac{1}{n}\sum_{i=1}^n (A_{k\bullet} V_i)^4 - 1 \end{pmatrix}.$$

- *Log density score estimates*: $\hat{\phi}_k$ estimates $\phi_k(z) = \partial \log \eta(z)/\partial z$, computed using B-spline regressions [Chen and Bickel, 2006]

Implementing efficient score test (ii)

Given $\hat{\ell}_\gamma(V_i)$ we compute information matrix

$$\hat{I}_\gamma = \frac{1}{n}\sum_{i=1}^{n} \hat{\ell}_\gamma(V_i)\hat{\ell}_\gamma(V_i)' \quad \text{with partitioning} \quad \hat{I}_\gamma = \begin{bmatrix} \hat{I}_{\gamma,\alpha\alpha} & \hat{I}_{\gamma,\alpha\beta} \\ \hat{I}_{\gamma,\beta\alpha} & \hat{I}_{\gamma,\beta\beta} \end{bmatrix}.$$

Implementing efficient score test (iii)

Efficient score test

$$\hat{S}_\gamma = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\kappa}_\gamma(V_i) \right)' \hat{\mathcal{I}}_\gamma^{t,\dagger} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\kappa}_\gamma(V_i) \right) ,$$

where

$$\hat{\kappa}_\gamma(V_i) = \hat{\ell}_{\gamma,\alpha}(V_i) - \hat{I}_{\gamma,\alpha\beta} \hat{I}_{\gamma,\beta\beta}^{-1} \hat{\ell}_{\gamma,\beta}(V_i) , \qquad \hat{\mathcal{I}}_\gamma = \hat{I}_{\gamma,\alpha\alpha} - \hat{I}_{\gamma,\alpha\beta} \hat{I}_{\gamma,\beta\beta}^{-1} \hat{I}_{\gamma,\beta\alpha} .$$

and $\hat{\mathcal{I}}_\gamma^{t,\dagger}$ denotes the generalized inverse of eigenvalue truncated $\hat{\mathcal{I}}_\gamma$ [Lütkepohl and Burda, 1997,Andrews and Guggenberger, 2019]

Main result

We prove (given assumptions), under H_0 for any $a \in (0,1)$ we have for $\hat{\gamma} = (\alpha_0, \hat{\beta})$, with $\hat{\beta}$ some $\sqrt{n}$-consistent estimate, that

$$\lim_{n \to \infty} P(\hat{S}_{\hat{\gamma}} > c_n) \leq a \, ,$$

where c_n is the $1 - a$ quantile of the $\chi^2_{r_n}$ distribution with $r_n = \text{rank}(\hat{\mathcal{I}}^t_{\hat{\gamma}})$.

- Crucially, no assumption regarding the shape of η is needed, only regularity conditions
- $\hat{S}_{\hat{\gamma}}$ is uniformly most powerful when $\hat{\mathcal{I}}$ is non-singular, and minimax optimal for singular $\hat{\mathcal{I}}$.

Algorithm 1: Efficient score test for LSEM

1 Obtain $\sqrt{n}$-consistent estimates $\hat{\beta} = (\hat{\sigma}, \hat{b})$ and residuals $\hat{V}_i = Y_i - \hat{B} X_i$;

2 For $k = 1, \ldots, K$, compute the log density scores $\hat{\phi}_k(\hat{V}_i)$ using B-spline regressions;

3 Compute the efficient scores $\hat{\ell}_{\hat{\gamma}}(\hat{V}_i)$ and the information matrix $\hat{I}_{\hat{\gamma}}$ with $\hat{\gamma} = (\alpha_0, \hat{\beta})$;

4 Compute $\hat{\kappa}_{\hat{\gamma}}(\hat{V}_i)$ and $\hat{\mathcal{I}}_{\hat{\gamma}}$.

5 Compute the score statistic $\hat{S}_{\hat{\gamma}}$ and reject $H_0 : \alpha = \alpha_0$ if $\hat{S}_{\hat{\gamma}} > c_n$, where c_n is the $1 - a$ quantile of the $\chi^2_{r_n}$ distribution with $r_n = \text{rank}(\hat{\mathcal{I}}^t_{\hat{\gamma}})$.

