Robust Non-Gaussian Inference for Linear Simultaneous Equations Models

Adam Lee & Geert Mesters (UPF & BSE)

October 12, 2021

Introduction

- Exploiting non-Gaussian distributions is an attractive way to (partially) identify structural parameters [Peters et al., 2018]
- Conducting inference on such parameters without assuming the distributions are indeed non-Gaussian is harder

Simple example

Consider

$$Y = A^{-1}\epsilon,$$

where

- Y is a $K \times 1$ vector
- A is a $K \times K$ invertible matrix
- the K components of ϵ are independent
- $\mathbb{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = I$

Goal: recover the matrix A

Identification

$$Y = A^{-1}\epsilon$$

- if ϵ Gaussian A is identified up to orthogonal transformation

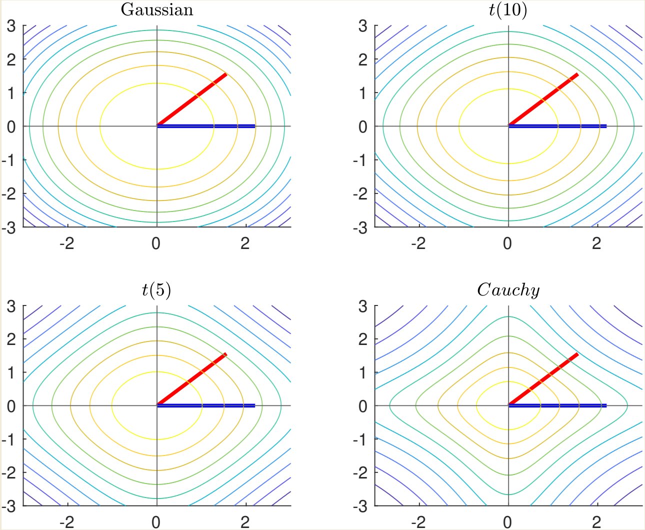
$$\{A^{-1}Q : Q \in \mathrm{O}(K)\}$$

- if *at most* one element of ϵ is Gaussian A is identified up to permutation and sign [Comon, 1994].

$$\{A^{-1}PD : P \in \Pi(K), D \in \mathrm{S}(K)\}$$

$\Rightarrow$ non-Gaussianity shrinks the identified set

Identification



Red and blue lines represent different $Q(\theta)$, θ angle

Standard inference approach

$$Y = A^{-1}\epsilon$$

(i) **Assume ϵ is non-Gaussian**

(ii) Estimate A

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

(iii) Confidence intervals based on sampling variation of estimator

Standard inference approach

$$Y = A^{-1}\epsilon$$

(i) **Assume ϵ is non-Gaussian**

(ii) Estimate A

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

(iii) Confidence intervals based on sampling variation of estimator

Key problem :

- size distortions when ϵ is "close" to Gaussian

Table: EMPIRICAL REJECTION FREQUENCIES $a = 0.05$

Test	$\mathcal{N}$	$t(15)$	$t(10)$	$t(5)$	SU	KU	OT	BI
W	0.25	0.23	0.18	0.09	0.28	0.07	0.02	0.17
LM	0.09	0.09	0.07	0.06	0.10	0.06	0.06	0.06
LR	0.01	0.03	0.04	0.05	0.04	0.03	0.02	0.000
W^G	0.23	0.09	0.05	0.01	0.03	0.02	0.03	0.98
LR^L	0.16	0.14	0.10	0.09	0.14	0.16	0.34	0.11

Linear simultaneous equations models (LSEM)

Perhaps more economically relevant

$$Z = BX + Y \qquad Y = A^{-1}\epsilon$$

where Z and X are observable variables

- If object of interest is A or $f(A, B)$ size distortions carry over when ϵ is "close" to Gaussian

Analogy to weak IV

The following analogy is perhaps useful

	IV
Theoretical assumption	rank condition
Relevant criteria	F-statistic
Robust inference	AR, LM, CLR

Analogy to weak IV

The following analogy is perhaps useful

	IV	**LSEM**
Theoretical assumption	rank condition	non-Gaussianity
Relevant criteria	F-statistic	distance to Gaussianity
Robust inference	AR, LM, CLR	This paper

