Robust and efficient inference for non-regular semiparametric models

EEA 2022, Università Bocconi

Adam Lee

22 August 2022

Motivation

Motivation

- Estimation theory in (semi-)parametric models is developed under regularity conditions
- A ubiquitous condition is non-singularity of the (efficient) information matrix, $\tilde{\mathcal{I}}$
  - ▷ In the parametric case, this non-singularity $\approx$ (local) identification
  - ▷ In the semiparametric case, the connection is less direct: singularity may imply failure of (local) identification or irregular (local) identification (Escanciano, 2022)
- Singularity of $\tilde{\mathcal{I}} \approx$ non-existence of regular estimators (Chamberlain, 1986)
  - ▷ Regularity requires estimators to converge (weakly) in a locally uniform manner
  - ▷ If we are "too close" to a point where $\tilde{\mathcal{I}}$ is singular, this breaks down
  - ▷ Hypothesis tests based on these estimators perform poorly
    - ■ E.g. weak identification
- Today: we can still construct tests with desirable properties

Example: Single-index model

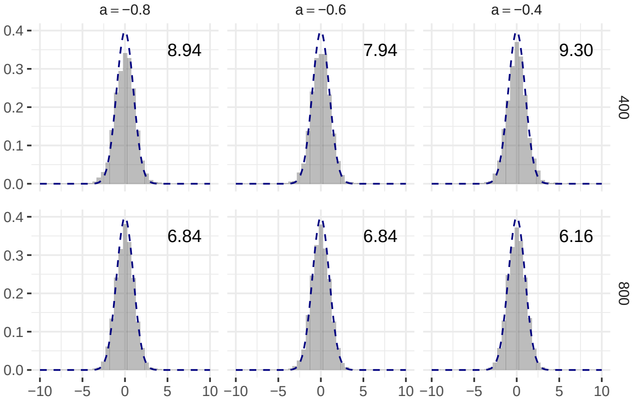
- Observe i.i.d. samples of $(Y, X_1, X_2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^K$:

$$Y = f\left(X_1 + \theta' X_2\right) + \epsilon, \qquad \mathbb{E}\left[\epsilon | X\right] = 0.$$

- $+$ moment and smoothness conditions
- Goal: test $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$
  - ▷ If $f' = 0$, then efficient information is singular (identically 0)
    - ■ In this case, θ is not identified (Ichimura, 1993; Horowitz, 2009)
  - ▷ If f' is **close to** 0, θ is **weakly** identified
    - ■ Standard asymptotic approximations are poor
    - ■ As is typical in cases of weak identification, size distortions occur

t-statistics in the single index model

Figure 1: t-statistics, $f(v) = b\left(a + \frac{1-a}{1+\exp(-v)}\right)$

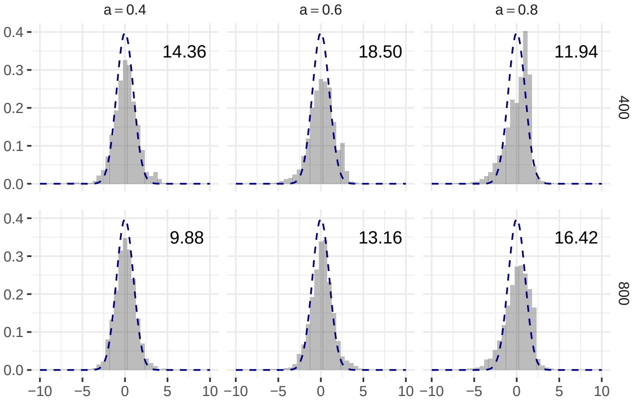


Based on 5000 Monte Carlo replications. t-statistic computed from Ichimura (1993)-style estimate of θ.
$\theta = 1$, $\epsilon \sim \mathcal{N}(0,1)$, $X \sim U[-1,1]$, $b = 10$.

▸ t-test details ▸ f plot ▸ f' plot

t-statistics in the single index model

Figure 2: t-statistics, $f(v) = b\left(a + \frac{1-a}{1+\exp(-v)}\right)$



Based on 5000 Monte Carlo replications. *t*-statistic computed from Ichimura (1993)-style estimate of θ.
$\theta = 1$, $\epsilon \sim \mathcal{N}(0,1)$, $X \sim U[-1,1]$, $b = 10$.

▸ t-test details ▸ f plot ▸ f' plot

This paper

1. A class of generalised C(α) statistics converge (weakly) locally uniformly
   - Including when $\tilde{\mathcal{I}}$ is singular
   - C(α) tests based on these statistics control size (locally uniformly)
   - Class includes the efficient score function
2. C(α) based on the efficient score function is minimax optimal whenever rank($\tilde{\mathcal{I}}$) > 0.
   - Includes both the classical case of full rank $\tilde{\mathcal{I}}$ and when it is rank deficient
   - Latter case applies in e.g. models underidentified for certain nuisance parameter values
3. Two examples covered in detail:
   - Single index model
   - Independent components analysis

▸ Related literature

Testing approach

Setup & score functions

- Semiparametric likelihood model for (i.i.d.) observations of $W \in \mathcal{W}$
- Model: $\mathcal{P} = \{P_\gamma : \gamma \in \Gamma\}$, with corresponding densities $p_\gamma \ll \nu$
- Parameters: $\gamma = (\theta, \eta) \in \Gamma = \Theta \times \mathcal{H}$, $\Theta \subset \mathbb{R}^{d_\theta}$ open, $\mathcal{H}$ a metric space
  - ▷ θ: parameter of interest; η: nuisance parameter
- Analogously to the parametric case, we can define score functions for all parameters

$$\underbrace{\dot{\ell}_\gamma = \nabla_\theta \log p_\gamma}_{\text{vector of score functions for } \theta} \qquad \underbrace{\{B_\gamma h : h \in H_\eta\}}_{\text{collection of score functions for } \eta}$$

  - ▷ Each $B_\gamma h$ is a pathwise directional derivative of $\eta \mapsto \log p_\gamma$ in direction $h \in H_\eta$

▸ Some more technical details

Generalised C(α)-statistics

- There exist functions $g_\gamma : \mathcal{W} \to \mathbb{R}^{d_\theta}$ such that $\mathbb{E}_\gamma\, g_\gamma(W) = 0$ and

$$\mathbb{E}_\gamma \left[g_\gamma(W) \, [B_\gamma h](W) \right] = 0 \qquad \text{for all} \quad h \in H_\eta. \tag{1}$$

  - ▷ If $f_\gamma : \mathcal{W} \to \mathbb{R}^{d_\theta}$ is square-integrable & mean-zero, its orthogonal projection onto the orthogonal complement of $\overline{\text{lin}}\, \{B_\gamma h : h \in H_\eta\} \subset L_2(P_\gamma)$ satisfies (1).

- Under local asymptotic normality,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_\gamma(W_i) \overset{P_{n,\tau_n,h_n}}{\rightsquigarrow} \mathcal{N}\left(\mathbb{E}_\gamma \left[g_\gamma \dot{\ell}'_\gamma \right] \tau,\ V_\gamma \right), \quad V_\gamma := \text{Var}_\gamma(g_\gamma(W_i)),$$

where P_{n,τ_n,h_n} is a contiguous perturbation of P_γ in direction $(\tau_n, h_n) \to (\tau, h)$.

  - ▷ This holds regardless of the rank of V_γ
  - ▷ h does not appear in the limit because of (1)
    - ■ Consequence of Le Cam's third lemma

Generalised C(α)-statistic

- g_γ and V_γ will generally be unknown, so replace with estimates $\hat{g}_{n,\theta}$, $\hat{V}_{n,\theta}$,

$$\hat{S}_{n,\theta} := \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{g}_{n,\theta}(W_i) \right)' \hat{V}_{n,\theta}^{\dagger} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{g}_{n,\theta}(W_i) \right).$$

- Provided $\hat{g}_{n,\theta}$, $\hat{V}_{n,\theta}^{\dagger}$ are sufficiently good estimators, by the previous convergence result,

$$\hat{S}_{n,\theta} \overset{P_{n,0,h_n}}{\rightsquigarrow} \chi^2_r, \qquad \text{for } r = \operatorname{rank}(V_\gamma).$$

  - ▷ This requires $\hat{V}_{n,\theta}^{\dagger} \xrightarrow{P} V_\gamma^{\dagger}$, the Moore – Penrose inverse of V_γ
  - ▷ Can be achieved by truncating the eigenvalues of an initial estimate of V_γ at a certain rate (Dufour and Valéry, 2016)

▸ Eigenvalue truncation

Generalised C(α) test

- We can use this result to form a C(α) – style test
  - Score type test $\implies$ everything evaluated / estimated under the null hypothesis
  - Circumvents the issue that θ may not be consistently estimable (e.g. weak identification)
- Reject if $\hat{S}_{n,\theta_0}$ exceeds $1-\alpha$ quantile of $\chi^2_{\hat{r}_n}$ where $\hat{r}_n \xrightarrow{P} r$.
  - Same eigenvalue truncation idea yields a $\hat{r}_n$ with this property
- This test is correctly sized (locally uniformly)
  - Follows from the previous convergence result
- If $g_\gamma = \tilde{\ell}_\gamma$ the resulting test is minimax optimal whenever $r > 0$ & H_η is a linear space.
  - $\tilde{\ell}_\gamma := \dot{\ell}_\gamma - \Pi\left(\dot{\ell}_\gamma \middle| \overline{\text{lin}}\,\{B_\gamma h : h \in H_\eta\}\right)$
  - However, $\tilde{\ell}_\gamma$ can be difficult to derive / estimate in some models
  - Alternative (inefficient) choices may be much simpler than $\tilde{\ell}_\gamma$ in practice

▸ Size result ▸ Power result

Single index model

Single index model

- Consider the single index model:

$$Y = f(X_1 + X_2'\theta) + \epsilon, \quad \mathbb{E}[\epsilon|X] = 0, \quad f \in \mathscr{F}$$

- $\mathscr{F}$: functions bounded and continuously differentiable with bounded derivative a.e. on $\mathscr{D} := \{x_1 + x_2'\theta : \theta \in \Theta, x \in \text{supp}(X)\}$.
- + Regularity conditions: smoothness, finite moments etc.
- Notation:

$$V_\theta := X_1 + X_2'\theta, \qquad \sigma^2 := \mathbb{E}[\epsilon^2].$$

▸ Regularity conditions ▸ Example: Independent components analysis

Single index model

- We will base our statistic on the function

$$g_\gamma(W) := \sigma^{-2}(Y - f(V_\theta))f'(V_\theta)\left[X_2 - \mathbb{E}\left[X_2 | V_\theta\right]\right]$$