(ii) Assumptions and formal results

Assumptions

Two types of assumptions

- "Unavoidable" model restrictions (moments, differentiability, etc)
- "Avoidable" regularity conditions for density score estimation

Assumption 1: model

For $\epsilon_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,K})'$, each component $\epsilon_{i,k}$ has a continuously differentiable root density (where the density is with respect to Lebesgue measure on $\mathbb{R}$). We write the density as η_k with log density score $\phi_k(x) = \partial \log \eta_k(x)/\partial x$. We assume that for all $k = 1, \ldots, K$ and some $\delta > 0$

(i) $\mathbb{E}\epsilon_{i,k} = 0$, $\mathbb{E}\epsilon_{i,k}^2 = 1$, $\mathbb{E}\epsilon_{i,k}^{4+\delta} < \infty$, $\mathbb{E}(\epsilon_{i,k}^4) - 1 > \mathbb{E}(\epsilon_{i,k}^3)^2$, $\mathbb{E}{\phi_k}^{4+\delta}(\epsilon_{i,k}) < \infty$

(ii) $\mathbb{E}\phi_k(\epsilon_{i,k}) = 0$, $\mathbb{E}\phi_k(\epsilon_{i,k})\epsilon_{i,k} = -1$, $\mathbb{E}\phi_k(\epsilon_{i,k})\epsilon_{i,k}^2 = 0$, $\mathbb{E}\phi_k(\epsilon_{i,k})\epsilon_{i,k}^3 = -3$

(iii) $\epsilon_{i,k}$ is independent of $\epsilon_{i,l}$ for all $k \neq l$

(iv) Let $X_i = (1, \tilde{X}_i')'$, the density $\eta_0 \in \mathscr{X}$ is a density function (with respect to Lebesgue measure on $\mathbb{R}^{d-1}$) such that if $\tilde{X}_i \sim \eta_0$, then $\mathbb{E}\tilde{X}_i\tilde{X}_i'$ is positive definite and $\mathbb{E}[|\tilde{X}_{i,l}|^{4+\delta}] < \infty$ for all $l = 1, \ldots, d-1$

(v) ϵ_i and $\tilde{X}_i$ are independent.

Assumption 2: log density score

Let $\phi_{k,n} := \phi_k 1_{[\Xi^L_{k,n},\Xi^U_{k,n}]}$ and $\Delta_{k,n} := \Xi^U_{k,n} - \Xi^L_{k,n}$ and suppose that for ν_n defined by $\nu^2_{n,p} = o(\nu_n)$ with $p := \min\{1+\delta/4, 2\}$ and $\nu_{n,p} = n^{(1-p)/p}$ if $p \in (1,2)$ or $\nu_{n,p} = n^{-1/2}\log(n)^{1/2+\rho}$, for some $\rho > 0$, if $p = 2$, $[\Xi^L_{k,n}, \Xi^U_{k,n}] \uparrow \tilde{\Xi} \supset \mathrm{supp}(\eta_k)$ and $\delta_{k,n} \downarrow 0$ such that

(i) $P(\epsilon_{i,k} \notin [\Xi^L_{k,n}, \Xi^U_{k,n}]) = o(\nu_n^2)$;

(ii) For some $\iota > 0$, $n^{-1}\Delta^{2+2\iota}_{k,n}\delta^{-(8+2\iota)}_{k,n} = o(\nu_n)$;

(iii) η_k is bounded ($\|\eta_k\|_\infty < \infty$) and differentiable, with a bounded derivative: $\|\eta'_k\|_\infty < \infty$;

(iv) For each n, $\phi_{k,n}$ is three-times continuously differentiable on $[\Xi^L_{k,n}, \Xi^U_{k,n}]$ and $\|\phi^{(3)}_{k,n}\|^2_\infty \delta^6_{k,n} = o(\nu_n)$;

(v) There are $c > 0$ and $N \in \mathbb{N}$ such that for $n \geq N$ we have $\inf_{t\in[\Xi^L_{k,n},\Xi^U_{k,n}]} |\eta_k(t)| \geq c\delta_{k,n}$.