This paper

- Consider $Y = A^{-1}\epsilon$ as **semi-parametric** model
  - α, β in $A = A(\alpha, \beta)$ parametric
  - densities ϵ non-parametric
- 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
- Extend tests to cover broader class $Z = BX + A^{-1}\epsilon$

Related literature - a very brief summary

- Identification robust testing: [Stock and Wright, 2000, Kleibergen, 2005, Andrews and Mikusheva, 2016]
- Semi-parametric inference: [Bickel et al., 1998, Choi et al., 1996, van der Vaart, 2002]
- ICA model: [Amari and Cardoso, 1997, Hyvärinen et al., 2001, Bach and Jordan, 2002, Chen and Bickel, 2006]
- non-Gaussian SVAR: [Hyvärinen et al., 2010, Moneta et al., 2013, Lanne et al., 2017, Maxand, 2018, Lanne and Luoto, 2019, Gouriéroux et al., 2017, Gouriéroux et al., 2019, Bekaert et al., 2019, Bekaert et al., 2020]

Road map

(i) High level general framework

(ii) Implementation details for $Y = A^{-1}\epsilon$

(iii) Simulation results

(i) Semi-parametric models and identification robust testing

Semiparametric model

Suppose we observe random vectors Y_i on $\mathbb{R}^K$ which are distributed according to the law P_{θ_0} where $\theta_0 \in \Theta$ with

$$\Theta = \mathcal{A} \times \mathcal{B} \times \mathcal{H} = \{\theta = (\alpha, \beta, \eta) : \alpha \in \mathcal{A},\ \beta \in \mathcal{B},\ \eta \in \mathcal{H}\}$$

where $\mathcal{A} \subset \mathbb{R}^{L_\alpha}$, $\mathcal{B} \subset \mathbb{R}^{L_\beta}$ and $\mathcal{H}$ is metric space.

The *model* is the collection

$$\mathcal{P}_\Theta := \{P_\theta : \theta \in \Theta\}.$$

Hypothesis testing

The goal is to test

$$H_0 : \alpha = \alpha_0 \, , \beta \in \mathcal{B} \, , \eta \in \mathcal{H} \qquad \text{vs} \qquad H_1 : \alpha \neq \alpha_0 \, , \beta \in \mathcal{B}, \; \eta \in \mathcal{H}$$

$\Rightarrow$ Without assuming that α is identified

High level overview

(i) Construct efficient score function for α

(ii) Compute score statistic under $H_0 : \alpha = \alpha_0$

$\Rightarrow$ Intuition: semiparametric generalization of Neyman's C(α) test

Efficient score function

The efficient score function for $\gamma = (\alpha, \beta)$ is defined as

$$\tilde{\ell}_\theta = \dot{\ell}_\theta - \Pi\left(\dot{\ell}_\theta \middle| \operatorname{cl} \mathcal{T}^{\eta}_{P_\theta, H}\right)$$

where

- $\dot{\ell}_\theta = \frac{\partial}{\partial\gamma} \log \mathrm{d}P_\theta : (L_\alpha + L_\beta) \times 1$ score vector of γ
- $\Pi\left(\dot{\ell}_\theta \middle| \operatorname{cl} \mathcal{T}^{\eta}_{P_\theta, H}\right)$ projection of $\dot{\ell}_\theta$ on the closure of $\mathcal{T}^{\eta}_{P_\theta, H}$
- $\mathcal{T}^{\eta}_{P_\theta, H}$ tangent space of η: collection of scores with respect to η

Efficient information matrix

Given the efficient score function

$$\tilde{\ell}_\theta = \dot{\ell}_\theta - \Pi\left(\dot{\ell}_\theta \middle| \operatorname{cl} \mathcal{T}^{\eta}_{P_\theta, H}\right)$$

The efficient information matrix becomes

$$\tilde{I}_\theta = \mathbb{E}\tilde{\ell}_\theta \tilde{\ell}'_\theta$$

Intuitively,

- $\tilde{\ell}_\theta$ is population residual of regressing $\dot{\ell}_\theta$ on scores of η
- $\tilde{I}_\theta$ is variance of this residual