  - ▷ The efficient score function in this model was derived by Newey and Stoker (1993):

$$\tilde{\ell}_\gamma = \omega(X)(Y - f(V_\theta))f'(V_\theta)\left[X_2 - \frac{\mathbb{E}\left[\omega(X)X_2 | V_\theta\right]}{\mathbb{E}\left[\omega(X) | V_\theta\right]}\right], \quad \omega(X) := \mathbb{E}[\epsilon^2 | X]^{-1}.$$

  - ▷ g_γ is a special case of $\tilde{\ell}_\gamma$ corresponding to $\omega(X) = \sigma^{-2}$, i.e. conditional homoskedasticity
- g_γ satisfies our key uncorrelatedness condition (1)
- We need to estimate g_γ and V_γ

Single index model

- Our moment condition

$$g_\gamma(W) := \sigma^{-2}(Y - f(V_\theta))f'(V_\theta)\left[X_2 - \mathbb{E}\left[X_2|V_\theta\right]\right]$$

- Unknowns: σ^2, f, f' and Z where $Z(V) := \mathbb{E}[X_2|V]$.
- Given an estimate $\hat{f}_n$ of f, put

$$\hat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{f}_n(V_{\theta,i}))^2.$$

- For the nonparametric parts, a wide range of non-parametric estimators are possible
  - ▷ Theory given for generic split – sample estimators in paper

▸ Split sample estimators ▸ Example: Series estimators ▸ Nonparametric estimation conditions

Single index model: Monte Carlo

- Draw 5000 Monte Carlo replications from

$$Y = f(X_1 + X_2\theta) + \epsilon, \quad \mathbb{E}[\epsilon|X] = 0,$$

where $f \in \{f_1, f_2, f_3\}$,

$$f_1(v) \propto a + \frac{1-a}{1+\exp(-v)}, \quad f_2(v) \propto \exp\left(\frac{-v^2}{2a^2}\right), \quad f_3(v) = av^2,$$

for various choices of a.

- Test the null $H_0 : \theta = 1$
- Compare size of C(α) – style test and Wald test based on estimator of Ichimura (1993)
- When $a \approx a_\star$, $f'(v) \approx 0 \implies \tilde{\mathcal{I}}$ is close to singularity

▸ Wald test details

Single index model: Monte Carlo

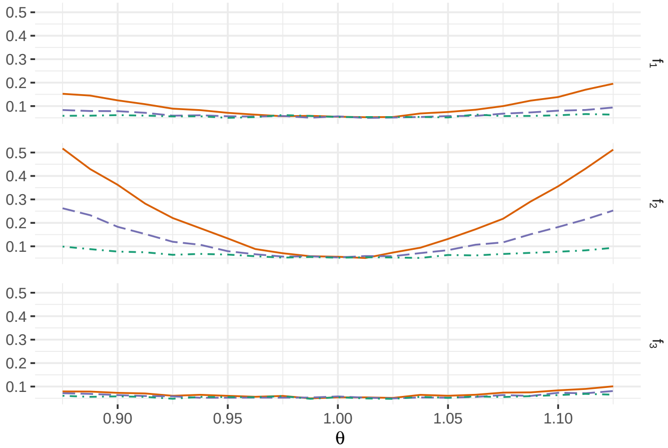
Table 1: Empirical rejection frequencies (%), $\epsilon \sim \mathcal{N}(0,1)$, $X_i \sim U[-1,1]$, $\theta = 1$

	$f_1(v) \propto a + (1-a)/(1+\exp(-v))$			$f_2(v) \propto \exp(-v^2/(2a^2))$			$f_3(v) = av^2$		
n	$a = 0.4$	$a = 0.6$	$a = 0.8$	$a = 3$	$a = 4$	$a = 5$	$a = 0.45$	$a = 0.35$	$a = 0.25$
$\hat{S}$									
400	5.18	4.84	5.48	5.92	5.42	5.26	5.78	5.60	5.50
600	4.82	5.68	5.32	5.68	5.28	5.88	5.56	5.64	5.76
800	5.34	4.96	5.44	5.44	4.84	5.68	5.44	4.66	4.76
1000	5.08	5.84	5.16	5.36	5.62	4.98	5.56	4.94	4.98
W									
400	14.08	18.68	12.22	21.54	19.96	17.86	21.38	18.36	17.54
600	11.32	14.40	13.76	18.30	21.40	16.44	23.30	19.90	18.12
800	10.82	12.90	15.88	15.14	21.34	19.28	21.32	21.34	19.42
1000	9.86	12.36	17.60	14.72	20.34	20.26	21.70	22.28	18.18

Based on 5000 Monte Carlo replications. $\hat{S}$ is the C(α) – style test.

Single index model: Monte Carlo

Figure 3: Empirical rejection frequency, $\epsilon \sim \mathcal{N}(0,1)$, $X_i \sim U[-1,1]$



$f_1(v) =$ generalised logistic, $f_2(v) =$ Gaussian function, $f_3(v) =$ quadratic

◂ Simulation design ▸ f plots ▸ f' plots ▸ Strong identification ▸ Heteroskedastic

Summary

- Many models are non-regular for certain values of nuisance parameters
- These nuisance parameters can be either finite or ∞-dimensional
  - ▷ E.g.: Single index model with (close to) flat link function
  - ▷ Many other ∞-dimensional examples, e.g:
    1. Independent components analysis model with (weakly) non-Gaussian errors
    2. Mixed proportional hazard model with (close to) Weibull baseline hazard
- In many such models, a class of generalised C(α) tests have good properties
  - ▷ Correctly sized (locally uniformly)
  - ▷ If based on efficient score function, minimax optimal provided variance has positive rank
  - ▷ Simulation evidence suggests these asymptotics are a good guide to finite sample performance

Thank you!

References

Amari, S.-I. and Cardoso, J.-F. (1997), "Blind source separation-semiparametric statistical approach," *IEEE Transactions on Signal Processing*, 45, 2692–2700.

Andrews, D. W. K. and Cheng, X. (2012), "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80, 2153–2211.

Andrews, I. and Mikusheva, A. (2015), "Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models," *Quantitative Economics*, 6, 123–152.

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

Bravo, F., Escanciano, J. C., and Van Keilegom, I. (2020), "Two-step semiparametric empirical likelihood inference," *Ann. Statist.*, 48, 1–26.

Cattaneo, M. D., Farrell, M. H., and Feng, Y. (2020a), "Large sample properties of partitioning-based series estimators," *The Annals of Statistics*, 48, 1718 – 1741.

— (2020b), "Supplement to "Large sample properties of partitioning-based series estimators"," Https://doi.org/10.1214/19-AOS1865SUPP.

Chamberlain, G. (1986), "Asymptotic efficiency in semi-parametric models with censoring," *Journal of Econometrics*, 32, 189–218.

Chen, A. and Bickel, P. J. (2006), "Efficient independent component analysis," *The Annals of Statistics*, 34, 2825 – 2855.

Chen, X. (2007), "Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models," Elsevier, vol. 6 of *Handbook of Econometrics*, pp. 5549–5632.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2022), "Locally Robust Semiparametric Estimation," *Econometrica*, 90, 1501–1535.

Chernozhukov, V., Hansen, C., and Spindler, M. (2015), "Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach," *Annual Review of Economics*, 7, 649–688.

Comon, P. (1994), "Independent component analysis, A new concept?" *Signal Processing*, 36, 287–314.

Dufour, J.-M. and Valéry, P. (2016), "Rank-robust Regularized Wald-type tests," Working paper.

Escanciano, J. C. (2022), "SEMIPARAMETRIC IDENTIFICATION AND FISHER INFORMATION," *Econometric Theory*, 38, 301–338.

Fernandez, C. and Steel, M. F. J. (1998), "On Bayesian Modeling of Fat Tails and Skewness," *Journal of the American Statistical Association*, 93, 359–371.

Fiorentini, G. and Sentana, E. (2022), "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.

Hoesch, L., Lee, A., and Mesters, G. (2022), "Robust Inference in Structural VAR models Identified by Non-Gaussianity," Working Paper.

Horowitz, J. L. (2009), *Semiparametric and Nonparametric Methods in Econometrics*, Springer-Verlag New York.

Ichimura, H. (1993), "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of Econometrics*, 58, 71–120.

Lanne, M. and Luoto, J. (2021), "GMM Estimation of Non-Gaussian Structural Vector Autoregression," *Journal of Business & Economic Statistics*, 39, 69–81.

McCloskey, A. (2017), "Bonferroni-based size-correction for nonstandard testing problems," *Journal of Econometrics*, 200, 17–35.

Newey, W. K. (1990), "Semiparametric efficiency bounds," *Journal of Applied Econometrics*, 5, 99–135.

Newey, W. K. and Stoker, T. M. (1993), "Efficiency of Weighted Average Derivative Estimators and Index Models," *Econometrica*, 61, 1199–1223.

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

Independent components analysis

Independent components analysis

- Observe i.i.d. samples of K-dimensional data:

$$Y = A(\theta, \sigma)^{-1}\, \epsilon, \quad \mathbb{E}[\epsilon] = 0, \;\; \mathrm{Var}[\epsilon] = I_K, \;\; \epsilon_j \perp\!\!\!\perp \epsilon_k \; (k \neq j), \; \epsilon_k \sim \eta_k \tag{2}$$

- (First and) Second moments of Y are insufficient to identify $A^{-1} = A(\theta, \sigma)^{-1}$:

$$\mathrm{Var}(Y) = A^{-1}[A^{-1}]' = A^{-1}Q'Q[A^{-1}]' = (QA)^{-1}[(QA)^{-1}]' \;\text{ for any } Q \in O(K)$$

- Higher moment information can be used to shrink the identified set for A^{-1} and so θ
  - ▷ **ICA result**: If at least $K-1$ components of ϵ follow non-Gaussian distributions, then A^{-1} is identified up to sign and permutation of columns (Comon, 1994)
  - ▷ If ϵ is Gaussian or has multiple Gaussian components, $\tilde{\mathcal{I}}$ for θ is singular

◂ Example: Single index model

Independent components analysis

- A number of recent works in econometrics use the ICA approach to conduct inference in the SVAR model:[1]
  - e.g. Gouriéroux, Monfort, and Renne (2017); Lanne and Luoto (2021); Fiorentini and Sentana (2022)

- When ϵ is close to Gaussian, such methods break down

- In this model we can base robust inference on the efficient score function: $g_\gamma = \tilde{\ell}_\gamma$
  - Under regularity conditions (smoothness & moment conditions), form of $\tilde{\ell}_\gamma$ derived by Amari and Cardoso (1997); Chen and Bickel (2006)

- The components of $g_\gamma(W)$ are known up to $\phi_k(e) := \frac{\mathrm{d} \log \eta_k(e)}{\mathrm{d}e}$ and some moments
  - Moments: estimate via sample analogue
  - ϕ_k: use B-spline based estimator of Chen and Bickel (2006)

[1] Hoesch, Lee, and Mesters (2022) discuss robust inference in the SVAR model using this approach.