Main result

Theorem

Suppose that Assumptions 1 and 2 hold, that $(\alpha, \sigma) \mapsto A(\alpha, \sigma)$ *is continuously differentiable and the maps* $(\alpha, \sigma) \to \zeta^{\alpha}_{l,k,j}$ *and* $(\alpha, \sigma) \to \zeta^{\sigma}_{l,k,j}$ *are Lipschitz continuous. Let* $r_n = \mathrm{rank}(\hat{\mathcal{I}}^{t}_{\bar{\gamma}})$ *and denote by* c_n *the* $1 - a$ *quantile of the* $\chi^2_{r_n}$ *distribution, for any* $a \in (0, 1)$*. Then, under* H_0

$$\lim_{n\to\infty} P_{\theta_0}(\hat{S}_{\bar{\gamma}} > c_n) \leq a,$$

with inequality only if $\mathrm{rank}(\tilde{\mathcal{I}}_{\gamma_0}) = 0$ *where* $\gamma_0 = (\alpha_0, \sigma)$*.*

(iii) Simulations

Baseline model

$$Y_i = R' \epsilon_i \ ,$$

where

- R is orthogonal parametrized by Cayley transform of skew-symmetric matrix
- $\epsilon_{i,1} \sim N(0,1)$, $\epsilon_{i,k} \sim$ different densities
- $K = 2, 3, 5$ and $n = 200, 500, 1000$
- Simulate 5000 data sets, compute $\hat{S}_{\hat{\gamma}}$, report rejection frequencies
- No additional nuisance parameters β

Errors

Table: TRUE ERROR DISTRIBUTIONS $\epsilon_{i,k}$

	Distribution
1	$\mathcal{N}(0, 1)$
2	$t(15)$
3	$t(10)$
4	$t(5)$
5	"skewed unimodal"
6	"kurtotic unimodal"
7	"outlier"
8	"bimodal"
9	"separate bimodal"
10	"skewed bimodal"

Table: Rejection Frequencies $\hat{S}_{\hat{\gamma}}$ test for Baseline model

n	K	B	1	2	3	4	5	6	7	8	9	10
200	2	4	4.9	4.9	4.8	4.0	4.7	4.9	4.7	5.0	5.0	5.2
200	2	6	4.8	4.5	4.9	4.4	4.8	5.3	5.4	4.5	5.1	5.0
200	2	8	5.0	4.9	4.7	4.4	4.8	4.8	5.0	5.0	5.1	4.6
200	3	4	4.3	3.9	3.9	3.9	4.4	4.8	4.5	4.9	4.5	4.5
200	3	6	4.5	3.8	4.0	4.4	4.1	4.8	4.8	4.7	4.6	4.5
200	3	8	4.7	4.6	4.1	4.0	4.4	4.8	4.3	4.9	5.1	4.5
200	5	4	3.2	3.5	3.4	3.4	3.5	4.0	3.0	4.2	4.3	4.2
200	5	6	4.7	5.4	3.7	3.2	3.3	4.0	3.2	4.5	4.4	4.0
200	5	8	4.9	4.9	3.9	5.1	3.5	4.4	3.1	4.1	4.1	4.2

Table: REJECTION FREQUENCIES $\hat{S}_{\hat{\gamma}}$ TEST FOR BASELINE MODEL

n	K	B	1	2	3	4	5	6	7	8	9	10
500	2	4	5.2	4.3	4.7	4.7	4.2	5.1	5.2	4.9	5.3	5.0
500	2	6	5.3	4.3	4.2	5.0	4.7	4.8	5.0	4.8	4.6	4.9
500	2	8	5.2	4.9	4.8	4.6	4.7	4.7	4.9	5.6	4.7	4.9
500	3	4	4.4	3.9	4.5	4.3	4.8	5.1	5.0	4.5	5.1	5.0
500	3	6	4.3	4.6	4.0	4.1	4.6	4.9	5.0	4.8	5.0	5.1
500	3	8	5.0	4.2	4.1	4.2	4.7	4.7	5.1	4.8	4.6	4.6
500	5	4	4.4	3.4	3.5	3.6	4.3	4.7	3.7	4.3	4.4	4.9
500	5	6	4.6	4.1	3.7	4.3	4.6	5.0	4.5	5.1	4.9	4.8
500	5	8	4.0	4.0	4.2	3.6	4.6	4.7	4.1	4.6	4.6	4.8