Efficient score function for α

Note that

$$\tilde{\ell}_\theta = \left(\tilde{\ell}'_{\theta,\alpha}, \tilde{\ell}'_{\theta,\beta}\right)' \quad \text{and} \quad \tilde{I}_\theta = \begin{bmatrix} \tilde{I}_{\theta,\alpha\alpha} & \tilde{I}_{\theta,\alpha\beta} \\ \tilde{I}_{\theta,\beta\alpha} & \tilde{I}_{\theta,\beta\beta} \end{bmatrix} .$$

Efficient score function for α:

$$\tilde{\kappa}_\theta := \tilde{\ell}_{\theta,\alpha} - \tilde{I}_{\theta,\alpha\beta} \tilde{I}^{-1}_{\theta,\beta\beta} \tilde{\ell}_{\theta,\beta} ,$$

corresponding efficient information matrix

$$\tilde{\mathcal{I}}_\theta := \tilde{I}_{\theta,\alpha\alpha} - \tilde{I}_{\theta,\alpha\beta} \tilde{I}^{-1}_{\theta,\beta\beta} \tilde{I}_{\theta,\beta\alpha} .$$

- Importantly, $\tilde{\mathcal{I}}_\theta$ can be singular
- e.g. when ϵ is exactly Gaussian in $Y = A^{-1}\epsilon$

Efficient score test

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

where

$$\hat{\kappa}_{\theta} = \hat{\ell}_{\theta,\alpha} - \hat{I}_{\theta,\alpha\beta} \hat{I}_{\theta,\beta\beta}^{-1} \hat{\ell}_{\theta,\beta} \;, \qquad \text{and} \qquad \hat{\mathcal{I}}_{\theta} = \hat{I}_{\theta,\alpha\alpha} - \hat{I}_{\theta,\alpha\beta} \hat{I}_{\theta,\beta\beta}^{-1} \hat{I}_{\theta,\beta\alpha}$$

- $\hat{\ell}_{\theta_0}, \hat{I}_{\theta_0}$ estimates for $\tilde{\ell}_{\theta_0}, \tilde{I}_{\theta_0}$
- $\hat{I}_{\theta}^{t} = \hat{U} \hat{\Lambda}(\nu_n) \hat{U}'$ with $\hat{\Lambda}(\nu_n)$ diagonal matrix of ν_n-truncated eigenvalues of $\hat{I}_{\theta}$ details

High level assumptions

Let $\gamma_n = \{(\alpha_0, \beta_n)\}_{n\in\mathbb{N}}$ be a deterministic sequence such that $\sqrt{n}(\gamma_n - \gamma_0) = O(1)$ and define $\theta_n = (\gamma_n, \eta)$ for each $n \in \mathbb{N}$.

(i) $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\ell}_{\theta_0}(Y_i) \rightsquigarrow Z \sim \mathcal{N}(0, \tilde{I}_{\theta_0})$

(ii) $\frac{1}{n} \sum_{i=1}^{n} \left(\hat{\ell}_{\theta_n}(Y_i) - \tilde{\ell}_{\theta_n}(Y_i)\right) = o_{P_{\theta_n}}(n^{-1/2})$

(iii) $\|\hat{I}_{\theta_n} - \tilde{I}_{\theta_0}\|_2 = o_{P_{\theta_n}}(\nu_n)$ for $\nu_n \to 0$

(iv) $\int \left\|\tilde{\ell}_{\theta_n} p_{\theta_n}^{1/2} - \tilde{\ell}_{\theta_0} p_{\theta_0}^{1/2}\right\|^2 \mathrm{d}\mu \to 0.$

Main result

Let $\theta_0 = (\alpha_0, \beta, \eta)$ for any $(\beta, \eta) \in \mathcal{B} \times \mathcal{H}$. Suppose that $\hat{\beta}_n$ is a $\sqrt{n}$-consistent estimator of β under P_{θ_0}. Let $B_n = n^{-1/2} C \mathbb{Z}^{L_\beta}$ for some $C > 0$ and let $\bar{\beta}_n$ be a discretised version of $\hat{\beta}_n$ which replaces its value with the closest point in B_n. Suppose assumptions (i)-(iv) hold and let $\bar{\theta}_n = (\alpha_0, \bar{\beta}_n, \eta)$. Let $r_n = \mathrm{rank}(\hat{\mathcal{I}}^t_{\bar{\theta}_n})$ and denote by c_n the $1 - a$ quantile of the $\chi^2_{r_n}$ distribution for any $a \in (0, 1)$.[1] Then

$$\lim_{n \to \infty} P_{\theta_0} \left(\hat{S}^{SR}_{\bar{\theta}_n} > c_n \right) \leq a,$$

with inequality only if $\mathrm{rank}(\tilde{\mathcal{I}}_{\theta_0}) = 0$.