▸ Regularity conditions ▸ Form of $\tilde{\ell}_\gamma$

Independent components analysis - Monte Carlo

- Draw 5000 Monte Carlo samples from

$$Y = A^{-1}(\theta, \sigma)\epsilon, \quad A(\theta, \sigma) = \begin{bmatrix} {\sigma_1}^{-1} & 0 \\ 0 & {\sigma_2}^{-1} \end{bmatrix} \begin{bmatrix} 1 & -a \\ 1 & -b \end{bmatrix}, \quad \theta = (a, b)$$

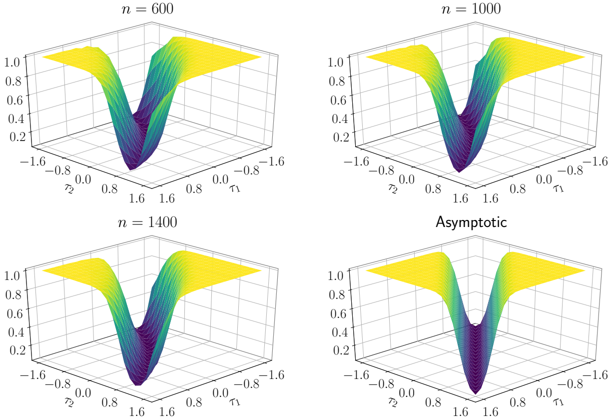
- Suppose that $c := (\sigma_2/\sigma_1)^2 = 1$. The efficient information matrix is

$$\tilde{\mathcal{I}}_\gamma = \frac{1}{(a-b)^2} \begin{bmatrix} \mathbb{E}[\phi_1(\epsilon_1)^2] & -1 \\ -1 & \mathbb{E}[\phi_2(\epsilon_2)^2] \end{bmatrix}$$

- Depending on the η_k, θ may be fully- or under- identified:

η_k	$\mathbb{E}[\phi_k(\epsilon_k)^2]$	$\text{rank}(\tilde{\mathcal{I}}_\gamma)$	$N(\tilde{\mathcal{I}}_\gamma)$
$\mathcal{N}(0,1)$	1	1	$\{\tau \in \mathbb{R}^2 : \tau_1 = \tau_2\}$
$t'(5)$	1.25	2	$\{0\}$
$st'(5,2)$	2.54	2	$\{0\}$

Figure 4: Power surfaces for ICA (ii), $\eta_1 \sim \mathcal{N}(0,1)$, $\eta_2 \sim \mathcal{N}(0,1)$

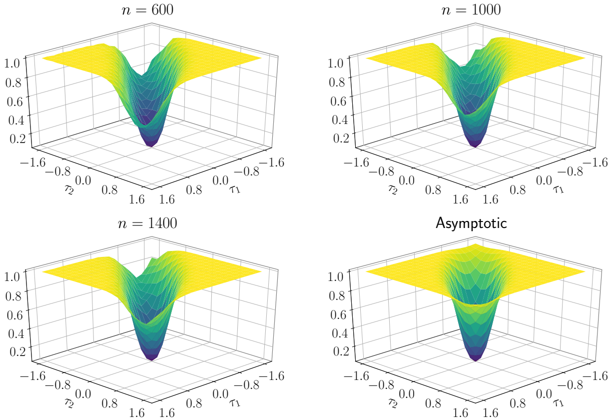


$\theta = (a, b) = (1/2, 1/4)$ and $\sigma_1 = \sigma_2 = 1$. The top-left, top-right and bottom-left panels are Monte Carlo version based on 5000 replications.

▸ Densities ▸ Simulation study 2

Independent components analysis - Monte Carlo

Figure 5: Power surfaces for ICA (ii), $\eta_1 \sim t'(5)$, $\eta_2 \sim t'(5)$

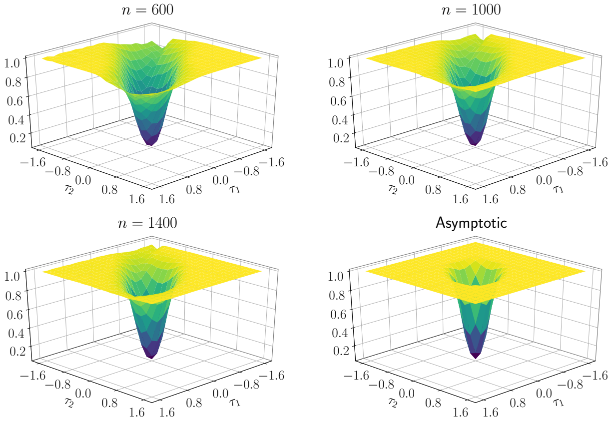


$\theta = (a, b) = (1/2, 1/4)$ and $\sigma_1 = \sigma_2 = 1$. The top-left, top-right and bottom-left panels are Monte Carlo version based on 5000 replications. $\eta_k \sim t'(5)$ indicates that each ϵ_k is drawn from a (standardised) t distribution with 5 df.

▸ Densities ▸ Simulation study 2

Independent components analysis - Monte Carlo

Figure 6: Power surfaces for ICA (ii), $\eta_1 \sim st'(5,2)$, $\eta_2 \sim st'(5,2)$



$\theta = (a, b) = (1/2, 1/4)$ and $\sigma_1 = \sigma_2 = 1$. The top-left, top-right and bottom-left panels are Monte Carlo version based on 5000 replications. $\eta_k \sim st'(5,2)$ indicates that each ϵ_k is drawn from a (standardised) skew t distribution Fernandez and Steel (1998) with 5 df and skewness 2.

▸ Densities ▸ Simulation study 2

Additional slides

Ichimura (1993)-style Wald test

The t / Wald test based on Ichimura (1993) uses the estimate $\hat{\theta}$ which minimises

$$\theta \mapsto \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{f}_n(V_{\theta,i}))^2, \quad V_\theta = X_1 + X_2\theta$$

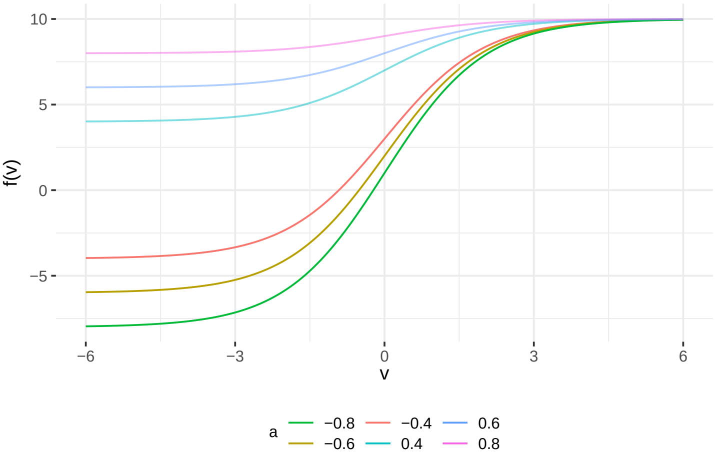
Under regularity conditions and the null, such an estimate satisfies

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \rightsquigarrow \mathcal{N}\left(0, \Omega\right),$$

for a known Ω.

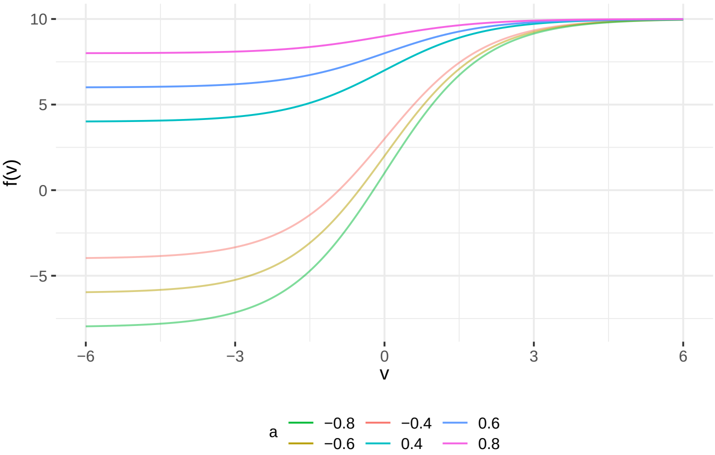
◂ SIM - t statistic   ◂ Simulation study

Figure 7: $f(v) = b\left(a + \frac{1-a}{1+\exp(-v)}\right)$, $b = 10$



SIM - t statistic
f' plot

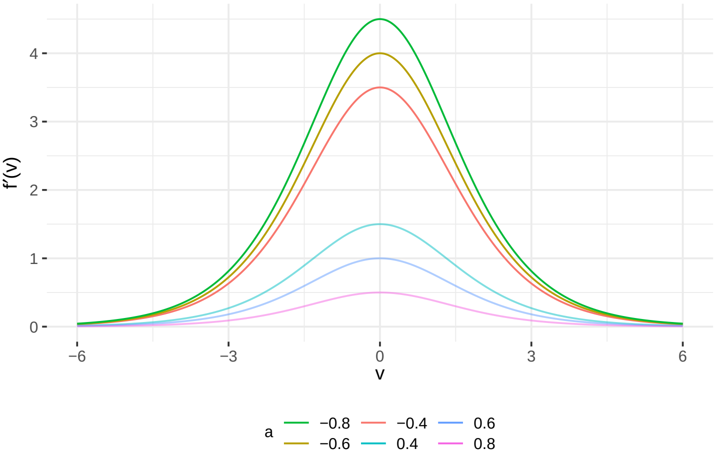
Figure 8: $f(v) = b\left(a + \frac{1-a}{1+\exp(-v)}\right),\ b = 10$



◂ SIM - t statistic
▸ f' plot

$f'(v)$

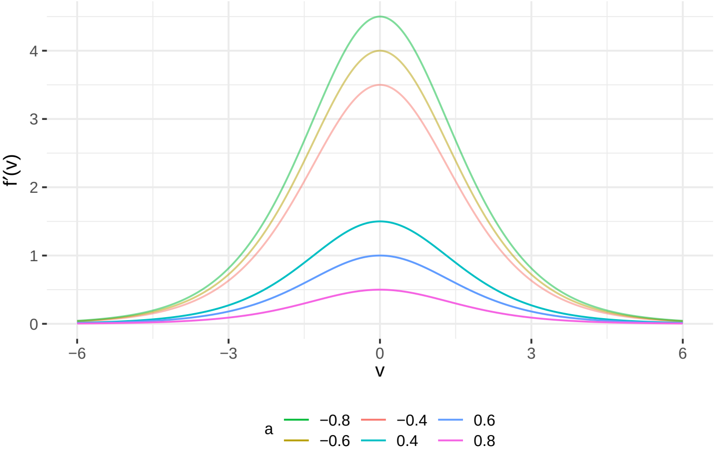
Figure 9: $f(v) = b\left(a + \frac{1-a}{1+\exp(-v)}\right),\ b = 10$