Table: Rejection Frequencies $\hat{S}_{\hat{\gamma}}$ test for Baseline model

n	K	B	1	2	3	4	5	6	7	8	9	10
1000	2	4	4.7	5.4	5.2	5.0	4.6	4.9	5.8	5.6	4.8	5.0
1000	2	6	4.6	4.9	4.9	5.1	5.2	5.0	5.6	5.0	4.8	5.3
1000	2	8	4.8	4.9	4.9	5.1	5.1	5.5	5.1	4.9	4.8	4.8
1000	3	4	4.5	4.1	4.6	4.2	4.6	5.0	5.6	4.6	4.9	5.1
1000	3	6	4.7	4.6	4.1	4.5	4.9	4.9	4.7	5.0	4.9	5.7
1000	3	8	5.1	4.6	5.0	4.2	4.6	5.0	5.2	5.2	4.9	4.6
1000	5	4	3.6	3.7	3.5	4.4	4.4	4.7	4.0	4.7	5.1	5.1
1000	5	6	3.8	3.8	3.9	4.2	4.5	4.6	5.0	5.3	4.5	4.8
1000	5	8	4.9	4.2	4.3	4.1	5.1	4.2	4.4	5.1	4.5	4.6

Alternative methods

Two categories:

(i) Should fail:

- W^{mle} and LR^{mle}
- W^{pmle} and LR^{pmle}, from [Gouriéroux et al., 2017]
- LR^{gmm} from [Lanne and Luoto, 2019]

(ii) Promising:

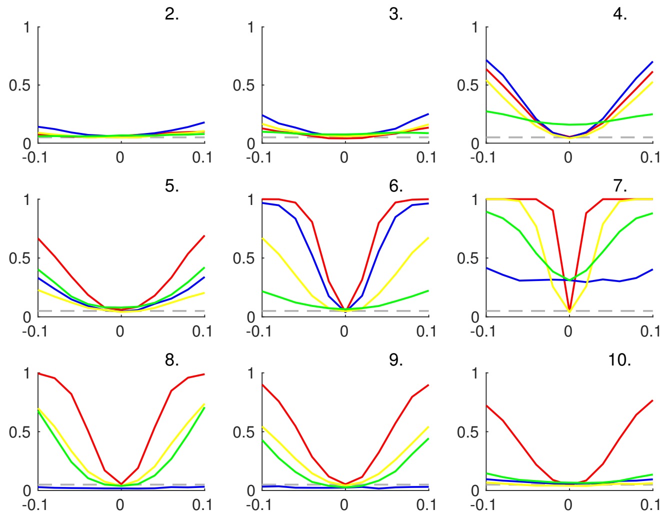
- LM^{mle}
- LM^{pmle}, inspired by [Gouriéroux et al., 2017]
- S^{gmm} from [Drautzburg and Wright, 2021]

Table: Rejection Frequencies Alternative Tests for Baseline model

		Category (i)								
Test	n	1	2	3	4	5	6	7	8	9
W^{mle}	500	18.0	13.4	11.5	11.5	9.5	16.7	59.2	11.5	12.0
	1000	18.8	10.1	8.0	7.5	6.2	40.5	46.0	12.4	11.8
LR^{mle}	500	2.3	5.7	6.8	6.4	6.5	2.3	22.9	1.6	1.5
	1000	4.9	6.5	6.3	6.2	5.3	3.1	15.8	2.2	2.0
W^{pmle}	500	49.0	23.6	16.0	6.0	16.2	2.9	0.9	98.9	59.9
	1000	56.4	27.4	18.0	5.3	13.8	2.6	1.5	99.9	97.2
LR^{pmle}	500	40.9	24.3	15.5	5.7	13.2	2.5	0.8	98.0	95.1
	1000	51.8	24.8	16.0	4.8	11.9	2.3	1.4	99.9	99.5
LR^{gmm}	500	29.3	24.7	24.7	28.7	14.1	17.2	38.5	10.9	13.6
	1000	23.2	18.1	15.6	17.7	7.4	11.5	22.9	6.9	9.2