[1] If $r_n = 0$ we take $c_n = 0$.

Remarks

- Intuitively,
  - Fixing $\alpha = \alpha_0$ avoids need for identification
  - Orthogonalizing allow to handle infinite dimensional η
- Confidence regions by test inversion
- Optimal under non-singularity [Choi et al., 1996]
- Discretizing $\hat{\beta}_n$ simplifies verifying (i)-(iv)
- Can use any $\sqrt{n}$-consistent estimate for β, but one-step efficient estimates increase finite sample power

$$\hat{\beta} = \hat{\beta}^{\text{init}} + \hat{I}_{\theta,\beta\beta}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\ell}_{\theta,\beta}(Y_i)$$

More generally

- Efficient score test provides general approach to robust inference with **infinite dimensional** nuisance parameters
- Efficient score test can handle
  - (i) Identification problems [Stock and Wright, 2000]
  - (ii) Boundary problems [Andrews, 2001]
  - (iii) Regularization problems [Chernozhukov et al., 2015]
- Uniform size control over local perturbations and always minimax optimal.
- Trivial to implement for many models with χ^2 critical values

$\Rightarrow$ see Lee (2021):
Robust and efficient inference for non-regular semiparametric models

(ii) Robust non-Gaussian inference

Primitive assumptions

$$Y = A^{-1}\epsilon$$

Assumption 1

Each component ϵ_k in $\epsilon = (\epsilon_1, \ldots, \epsilon_K)'$ has a continuously differentiable density function, which we write as η_k with log density score $\phi_k(x) = \partial \log \eta_k(x)/\partial x$. We assume that for all $k = 1, \ldots, K$

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

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

(iii) ϵ_k is independent of ϵ_j for all $k \neq j$

Semi-parametric formulation

$$Y = A^{-1}\epsilon\,, \qquad \epsilon = (\epsilon_1, \ldots, \epsilon_K)\,, \qquad \epsilon_k \sim \eta_k$$

- $\gamma = (\alpha, \beta)$ is $L \times 1$ and models $A(\gamma)$
- $\mathcal{H} = \{\eta = (\eta_1, \ldots, \eta_K) : \eta_k \in \mathscr{H}\}$, with

$$\mathscr{H} = \left\{ g : g \in \mathcal{C}^1(\lambda), \text{satisfies assumption 1} \right\}$$

- P_θ law defined by density

$$p_\theta(y) = |\det A| \prod_{k=1}^{K} \eta_k(A_{k\bullet} y)$$

Efficient score function

Proposition

Given Assumption 1, if $\gamma \mapsto A(\gamma)$ is continuously differentiable, for $l = 1, \ldots, L$,

$$\tilde{\ell}_{\theta,l}(y) = \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta_{l,k,j}\phi_k(A_{k\bullet}y)A_{j\bullet}y + \sum_{k=1}^{K} \zeta_{l,k,k}\left[\tau_{k,1}A_{k\bullet}y + \tau_{k,2}\kappa(A_{k\bullet}y)\right]$$

where $\zeta_{l,k,j} := [D_l(\gamma)]_{k\bullet}A^{-1}_{\bullet j}$ with $D_l(\gamma) = \partial A(\gamma)/\partial\gamma_l$, $A^{-1}_{\bullet j}$ is the j-th column of $A(\gamma)^{-1}$ and

$$\tau_k := M_k^{-1}\begin{pmatrix} 0 \\ -2 \end{pmatrix}, \quad \text{where } M_k := \begin{pmatrix} 1 & \mathbb{E}_\theta(A_{k\bullet}y)^3 \\ \mathbb{E}_\theta(A_{k\bullet}y)^3 & \mathbb{E}_\theta(A_{k\bullet}y)^4 - 1 \end{pmatrix}.$$