◂ SIM - t statistic
◂ f plot

$f'(v)$

Figure 10: $f(v) = b\left(a + \frac{1-a}{1+\exp(-v)}\right),\ b = 10$



Relation to the literature

1. Semiparametric estimation theory
   - Newey (1990); Bickel, Klaassen, Ritov, and Wellner (1998); van der Vaart (2002)
   - Requires nonsingular efficient information matrix
2. C(α)-style tests to handle e.g. machine learning first steps
   - Chernozhukov, Hansen, and Spindler (2015); Bravo, Escanciano, and Van Keilegom (2020); Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2022)
3. Weak identification robust inference
   - Andrews and Cheng (2012); Andrews and Mikusheva (2015); McCloskey (2017)
   - Focus on identification failure caused by a *finite* dimensional nuisance parameter
   - This paper: identification failure caused by *infinite* dimensional nuisance parameter

◂ This paper

Score functions

- Parameters: $\gamma = (\theta, \eta) \in \Gamma = \Theta \times \mathcal{H}$, $\Theta \subset \mathbb{R}^{d_\theta}$ open, $\mathcal{H}$ a metric space
- Model: $\mathcal{P} = \{P_\gamma : \gamma \in \Gamma\}$, with corresponding densities $p_\gamma \ll \nu$
- Score function for θ is $\dot{\ell}_\gamma$
  - ▷ Typically $\dot{\ell}_\gamma = \nabla_\theta \log p_\gamma$
- Score operator for η is $B_\gamma : H_\eta \to L_2(P_\gamma)$, $H_\eta \subset H$ a Banach space
  - ▷ B_γ maps directions $h \in H_\eta$ into scores $B_\gamma h$
  - ▷ Typically each $B_\gamma h$ is a pathwise directional derivative of $\eta \mapsto \log p_\gamma$ in direction h
- Can be formalised by e.g. requiring $\gamma \mapsto \sqrt{p_\gamma}$ to be pathwise directionally differentiable in $L_2(\nu)$ tangentially to $\mathbb{R}^{d_\theta} \times H_\eta$ with derivative

$$\frac{1}{2}\left[\tau'\dot{\ell}_\gamma + B_\gamma h\right]\sqrt{p_\gamma} \quad \text{in direction} \quad (\tau, h).$$

◂ Scores

Truncation

Proposition

Suppose that $0 \preceq M_n \to M$ are deterministic $L \times L$ matrices with $\mathrm{rank}(M_n) = \mathrm{rank}(M)$ for $n \geq N$ and for $\check{M}_n \succeq 0$ and $0 \leq \nu_n \to 0$,

$$\lim_{n\to\infty} P_n \left(\|\check{M}_n - M_n\|_2 < \nu_n \right) = 1.$$

Let $\check{M}_n = \check{U}_n \check{\Lambda}_n \check{U}_n'$ be the corresponding eigendecompositions and define

$$\hat{M}_n := \check{U}_n \Lambda_n(\nu_n) \check{U}_n', \quad \Lambda_n(\nu_n) := \mathrm{diag}\left(\check{\Lambda}_{n,ii} 1(\check{\Lambda}_{n,ii} \geq \nu_n) \right)_{i=1}^{L},$$

where the eigenvalues are ordered non-increasingly.

Then, $\hat{M}_n \xrightarrow{P_n} M$ and

$$\lim_{n\to\infty} P_n \left(\mathrm{rank}(\hat{M}_n) = \mathrm{rank}(M) \right) = 1.$$

◂ C(α)

Theoretical results - size

Assumptions

1. Local asymptotic normality
   - The model is asymptotically locally approximable by a Gaussian shift experiment
2. In P_γ – probability

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\hat{g}_{n,\theta_0}(W_i) - g_\gamma(W_i)] \to 0, \qquad \hat{V}^\dagger_{n,\theta_0} \to \hat{V}^\dagger_\gamma, \qquad \text{and} \qquad \hat{r}_n \to r.$$

Proposition

If assumptions 1 & 2 are satisfied and $h_n \to h$,

$$\hat{S}_{n,\theta_0} \overset{P_{n,0,h_n}}{\rightsquigarrow} \chi^2_r, \qquad r := \text{rank}(V_\gamma),$$

and the test is asymptotically (locally uniformly) correctly sized.

C(α) test | Assumption details | Result details

Theoretical results - power

Assumptions

1. Local asymptotic normality
   - The model is asymptotically locally approximable by a Gaussian shift experiment
2. In P_γ – probability

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\hat{g}_{n,\theta_0}(W_i) - \tilde{\ell}_\gamma(W_i)\right] \to 0, \quad \hat{V}^{\dagger}_{n,\theta_0} \to \tilde{\mathcal{I}}^{\dagger}_\gamma = \left[\mathbb{E}_\gamma\, \tilde{\ell}(W_i)\tilde{\ell}(W_i)'\right]^{\dagger}, \quad \text{and} \quad \hat{r}_n \to r.$$

3. The permitted directions H_η for h form a linear space
4. $\text{rank}(\tilde{\mathcal{I}}_\gamma) > 0$

Proposition

If assumptions 1 – 4 are satisfied, the efficient score test has locally asymptotically minimax optimal power

◂ C(α) test ▸ Assumption details ▸ Result details

Local Asymptotic Normality (LAN)

Suppose that $\eta_n : H_\eta \to \mathcal{H}$ is a sequence continuously convergent to η and define for $(\tau, h) \in \mathbb{R}^{d_\theta} \times H_\eta$,

$$\gamma_n(\tau, h) := (\theta + \tau/\sqrt{n}, \eta_n(h)).$$

For any $\tau_n \to \tau \in \mathbb{R}^{d_\theta}$ and any $h_n \to h \in H_\eta$,

1. $P_{\gamma_n(\tau_n, h_n)} \in \mathcal{P}$ for all sufficiently large n
2. for a function $\dot{\ell}_\gamma \in L_2^0(P_\gamma)$ and a linear map $B_\gamma : H_\eta \to L_2^0(P_\gamma)$,

$$\Lambda_n(\gamma_n(\tau_n, h_n), \gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} q(W_i) - \frac{1}{2} \mathbb{E}[g(W_i)^2] + o_{P_\gamma}(1),$$

with $q := \tau' \dot{\ell}_\gamma + B_\gamma h$.

◂ Main results ▸ DQM is sufficient for LAN

Differentiability in quadratic mean (DQM)

Suppose that $\eta_n : H_\eta \to \mathcal{H}$ is a sequence continuously convergent to η and define for $(\tau, h) \in \mathbb{R}^{d_\theta} \times H_\eta$,

$$\gamma_n(\tau, h) := (\theta + \tau/\sqrt{n}, \eta_n(h)).$$

For any $\tau_n \to \tau \in \mathbb{R}^{d_\theta}$ and any $h_n \to h \in H_\eta$,

1. $P_{\gamma_n(\tau_n,h_n)} \in \mathcal{P}$ for all sufficiently large n
2. for a function $\dot{\ell}_\gamma \in L_2^0(P_\gamma)$ and a linear map $B_\gamma : H_\eta \to L_2^0(P_\gamma)$,

$$\int \left[\sqrt{n}\left(\sqrt{p_{\gamma_n(\tau_n,h_n)}} - \sqrt{p_\gamma}\right) - \frac{1}{2} q \sqrt{p_\gamma}\right]^2 \mathrm{d}\nu \to 0,$$

with $q := \tau' \dot{\ell}_\gamma + B_\gamma h$.

Main results LAN

Equivariance in law

Proposition - equivariance in law

If assumption 1 holds then

1. (P^n_γ) and $(P^n_{\gamma_n(\tau_n,h_n)})$ are mutually contiguous
2. Under $P_{\gamma_n(\tau_n,h_n)}$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_\gamma(W_i) \rightsquigarrow \mathcal{N}\left(\mathbb{E}_\gamma\left[g_\gamma(W_i)\dot{\ell}_\gamma(W_i)'\right]\tau, V_\gamma\right)$$

If assumption 2 also holds, then

3. Under $P_{\gamma_n(\tau_n,h_n)}$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \hat{g}_{n,\theta}(W_i) \rightsquigarrow \mathcal{N}\left(\mathbb{E}_\gamma\left[g_\gamma(W_i)\dot{\ell}_\gamma(W_i)'\right]\tau, V_\gamma\right)$$

Size

Proposition - Test is correctly sized

Let ϕ_{n,θ_0} be the generalised C(α) test. If assumptions 1 & 2 hold with $\theta = \theta_0$, then for any $h_n \to h$ in H_η,

$$\limsup_{n\to\infty} \mathbb{E}_{\gamma_n(0,h_n)} \phi_{n,\theta_0} \leq \alpha.$$

- test is asymptotically locally uniformly correctly sized
- size is α unless $V_\gamma = 0$, in which case it is 0

◂ Return to main results

Minimax optimality for multivariate hypotheses

Proposition

Suppose assumptions 1, 3 & 4 hold, $\alpha \in (0,1)$, $H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$.

1. For any sequence (ψ_n) of asymptotically level-α tests of H_0 against H_1 and all $a > 0$,

$$\limsup_{n\to\infty} \inf_{(\tau,h)\in M_a} \mathbb{E}_{\gamma_n(\tau,h)} \psi_n \leq 1 - \mathbb{P}\left(\chi^2_r(a) \leq c_{r,\alpha}\right),$$

where $M_a := \{(\tau, h) \in N(\tilde{\mathcal{I}}_\gamma)^\perp \times H_\eta : \tau'\tilde{\mathcal{I}}_\gamma\tau \geq a\}$, $c_{r,\alpha}$ is the $1-\alpha$ quantile of the χ^2_r distribution and $\chi^2_r(a)$ denotes a non-central χ^2 random variable with $r := \text{rank}(\tilde{\mathcal{I}}_\gamma)$ degrees of freedom and non-centrality a.