Table: REJECTION FREQUENCIES ALTERNATIVE TESTS FOR BASELINE MODEL

		Category (ii)								
Test	n	1	2	3	4	5	6	7	8	9
$\hat{S}_{\hat{\gamma}}$	500	4.7	4.7	5.4	4.7	4.5	4.4	5.2	4.8	5.3
	1000	4.7	4.3	4.6	4.9	4.8	4.8	5.3	4.4	5.1
LM^{mle}	500	5.7	5.2	5.3	4.3	4.6	4.7	33.2	1.8	2.3
	1000	6.2	5.3	5.0	4.9	4.0	4.0	33.3	1.6	2.6
LM^{plme}	500	5.0	4.4	5.1	4.2	4.5	4.2	4.2	4.9	4.3
	1000	5.2	4.9	4.8	4.3	4.8	5.6	5.0	5.5	4.9
S^{gmm}	500	9.4	10.6	12.3	22.3	11.6	13.4	47.7	5.8	6.8
	1000	6.2	7.0	8.2	16.3	6.9	7.9	32.6	3.1	3.8



$\hat{S}_{\hat{\gamma}}$, $\mathrm{LM}^{\mathrm{mle}}$, $\mathrm{LM}^{\mathrm{plme}}$, $\mathrm{S}^{\mathrm{gmm}}$

LSEM model

$$Y_i = BX_i + \Sigma^{1/2} R' \epsilon_i \ .$$

where

- Same set-up, but now with nuisance parameters $\beta = (\text{vec}(B)', \text{vech}(\Sigma^{1/2})')'$
- X_i is $d \times 1$, with $d = 2, 3$
- Estimate β by OLS

Table: REJECTION FREQUENCIES $\hat{S}_{\hat{\gamma}}$ TEST FOR LSEM

n	K	d	1	2	3	4	5	6	7	8	9	10
500	2	2	4.2	4.8	5.3	5.6	3.9	3.1	5.7	4.9	3.5	4.2
500	2	3	5.2	4.5	4.4	4.8	4.5	1.9	4.9	5.3	3.9	4.4
500	3	2	4.9	5.2	5.2	4.8	4.1	2.2	5.2	4.1	3.9	4.1
500	3	3	5.0	5.0	5.1	4.8	4.7	3.0	6.6	3.6	4.3	3.6
500	5	2	6.4	7.2	7.1	7.2	5.6	2.3	5.2	4.4	3.9	4.3
500	5	3	7.2	7.6	7.4	7.9	6.2	2.0	5.4	4.2	4.1	4.2
1000	2	2	5.3	4.9	4.3	4.9	3.9	2.2	4.4	3.5	4.2	4.1
1000	2	3	4.7	4.6	5.2	5.0	4.4	1.4	5.2	4.3	4.4	3.2
1000	3	2	4.5	4.8	4.7	3.1	3.5	1.1	5.5	2.9	3.5	3.8
1000	3	3	4.4	4.4	4.1	4.1	3.6	2.1	6.3	3.9	4.1	3.4
1000	5	2	5.6	5.2	4.8	4.8	4.6	1.7	5.8	2.8	4.0	2.7
1000	5	3	5.5	6.1	5.1	5.4	4.2	2.0	5.1	3.4	4.2	3.3

Conclusions

- Independence and non-Gaussian distributions can provide information for identification
- Provide robust inference methods for non-Gaussian LSEMs with independent errors
- Simulation studies demonstrate that our asymptotic results seem to provide a good approximation to finite sample performance.

References I

Andrews, D. W. K. and Guggenberger, P. (2019).
Identification- and singularity-robust inference for moment condition models.
Quantitative Economics, 10(4):1703–1746.

Andrews, I. and Mikusheva, A. (2015).
Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models.
Quantitative Economics, 6.

Bekaert, G., Engstrom, E., and Ermolov, A. (2019).
Macro Risks and the Term Structure of Interest Rates.
Working paper.

Bekaert, G., Engstrom, E., and Ermolov, A. (2020).
Aggregate Demand and Aggregate Supply Effects of COVID-19: A Real-time Analysis.
Working paper.