Estimating the efficient score function

To estimate we need

- τ_k, use its sample analog:

$$\hat{\tau}_{k,n} = \begin{pmatrix} 1 & \frac{1}{n}\sum_{i=1}^{n}(A_{k\bullet}Y_i)^3 \\ \frac{1}{n}\sum_{i=1}^{n}(A_{k\bullet}Y_i)^3 & \frac{1}{n}\sum_{i=1}^{n}(A_{k\bullet}Y_i)^4 - 1 \end{pmatrix} \begin{pmatrix} 0 \\ -2 \end{pmatrix}.$$

- $\phi_k = \partial \log \eta_k(x)/\partial x$: requires log density score estimator
  - We follow [Jin, 1992] and [Chen and Bickel, 2006] and use B-spline based estimator
  - Show in the paper that this estimator allows to consistently estimate the efficient score under mild assumptions on η_k

Efficient score function estimate

Given estimates $\hat{\tau}_{k,n}$ and $\hat{\phi}_{k,n}$ we compute

$$\hat{\ell}_{\theta,l,n}(y) = \sum_{k=1}^{K} \sum_{j \neq k}^{K} \zeta_{l,k,j} \hat{\phi}_{k,n}(A_{k\bullet} y) A_{j\bullet} y + \sum_{k=1}^{K} \zeta_{l,k,k} \left[\hat{\tau}_{k,n,1} A_{k\bullet} y + \hat{\tau}_{k,n,2} \kappa(A_{k\bullet} y) \right]$$

for $l = 1, \ldots, L$ and

$$\hat{I}_{\theta_0,n} = \frac{1}{n} \sum_{i=1}^{n} \hat{\ell}_{\theta,l,n}(Y_i) \hat{\ell}_{\theta,l,n}(Y_i)' \qquad \hat{I}^t_{\theta_0,n} = (\nu_n)\text{- eigenvalue truncated}$$

Main result

Let $\hat{\beta}_n$ be a $\sqrt{n}-$consistent estimator of β and let $\bar{\beta}_n$ denote a discretised version. Define $\bar{\theta}_n = (\alpha_0, \bar{\beta}_n, \eta)$ and consider the statistic

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

where $\hat{\mathcal{I}}^{t}_{\bar{\theta}_n}$ is truncated at $\nu_n = \nu^2_{n,p}$ 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 $p = 2$. Then

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

Remarks

- Size control does not require a non-Gaussian assumption
- Implementation requires
    - K regressions to fit B-splines for log density score
    - and computing some sums ...

(iii) Simulation results

Simulation design

(i) Consider

$$Y_i = A^{-1}\epsilon_i \ , \qquad A = \begin{bmatrix} \beta_1 & 0 \\ \beta_2 & \beta_3 \end{bmatrix} \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

(ii) Consider various densities for $\epsilon_{i,k}$

Simulation design

Table: True Error Distributions ϵ_i

	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"

Simulation results: size

Table: EMPIRICAL REJECTION FREQUENCIES $\hat{S}_n^{SR}$ FOR BASELINE ICA

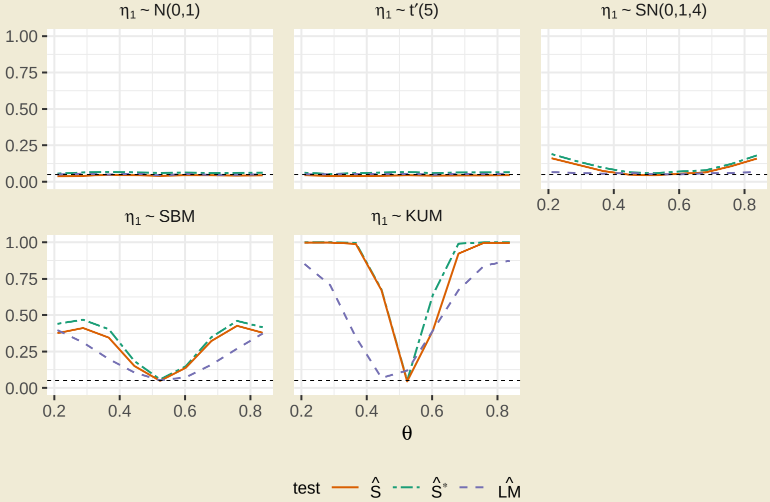
K	B	$\mathcal{N}$	$t(15)$	$t(10)$	$t(5)$	SU	KU	OT	BI
2	4	0.041	0.047	0.038	0.043	0.047	0.051	0.047	0.052
2	6	0.045	0.043	0.042	0.044	0.045	0.054	0.047	0.053
2	8	0.046	0.047	0.047	0.046	0.043	0.051	0.046	0.050
3	4	0.031	0.040	0.037	0.037	0.043	0.047	0.041	0.047
3	6	0.038	0.042	0.038	0.037	0.045	0.046	0.044	0.042
3	8	0.041	0.046	0.040	0.042	0.048	0.047	0.043	0.044