2. If assumption 2 also holds, for all compacts $K_\alpha \subset M_a$, the efficient score test satisfies

$$\lim_{n\to\infty} \inf_{(\tau,h)\in K_a} \mathbb{E}_{\gamma_n(\tau,h)} \phi_{n,\theta_0} = 1 - \mathbb{P}\left(\chi^2_r(a) \leq c_{r,\alpha}\right).$$

◂ Return to main results

SIM - Assumptions

Assumption SIM

$W = (Y, X)$ obeys the SIM with $\Theta \subset \mathbb{R}^K$ open and $\mathcal{H}$ such that

1. For $\phi(\epsilon, X) := \frac{\partial log\zeta(e,x)}{\partial e}(\epsilon, X)$ where $(\epsilon, X) \sim \zeta$ and some $\rho > 0$,

$$\mathbb{E}\left[\left(|\phi(\epsilon, X)|^{2+\rho} + 1\right) \|X\|_2^{2+\rho}\right] < \infty, \quad \mathbb{E}[XX'] \succ 0,$$

$$\mathbb{E}\left[\epsilon\phi(\epsilon, X)|X\right] = -1, \ \mathbb{E}[\phi(\epsilon, X)^2|X] < C < \infty, \ \mathbb{E}[\epsilon^2|X] \in (c, C).$$

2. $e \mapsto \sqrt{\zeta(e, X)}$ is continuously differentiable a.e.
3. f is bounded and continuously differentiable with bounded derivative a.e. on $\mathscr{D} := \{x_1 + x_2\theta : \theta \in \Theta, x \in \mathscr{X}\}$, where $\text{supp}(X) \subset \mathscr{X}$.
4. There exists a $\tilde{m} : \mathbb{R} \to \mathbb{R}$ bounded, continuously differentiable with bounded derivative such that $\mathbb{E}[\epsilon\tilde{m}(\epsilon)|X] \geq c > 0$.

Return to SIM

Split sample estimators

- Using **split sample estimators** removes dependence between (function) estimation error and evaluation data point
- Suppose we have a function estimator which depends on a parameter vector $\hat{\xi}_n \in \mathbb{R}^{K_n}$:

$$\hat{f}_n(v) = \check{f}_n(v, \hat{\xi}_n).$$

- Split the data in half (assume n even)

$$\underbrace{W_1, \ldots, W_{n/2}}_{\hat{\xi}_n^{(1)}} \qquad \& \qquad \underbrace{W_{n/2+1}, \ldots, W_n}_{\hat{\xi}_n^{(2)}}$$

- Our estimates of the $f(V_i)$ are:

$$\check{f}_n(V_1, \hat{\xi}_n^{(2)}), \ldots, \check{f}_n(V_{n/2}, \hat{\xi}_n^{(2)}) \qquad \& \qquad \check{f}_n(V_{n/2+1}, \hat{\xi}_n^{(1)}), \ldots, \check{f}_n(V_n, \hat{\xi}_n^{(1)})$$

◂ SIM estimation

Series estimators

- In the single index model, $f(V_\theta) = \mathbb{E}[Y|V_\theta]$. We can estimate f and f' as

$$\hat{f}_n(V_{\theta,i}) = \check{f}_n(v, \hat{\xi}_{1,n,i}) = q_n(V_{\theta,i})'\hat{\xi}_{1,n,i}$$
$$\widehat{f'}_n(V_{\theta,i}) = \check{f'}_n(V_{\theta,i}, \hat{\xi}_{2,n,i}) = \left[q_n'(V_{\theta,i})\right]' \hat{\xi}_{2,n,i},$$

where q_n is a K_n-vector of basis functions ($\mathbb{R} \to \mathbb{R}$), q_n' their derivatives,

$$\hat{\xi}_{1,n,i} = \hat{\xi}_{2,n,i} = \left(\sum_{j \in N_{-i}} q_n(V_{\theta,j}) q_n(V_{\theta,j})'\right)^{-1} \left(\sum_{j \in N_{-i}} q_n(V_{\theta,j}) Y_j\right).$$

- Examples of basis functions q_n include B-splines and (local) polynomials
  - ▷ see e.g. Chen (2007) for many examples
- Analogous estimators can be constructed for $Z(V_\theta) = \mathbb{E}[X_2|V_\theta]$.

◂ SIM estimation

SIM - Nonparametric estimation assumptions

Assumption SIM-NP

Suppose that $\mathscr{X}$ is a compact set, $\sigma^2 := \mathbb{E}[\epsilon^2|X] = \mathbb{E}[\epsilon^2]$, $\mathbb{E}[\epsilon^4] < \infty$ and with P_γ-probability approaching one for $l \in [3]$, each $i \in [n]$, some $r_n = o(n^{-1/4})$ and $\mathcal{V}$ the distribution of V_θ:

$$\mathcal{R}_{1,n,i} := \left(\int \left[\check{f}_n(v, \hat{\xi}_{1,n,i}) - f(v)\right]^2 \mathrm{d}\mathcal{V}(v)\right)^{1/2} \leq r_n,$$

$$\mathcal{R}_{2,n,i} := \left(\int \left[\check{f'}_n(v, \hat{\xi}_{2,n,i}) - f'(v)\right]^2 \mathrm{d}\mathcal{V}(v)\right)^{1/2} \leq r_n,$$

$$\mathcal{R}_{3,n,i} := \left(\int \left\|\check{Z}_n(v, \hat{\xi}_{3,n,i}) - Z(v))\right\|_2^2 \mathrm{d}\mathcal{V}(v)\right)^{1/2} \leq r_n.$$

- These rate conditions are attainable under mild regularity conditions, using e.g. series estimators e.g. Cattaneo, Farrell, and Feng (2020a).

◂ SIM estimation ▸ Regularity conditions for rates

Regularity conditions for NP estimation rates

If f, f' and Z are estimated using equally spaced third order (i.e. quadratic) B-spline series estimators with the number of splines $K \asymp n^{1/7}$ and

1. $\mathscr{D}$ is compact and connected and the density of V_θ is continuous and bounded away from zero,
2. $v \mapsto f(v)$ and $v \mapsto Z(v) = \mathbb{E}[X_2|v]$ are three-times continuously differentiable with Hölder continuous derivatives

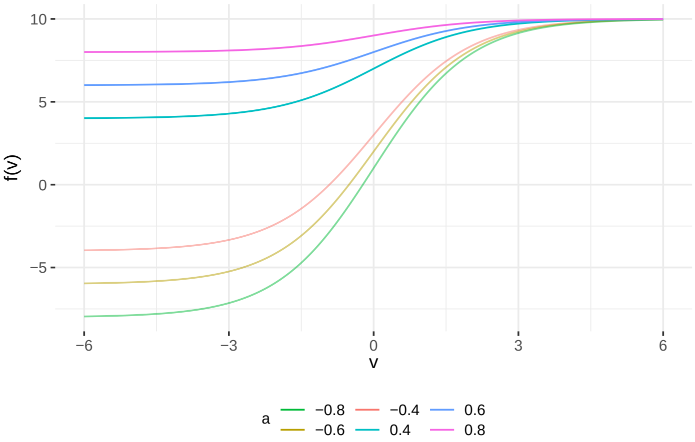
then with P_γ-probability approaching one for $l \in [3]$ and each $i \in [n]$, $\mathcal{R}_{l,n,i} \leq r_n = o(n^{-1/4})$.

- Follows from Cattaneo et al. (2020a) & Cattaneo, Farrell, and Feng (2020b).