Bickel, P. J., Klaasen, C. A. J., Ritov, Y., and Wellner, J. A. (1998).
Efficient and Adaptive Estimation for Semiparametric Models.
Springer, New York, NY, USA.

Bonhomme, S. and Robin, J.-M. (2009).
Consistent noisy independent component analysis.
Journal of Econometrics, 149.

Cardoso, J. and Souloumiac, A. (1994).
Blind beamforming for non gaussian signals.
Radar and Signal Processing, IEE Proceedings F, 140:362 – 370.

Chen, A. and Bickel, P. J. (2006).
Efficient Independent Component Analysis.
Annals of Statistics, 34(6).

References II

Comon, P. (1994).
Independent component analysis, A new concept?
Signal Processing, 36.

Dagenais, M. G. and Dagenais, D. L. (1997).
Higher moment estimators for linear regression models with errors in the variables.
Journal of Econometrics, 76(1-2):193–221.

Darmois, G. (1953).
Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire.
Revue de l'Institut International de Statistique / Review of the International Statistical Institute, 21(1/2):2–8.

Davis, R. and Ng, S. (2022).
Time Series Estimation of the Dynamic Effects of Disaster-Type Shocks.
Working paper.

Drautzburg, T. and Wright, J. H. (2021).
Refining set-identification in vars through independence.
Working Paper 29316, National Bureau of Economic Research.

Erickson, T. and Whited, T. M. (2002).
Two-step gmm estimation of the errors-in-variable model using higher order moments.
Econometric Theory, 18(3):776–799.

Fiorentini, G. and Sentana, E. (2020).
Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions.
working paper.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017).
Statistical inference for independent component analysis: Application to structural VAR models.
Journal of Econometrics, 196.

References III

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2019).
Identification and Estimation in Non-Fundamental Structural VARMA Models.
The Review of Economic Studies, 87(4):1915–1953.

Guay, A. (2020).
Identification of structural vector autoregressions through higher unconditional moments.
Journal of Econometrics.
forthcoming.

Herwartz, H. (2019).
Long-run neutrality of demand shocks: Revisiting blanchard and quah (1989) with independent structural shocks.
Journal of Applied Econometrics, 34(5):811–819.

Herwartz, H., Lange, A., and Maxand, S. (2019).
Statistical identification in svars - monte carlo experiments and a comparative assessment of the role of economic uncertainties for the us business cycle.
Discussion Paper 375, CEGE.

Hoesch, L., Lee, A., and Mesters, G. (2020).
Robust tests in structual VAR models identified by non-Gaussianity.
Working paper.

Ikeda, S. and Toyama, K. (2000).
Ica for noisy neurobiological data.
Proceedings of the IEEE-INNS-ENNS, 4:89 – 94.

Kapteyn, A. and Wansbeek, T. (1983).
Identification in the linear errors in variables model.
Econometrica, 51(6):1847–1849.

Kendall, M. G. and Stuart, A. (1970).
The Advanced Theory of Statistics (4th edition).
MacMillan.

References IV

Klein, R. and Vella, F. (2010).
Estimating a class of triangular simultaneous equations models without exclusion restrictions.
Journal of Econometrics, 154(2):154–164.

Lanne, M. and Luoto, J. (2019).
Gmm estimation of non-gaussian structural vector autoregression.
Journal of Business & Economic Statistics, 0(0):1–13.

Lanne, M. and Lütkepohl, H. (2010).
Structural vector autoregressions with nonnormal residuals.
Journal of Business & Economic Statistics, 28(1):159–168.

Lanne, M., Meitz, M., and Saikkonen, P. (2017).
Identification and estimation of non-Gaussian structual vector autoregressions.
Journal of Econometrics, 196.

Lee, A. (2022).
Robust and efficient inference for non-regular semiparametric models.
Working paper.

Lewbel, A. (2012).
Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models.
Journal of Business & Economic Statistics, 30(1):67–80.

Lewbel, A. (2019).
The identification zoo: Meanings of identification in econometrics.
Journal of Economic Literature, 57(4):835–903.

Lewbel, A., Schennach, S. M., and Zhang, L. (2021).
Identification of a triangular two equation system without instruments.
working paper.