Simulation results: comparison

Table: Empirical Rejection Frequencies $\alpha = 0.05$

Test	$\mathcal{N}$	$t(15)$	$t(10)$	$t(5)$	SU	KU	OT	BI
$\hat{S}^{SR}_{\bar{\theta}_n}$	0.04	0.04	0.05	0.04	0.05	0.05	0.05	0.05
W	0.25	0.23	0.18	0.09	0.28	0.07	0.02	0.17
LM	0.09	0.09	0.07	0.06	0.10	0.06	0.06	0.06
LR	0.01	0.03	0.04	0.05	0.04	0.03	0.02	0.000
W^G	0.23	0.09	0.05	0.01	0.03	0.02	0.03	0.98
LR^L	0.16	0.14	0.10	0.09	0.14	0.16	0.34	0.11

Simulation results: power



Conclusions

- We provide a general framework for conducting inference on the Euclidean parameter in semiparametric models without assuming this parameter is identified.
- Specialising to the LSEMs we provide full details of the theory and implementation of this procedure.
- Simulation studies demonstrate that our asymptotic results seem to provide a good approximation to finite sample performance.

References I

Amari, S. and Cardoso, J.-F. (1997).
Blind Source Separation - Semiparametric Statistical Approach.
IEEE Transactions On Signal Processing, 45(11).

Andrews, D. W. K. (2001).
Testing when a parameter is on the boundary of the maintained hypothesis.
Econometrica, 69(3):683–734.

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. (2016).
Conditional inference with a functional nuisance parameter.
Econometrica, 84(4).

Bach, F. R. and Jordan, M. I. (2002).
Kernel independent component analysis.
Journal of Machine Learning Research, 3:1–48.

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.

References II

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

Chernozhukov, V., Hansen, C., and Spindler, M. (2015).
Valid post-selection and post-regularization inference: An elementary, general approach.
Annual Review of Economics, 7(1):649–688.

Choi, S., Hall, W. J., and Schick, A. (1996).
Asymptotically uniformly most powerful tests in parametric and semiparametric models.
Ann. Statist., 24(2):841–861.

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

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

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.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001).
Independent Component Analysis.
John Wiley & Sons, Inc.

References III

Hyvärinen, A., Zhang, K., Shimizu, S., and Hoyer, P. O. (2010).
Estimation of a structural vector autoregression model using non-gaussianity.
Journal of Machine Learning Research, 11(56):1709–1731.

Jin, K. (1992).
Empirical Smoothing Parameter Selection In Adaptive Estimation.
Annals of Statistics, 20(4).

Kleibergen, F. (2005).
Testing parameters in GMM without assuming that they are identified.
Econometrica, 73(4).

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

Lee, A., Mesters, G., and Zwiernik, P. (2021).
Non-independent components analysis.

Matteson, D. S. and Tsay, R. S. (2017).
Independent component analysis via distance covariance.
Journal of the American Statistical Association, 112(518):623–637.

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

References IV

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.

Peters, J., Janzing, D., and Schölkopf, B. (2018).
Elements of Causal Inference: Foundations and Learning Algorithms.
MIT press.

Stock, J. H. and Wright, J. H. (2000).
GMM with weak identification.
Econometrica, 68(5).

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.

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 $\hat{I}_n \xrightarrow{p_n} I$ 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.

back