◂ High level conditions

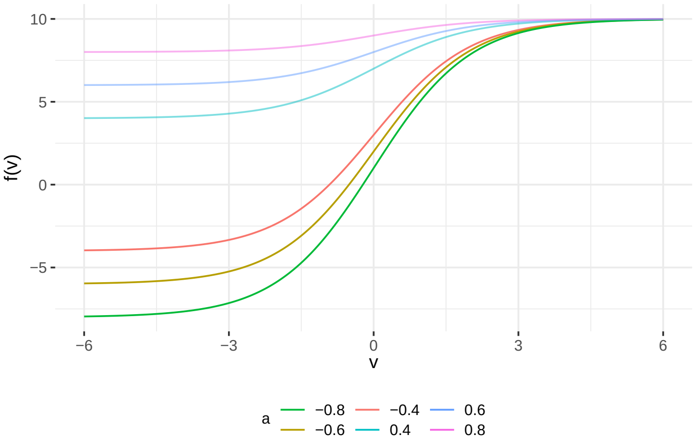
Single index model - f shapes

Figure 11: $f(v) \propto a + \frac{1-a}{1+\exp(-v)}$



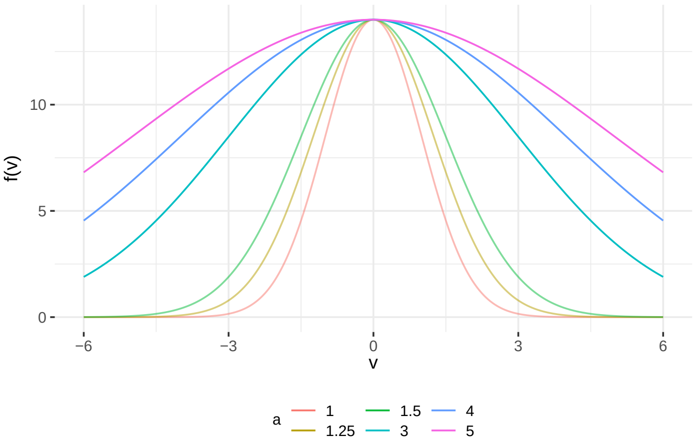
Single index model - f shapes

Figure 12: $f(v) \propto a + \frac{1-a}{1+\exp(-v)}$



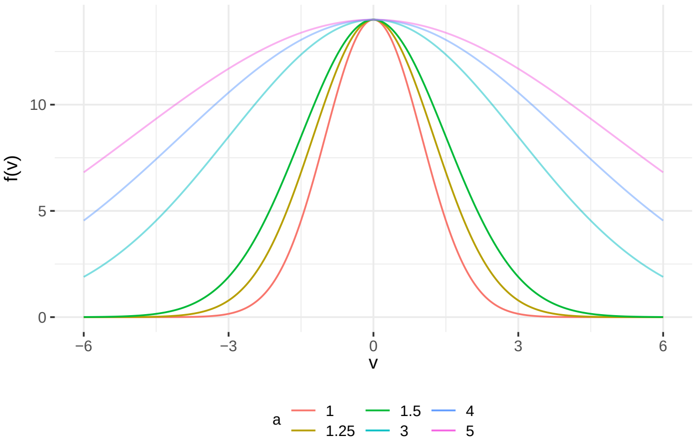
Single index model - f shapes

Figure 13: $f(v) \propto \exp\left(-v^2/(2a^2)\right)$



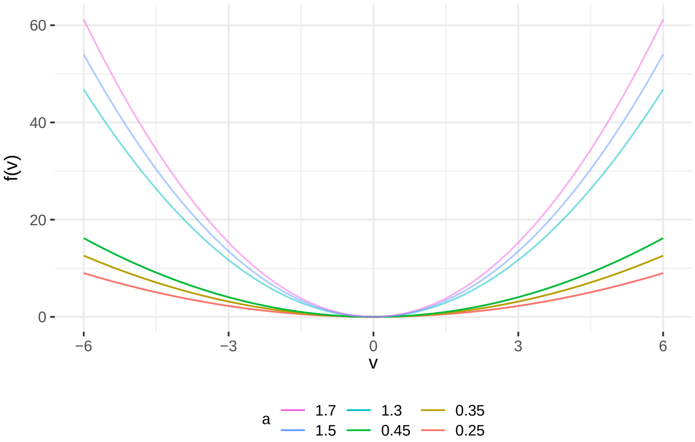
Single index model - f shapes

Figure 14: $f(v) \propto \exp\left(-v^2/(2a^2)\right)$



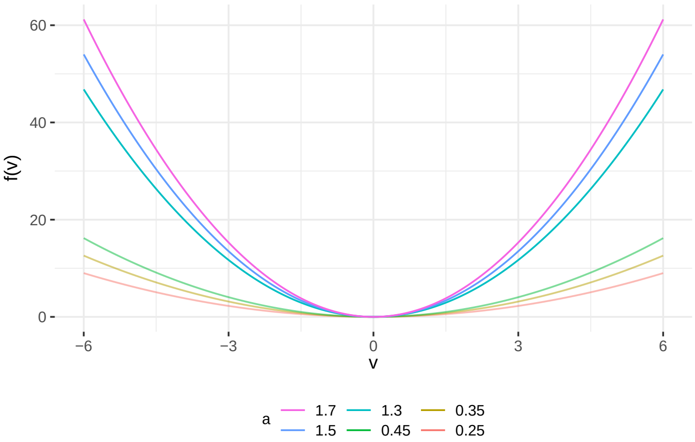
Single index model - f shapes

Figure 15: $f(v) = av^2$



Single index model - f shapes

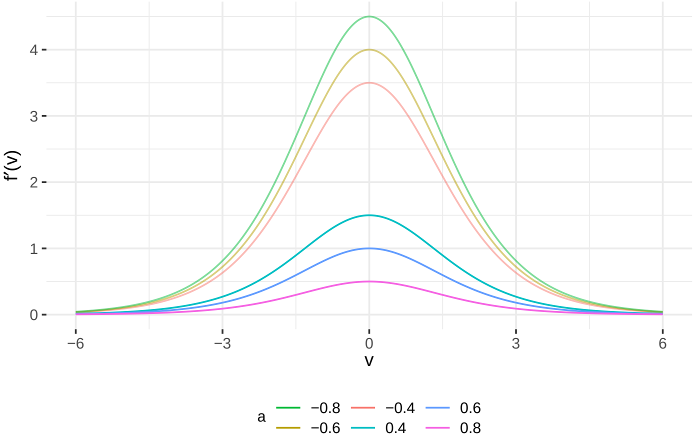
Figure 16: $f(v) = av^2$



t-statistic Size Power

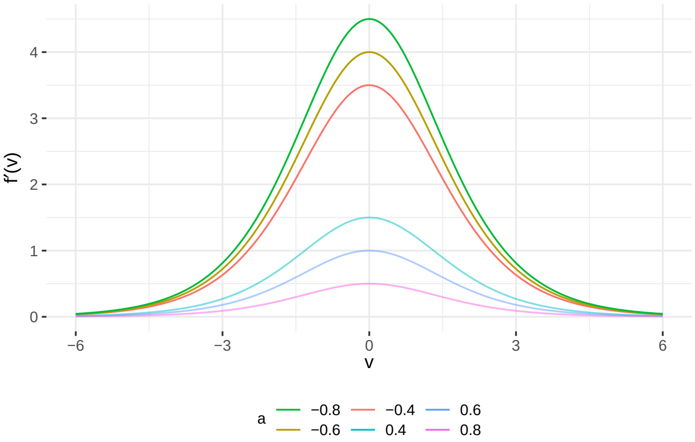
Single index model - f' shapes

Figure 17: $f(v) \propto a + \frac{1-a}{1+\exp(-v)}$



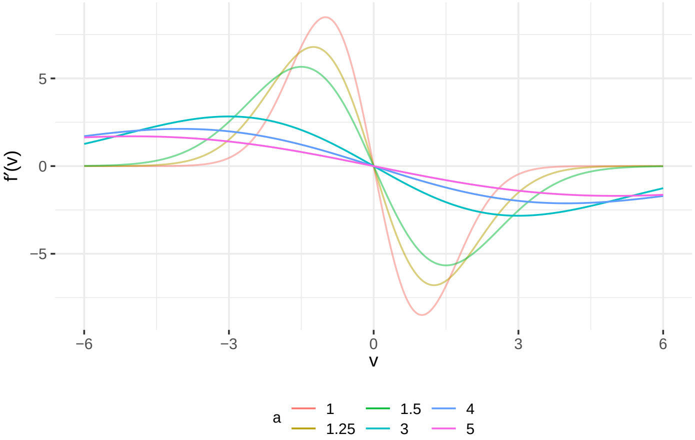
Single index model - f' shapes

Figure 18: $f(v) \propto a + \frac{1-a}{1+\exp(-v)}$



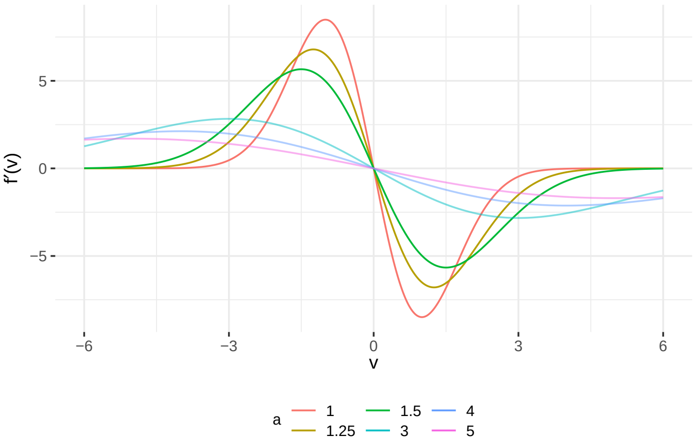
Single index model - f' shapes

Figure 19: $f(v) \propto \exp\left(-v^2/(2a^2)\right)$



Single index model - f' shapes

Figure 20: $f(v) \propto \exp\left(-v^2/(2a^2)\right)$



Single index model - f' shapes

Figure 21: $f(v) = av^2$

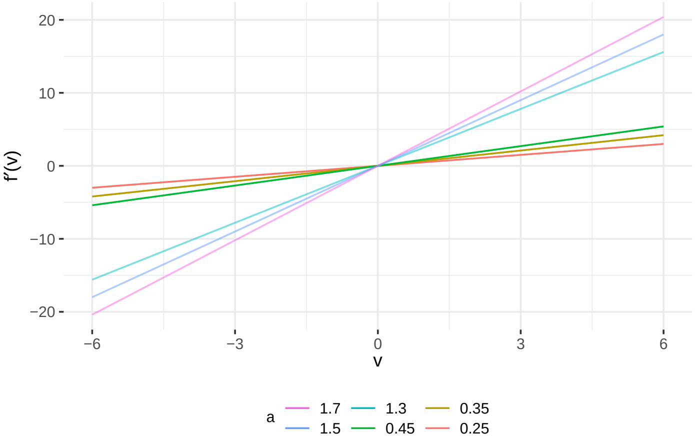
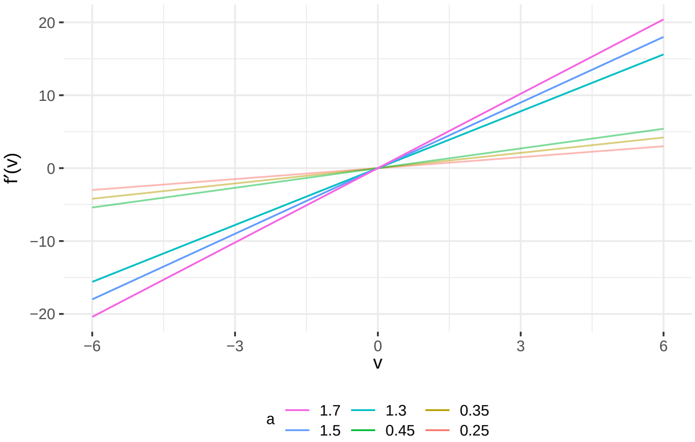


Figure 22: $f(v) = av^2$



t-statistic
Size
Power

Single index model: Monte carlo

Table 2: Empirical rejection frequencies (%), $\epsilon \sim \mathcal{N}(0,1)$, $X_i \sim U[-1,1]$, $\theta = 1$

	$f_1(v) \propto a + (1-a)/(1+\exp(-v))$			$f_2(v) \propto \exp(-v^2/(2a^2))$			$f_3(v) = av^2$		
n	$a = -0.8$	$a = -0.6$	$a = -0.4$	$a = 1$	$a = 1.25$	$a = 1.5$	$a = 1.7$	$a = 1.5$	$a = 1.3$
$\hat{S}$									
400	4.86	5.56	5.88	5.54	5.80	5.74	5.30	5.96	5.60
600	5.40	5.46	4.80	5.32	5.38	5.40	5.34	5.50	5.64
800	4.98	4.88	5.40	5.38	5.36	5.10	5.70	5.54	5.12
1000	5.48	5.14	5.82	5.06	5.20	5.26	5.54	5.18	5.44
W									
400	8.32	9.30	8.90	8.94	9.30	10.52	14.74	15.98	17.18
600	7.08	7.70	7.70	8.54	8.36	9.30	12.74	13.50	14.76
800	7.10	6.78	7.74	6.94	7.68	8.88	11.76	12.30	13.08
1000	5.88	6.88	6.18	6.32	7.80	7.70	10.56	11.76	11.46

Based on 5000 Monte Carlo replications. $\hat{S}$ is the C(α) – style test.

◂ Simulation design ▸ f plots ▸ f' plots ◂ Weak identification ▸ Heteroskedastic

Single index model: Monte Carlo

Table 3: Empirical rejection frequencies (%), $\epsilon \sim \mathcal{N}(0, \log(2 + V_{\theta}^2))$, $X_i \sim U[-1, 1]$, $\theta = 1$

	$f_1(v) \propto a + (1-a)/(1+\exp(-v))$			$f_2(v) \propto \exp(-v^2/(2a^2))$			$f_3(v) = av^2$		
n	$a = 0.4$	$a = 0.6$	$a = 0.8$	$a = 3$	$a = 4$	$a = 5$	$a = 0.45$	$a = 0.35$	$a = 0.25$
$\hat{S}$									
400	5.12	4.92	5.56	5.90	5.90	5.50	5.58	5.38	5.82
600	4.76	5.46	5.62	5.36	5.28	5.62	5.40	5.76	6.12
800	5.34	5.28	5.22	5.26	4.86	5.16	5.68	4.76	4.66
1000	5.12	5.84	5.14	5.22	5.78	5.04	5.38	5.26	5.24
W									
400	20.38	25.36	15.00	31.58	27.86	22.04	29.04	24.90	20.90
600	15.92	20.46	18.78	26.80	30.64	23.32	32.50	27.12	24.04
800	13.44	18.46	23.92	24.30	30.76	25.78	31.58	30.40	25.92
1000	12.48	16.10	26.76	22.80	29.46	28.50	31.42	31.38	26.40

Based on 5000 Monte Carlo replications. $\hat{S}$ is the C(α) – style test.