References V

Lütkepohl, H. and Burda, M. M. (1997).
Modified wald tests under nonregular conditions.
Journal of Econometrics, 78(2):315–332.

Maxand, S. (2018).
Identification of independent structural shocks in the presence of multiple gaussian components.
Econometrics and Statistics.

Moneta, A., Entner, D., Hoyer, P. O., and Coad, A. (2013).
Causal inference by independent component analysis: Theory and applications*.
Oxford Bulletin of Economics and Statistics, 75(5):705–730.

Moneta, A. and Pallante, G. (2020).
Identification of Structural VAR Models via Independent Component Analysis: A Performance Evaluation Study.
LEM Papers Series 2020/24, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy.

Olea, J. L. M., Plagborg-Møller, M., and Qian, E. (2021).
Svar identification from higher moments: Has the simultaneous causality problem been solved?
Working paper.

Pal, M. (1980).
Consistent moment estimators of regression coefficients in the presence of errors in variables.
Journal of Econometrics, 14(3):349–364.

Schennach, S. M. (2021).
Measurement systems.
Journal of Economic Literature.
forthcoming.

Sims, C. A. (2021).
Svar identification through heteroskedasticity with misspecified regimes.
working paper.

References VI

Skitivic, V. P. (1953).
On a property of the normal distribution.
Dokl. Akad. Nauk SSSR (N.S.), 89:217–219.

Tank, A., Fox, E. B., and Shojaie, A. (2019).
Identifiability and estimation of structural vector autoregressive models for subsampled and mixed-frequency time series.
Biometrika, 106(2):433–452.

van der Vaart, A. W. (2002).
Semiparametric statistics.
In Bernard, P., editor, *Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXIX - 1999*. Springer, Berlin, Germany.

Velasco, C. (2020).
Identification and estimation of structural varma models using higher order dynamics.
working paper.

Parametric approach

$\eta = \eta(\beta)$ with β finite dimensional

- Orthogonalize scores $\dot{\ell}_\alpha(Y_i) = \nabla_\alpha \ell(Y_i)$ wrt scores $\dot{\ell}_\beta(Y_i) = \nabla_\beta \ell(Y_i)$, i.e. Neyman's $C(\alpha)$ test
- Consider

$$C(\alpha) = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\kappa}(Y_i) \right)' \hat{\mathcal{I}}^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\kappa}(Y_i) \right) ,$$

with

$$\hat{\kappa}(Y_i) = \dot{\ell}_\alpha - \hat{I}_{\alpha\beta} \hat{I}_{\beta\beta}^{-1} \dot{\ell}_\beta \quad \text{and} \quad \hat{\mathcal{I}} = \hat{I}_{\alpha\alpha} - \hat{I}_{\alpha\beta} \hat{I}_{\beta\beta}^{-1} \hat{I}_{\beta\alpha} ,$$

and $\hat{I} = \frac{1}{n} \sum_{i=1}^{n} \dot{\ell}(Y_i) \dot{\ell}(Y_i)'$.

- [Andrews and Mikusheva, 2015] show $C(\alpha)$ tests retain correct size regardless whether α is identified

back

Truncation idea

Suppose that $I_n \to I$ are deterministic $L \times L$ matrices with $\text{rank}(I_n) = \text{rank}(I)$ for $n \geq N$ and for $\hat{I}_n$ and $0 \leq \nu_n \to 0$

$$\|\hat{I}_n - I_n\|_2 = o_{p_n}(\nu_n) \; .$$

Let $\hat{I}_n^t$ be the ν_n eigenvalue truncated version of $\hat{I}_n$

$$\hat{I}_n^t = \hat{U}\hat{\Lambda}(\nu_n)\hat{U}' \qquad \hat{\Lambda}(\nu_n) = \text{diag}(\{\lambda_{ii} 1(\lambda_{ii} \geq \nu_n)\}_{i=1}^n)$$

Then $\hat{I}_n \stackrel{p_n}{\to} I$ and

$$\lim_{n\to\infty} P_n(\text{rank}(\hat{I}_n) = \text{rank}(I)) = 1$$

see [Andrews and Guggenberger, 2019] for a similar construction.