◂ Simulation design ▸ f plots ▸ f' plots ▸ Strong identification ◂ Homoskedastic

Single index model: Monte Carlo

Table 4: Empirical rejection frequencies (%), $\epsilon \sim \mathcal{N}(0, \log(2 + V_\theta^2))$, $X_i \sim U[-1, 1]$, $\theta = 1$

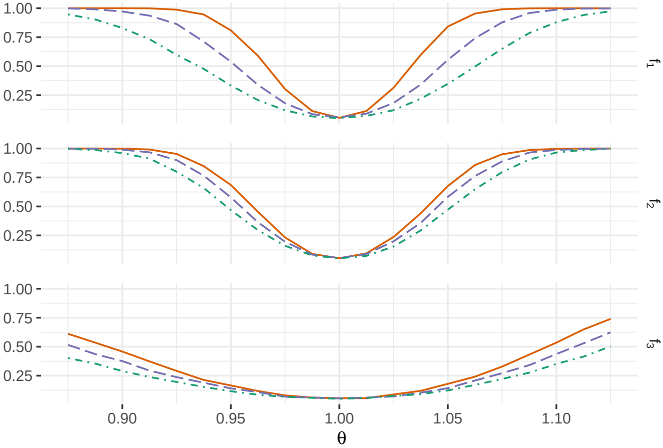
	$f_1(v) \propto a + (1-a)/(1+\exp(-v))$			$f_2(v) \propto \exp(-v^2/(2a^2))$			$f_3(v) = av^2$		
n	$a = -0.8$	$a = -0.6$	$a = -0.4$	$a = 1$	$a = 1.25$	$a = 1.5$	$a = 1.7$	$a = 1.5$	$a = 1.3$
$\hat{S}$									
400	5.00	5.66	5.86	5.42	6.16	5.74	5.42	5.92	5.96
600	5.52	5.48	4.94	5.52	5.58	5.40	5.32	5.58	5.68
800	5.00	4.84	5.08	5.42	5.40	4.82	5.80	5.34	5.26
1000	5.62	5.14	5.76	5.24	5.04	5.16	5.70	5.36	5.42
W									
400	9.40	10.18	10.26	12.08	13.46	15.74	22.00	22.90	24.54
600	7.76	8.02	8.42	10.16	12.42	13.88	18.40	20.04	21.40
800	6.78	6.90	7.62	8.70	10.88	13.08	16.94	17.96	19.20
1000	5.84	6.84	5.58	7.78	10.12	10.66	15.58	16.78	17.40

Based on 5000 Monte Carlo replications. $\hat{S}$ is the C(α) – style test.

◂ Simulation design ▸ f plots ▸ f' plots ▸ Weak identification ◂ Homoskedastic

Single index model: Monte carlo

Figure 23: Empirical rejection frequency, $\epsilon \sim \mathcal{N}(0,1)$, $X_i \sim U[-1,1]$

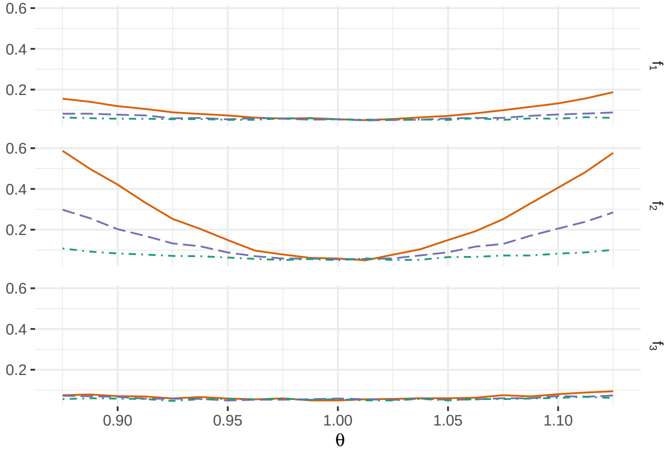


$f_1(v)$ = generalised logistic, $f_2(v)$ = Gaussian function, $f_3(v)$ = quadratic

◂ Simulation design ▸ f plots ▸ f' plots ▸ Weak identification ▸ Heteroskedastic

Single index model: Monte Carlo

Figure 24: Empirical rejection frequency, $\epsilon \sim \mathcal{N}(0, \log(2 + V_\theta^2))$, $X_i \sim U[-1, 1]$

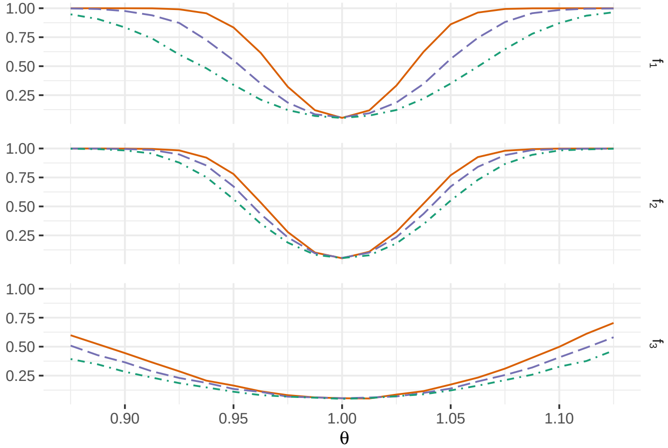


$f_1(v)$ = generalised logistic, $f_2(v)$ = Gaussian function, $f_3(v)$ = quadratic

◂ Simulation design ▸ f plots ▸ f' plots ▸ Strong identification ◂ Homoskedastic

Single index model: Monte Carlo

Figure 25: Empirical rejection frequency, $\epsilon \sim \mathcal{N}(0, \log(2 + V_\theta^2))$, $X_i \sim U[-1, 1]$



$f_1(v)$ = generalised logistic, $f_2(v)$ = Gaussian function, $f_3(v)$ = quadratic

◂ Simulation design ▸ f plots ▸ f' plots ▸ Weak identification ◂ Homoskedastic

ICA Assumption

Equation (2) holds, where each ϵ_k $(k = 1, \dots, K)$ has a (Lebesgue) density η_k, and for each $k = 1, \dots, K$,

1. $e \mapsto \sqrt{\eta_k(e)}$ and $(\theta, \sigma) \mapsto A(\theta, \sigma)$ are continuously differentiable,
2. $\mathbb{E}[|\epsilon_k|^{4+\delta}] < \infty, \quad \mathbb{E}[\epsilon_k^4] - 1 > \mathbb{E}[\epsilon_k^3]^2, \quad \mathbb{E}[|\phi_k|^{4+\delta}] < \infty,$
3. $\mathbb{E}[\phi_k(\epsilon_k)] = 0, \quad \mathbb{E}[\phi_k(\epsilon_k)\epsilon_k] = -1, \quad \mathbb{E}[\phi_k(\epsilon_k)\epsilon_k^2] = 0, \quad \mathbb{E}[\phi_k(\epsilon_k)\epsilon_k^3] = -3,$

for some $\delta > 0$ and $\phi_k := \frac{\mathrm{d} \log \eta_k(e)}{\mathrm{d} e}$.

◂ Efficient score

Form of efficient score function in ICA

- Under regularity conditions, in the ICA model:

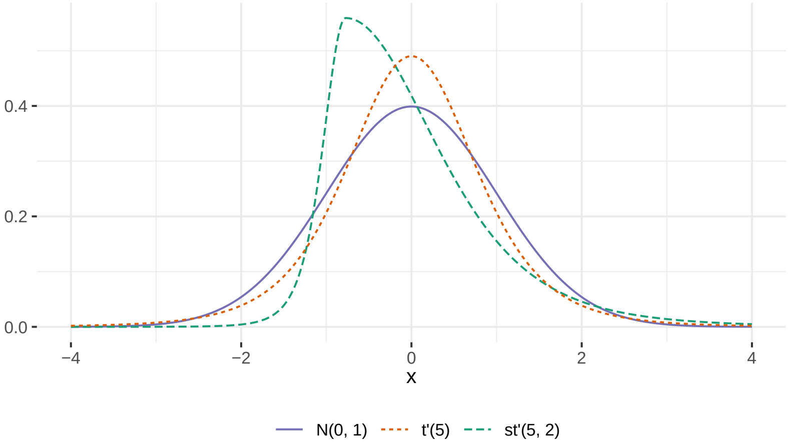
$$\tilde{\ell}_\gamma(W) = \sum_{k=1}^{K} \sum_{j=1, j\neq k}^{K} \zeta_{l,k,j} \phi_k(A_{k\bullet} y) A_{j\bullet} y + \sum_{k=1}^{K} \zeta_{l,k,k} \left[\tau_{k,1} A_{k\bullet} y + \tau_{k,2} \kappa(A_{k\bullet} y)\right],$$

where $\phi_k(e) := \frac{\mathrm{d} \log \eta_k(e)}{\mathrm{d}e}$, $\zeta_{l,k,j} := [D_l(\beta)]_{k\bullet} A^{-1}_{\bullet j}$ with $D_l(\beta) = \partial A(\beta)/\partial \beta_l$, $A^{-1}_{\bullet j}$ is the j-th column of $A(\theta, \sigma)^{-1}$, $\kappa(e) := e^2 - 1$ and

$$\tau_k := \begin{pmatrix} 1 & \mathbb{E}(\epsilon_k)^3 \\ \mathbb{E}(\epsilon_k)^3 & \mathbb{E}(\epsilon_k)^4 - 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ -2 \end{pmatrix}.$$

▸ Regularity conditions ◂ ICA

Figure 26: True densities used in ICA simulation study 2



◂ Simulation study 1

Independent components analysis - Monte Carlo

- Draw 5000 Monte Carlo replications from

$$Y = A^{-1}(\theta, \sigma)\epsilon, \quad A(\theta, \sigma) = R(\theta)'\Sigma^{1/2}, \quad R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \; \sigma = \text{vech}(\Sigma^{1/2}),$$

  for a range of different η_1, η_2.
  - ▷ As both become close to $\mathcal{N}(0, 1)$, we should expect weak identification.
- Compare finite sample performance with GMM Lanne and Luoto (2021) and PML Gouriéroux et al. (2017) approaches
  - ▷ Wald tests: unlikely to perform well (weak identification)
  - ▷ LM tests: fix θ under the null so likely to perform better

▸ PML approach ▸ GMM approach ◂ Simulation study 1

Independent components analysis - Monte Carlo

Table 5: Empirical rejection frequencies (%) for ICA, $\epsilon_1 \sim \mathcal{N}(0,1)$

$\epsilon_2 \sim$	$\mathcal{N}(0,1)$	$t'(\nu)$			$\mathcal{SN}'(0,1,\nu)$			ABM(ν)		
n		15	10	5	2	3	4	19/20	17/20	15/20
$\hat{S}$										
400	6.40	7.04	6.44	6.04	6.50	6.28	6.76	6.26	6.18	6.56
800	6.92	6.06	6.16	6.14	6.26	7.14	7.36	6.50	5.88	6.78
PML - LM										
400	5.26	5.82	4.92	6.40	5.08	5.18	5.04	5.68	5.76	5.42
800	5.24	4.94	5.58	7.86	5.52	6.08	5.90	5.48	4.80	5.20
GMM - LM										
400	3.88	4.04	4.60	6.40	3.94	3.98	4.56	6.10	5.08	3.82
800	3.46	3.36	3.86	6.04	3.00	3.36	3.30	7.10	5.62	2.98
PML - Wald										
400	25.88	14.04	10.68	2.84	20.40	16.80	15.70	56.80	49.66	33.62
800	25.84	9.92	6.26	2.22	16.98	13.34	12.26	70.66	59.78	37.74
GMM - Wald										
400	76.42	68.58	60.82	25.48	70.26	66.78	63.70	73.74	75.64	76.68
800	86.92	82.18	75.66	29.64	84.34	80.46	79.66	83.92	85.58	87.52

Based on 5000 Monte Carlo replications. $\hat{S}$ is the efficient score test.

▸ Densities - t ▸ Densities - SN ▸ Densities - ABM

Independent components analysis - Monte Carlo

Table 6: Empirical rejection frequencies (%) for ICA, $\epsilon_1 \sim t'(5)$

$\epsilon_2 \sim$	$\mathcal{N}(0,1)$	$t'(\nu)$			$\mathcal{SN}'(0,1,\nu)$			ABM(ν)		
n		15	10	5	2	3	4	19/20	17/20	15/20
$\hat{S}$										
400	6.38	5.76	5.24	5.44	6.16	6.80	6.46	5.24	5.72	5.86
800	5.90	6.20	5.56	5.28	5.76	5.64	5.44	5.94	5.40	5.78
PML - LM										
400	5.58	5.14	5.10	5.14	5.40	6.28	5.36	5.76	5.98	5.34
800	6.46	6.28	5.56	5.00	6.10	5.56	5.28	6.90	6.46	6.04
GMM - LM										
400	25.68	24.14	26.46	25.34	24.94	25.52	25.20	27.60	26.80	25.72
800	25.72	23.62	23.56	23.00	25.04	25.00	25.14	27.22	28.18	26.08
PML - Wald										
400	1.80	1.84	2.10	2.20	1.62	1.64	1.78	1.44	1.74	1.32
800	2.94	3.38	3.16	3.28	3.06	2.68	2.84	2.06	2.30	2.80
GMM - Wald										
400	4.18	3.70	3.32	2.18	3.94	3.74	3.70	3.86	3.52	3.80
800	1.48	1.52	1.50	1.40	1.10	1.72	1.22	1.12	1.18	1.06

Based on 5000 Monte Carlo replications. $\hat{S}$ is the efficient score test.

Densities - t | Densities - SN | Densities - ABM

Independent components analysis - Monte Carlo

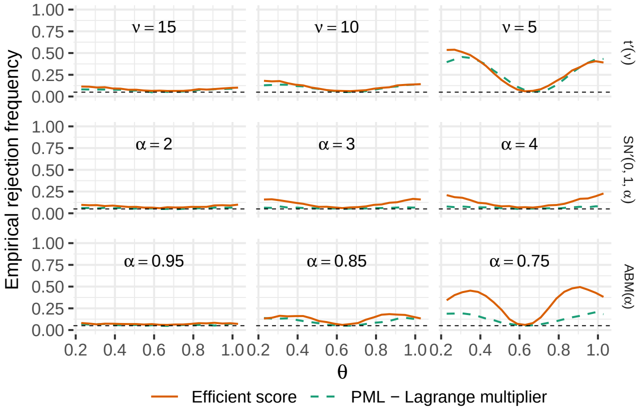
Table 7: Empirical rejection frequencies (%) for ICA, $\epsilon_1 \sim \mathcal{SN}'(0,1,4)$

$\epsilon_2 \sim$	$\mathcal{N}(0,1)$	$t'(\nu)$			$\mathcal{SN}'(0,1,\nu)$			ABM(ν)		
n		15	10	5	2	3	4	19/20	17/20	15/20
$\hat{S}$										
400	6.04	6.32	6.02	5.68	6.82	6.26	6.46	5.72	5.52	6.80
800	6.20	5.76	5.94	6.04	5.64	6.38	6.50	5.40	6.00	6.48
PML - LM										
400	4.78	5.28	5.26	6.60	5.64	5.60	5.72	5.54	5.32	4.76
800	5.32	5.62	5.68	7.64	4.84	5.32	4.92	5.20	5.84	5.16
GMM - LM										
400	9.28	9.68	10.02	10.06	9.24	8.78	9.24	11.62	10.70	10.42
800	8.22	7.96	7.74	9.18	7.62	7.72	7.92	12.26	10.92	9.08
PML - Wald										
400	13.12	6.66	5.12	2.14	9.84	8.38	8.20	31.82	26.12	16.92
800	9.26	3.60	2.12	2.40	5.96	4.78	3.66	34.18	27.48	13.64
GMM - Wald										
400	40.76	35.60	31.48	15.06	38.24	34.82	32.04	35.26	37.18	38.82
800	38.50	39.14	32.96	15.16	37.64	37.86	37.68	30.00	31.82	36.52

Based on 5000 Monte Carlo replications. $\hat{S}$ is the efficient score test.

▸ Densities - t ▸ Densities - SN ▸ Densities - ABM

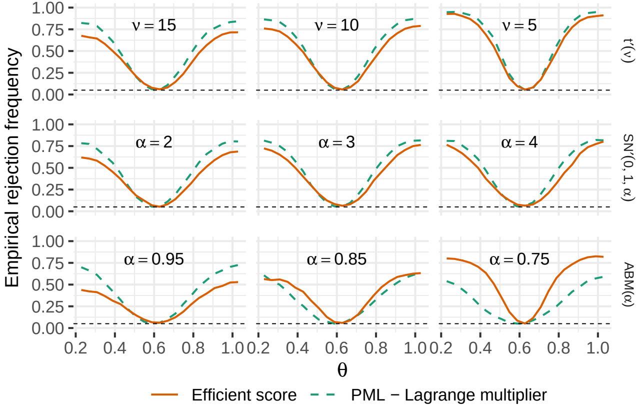
Figure 27: Empirical rejection frequency, $\epsilon_2 \sim N(0, 1)$



Simulation study 1

Independent components analysis - Monte Carlo

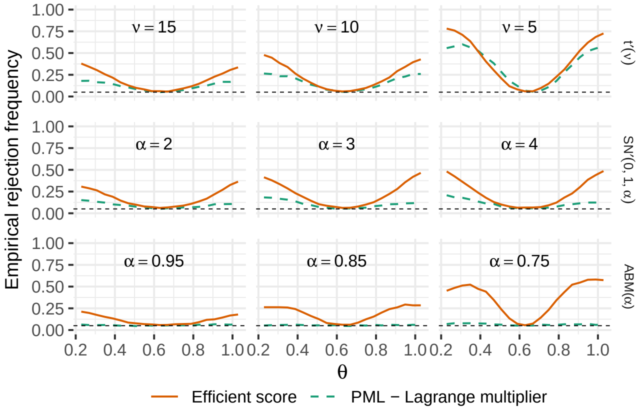
Figure 28: Empirical rejection frequency, $\epsilon_2 \sim t'(5)$



◂ Simulation study 1

Independent components analysis - Monte Carlo

Figure 29: Empirical rejection frequency, $\epsilon_2 \sim \mathcal{SN}'(0, 1, 4)$



◂ Simulation study 1

Pseudo-Maximum Likelihood approach

- Based on Gouriéroux et al. (2017) who proposed a similar approach in the SVAR setting
- Estimate (θ, σ) using pseudo-maximum likelihood
- Pseudo-densities: $t'(7)$ for ϵ_1, $t'(12)$ for ϵ_2
  - ▷ $t'(\nu)$ denotes a normalised t distribution:

$$\epsilon_k \sim t'(\nu) \quad \text{if} \quad \epsilon_k \sim \tilde{\epsilon}_k \Big/ \sqrt{\frac{\nu}{\nu - 2}} \quad \text{and} \quad \tilde{\epsilon}_k \sim t(\nu).$$

- Form standard pseudo-maximum likelihood Wald and LM tests

◂ Simulation study 2

GMM approach

- Based on Lanne and Luoto (2021) who proposed a similar approach in the SVAR setting
- Estimate (θ, σ) using GMM
- Moment conditions:

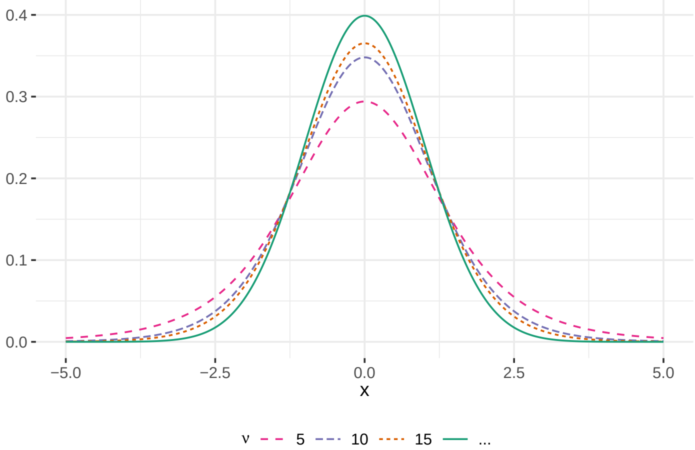
$$\mathbb{E}[\epsilon_k^2] = 1 \ (k = 1, 2), \quad \mathbb{E}[\epsilon_k^3 \epsilon_j] = 0, \quad \mathbb{E}[\epsilon_k^2 \epsilon_j^2] = 1 \quad \text{for} \quad k \neq j$$

  - ▷ OLS and variance normalisation
  - ▷ Asymmetric and symmetric co-kurtosis conditions
    - These co-kurtosis conditions provide no further information under Gaussianity
- Form standard GMM Wald and LM tests

◂ Simulation study 2

Independent components analysis - density shapes

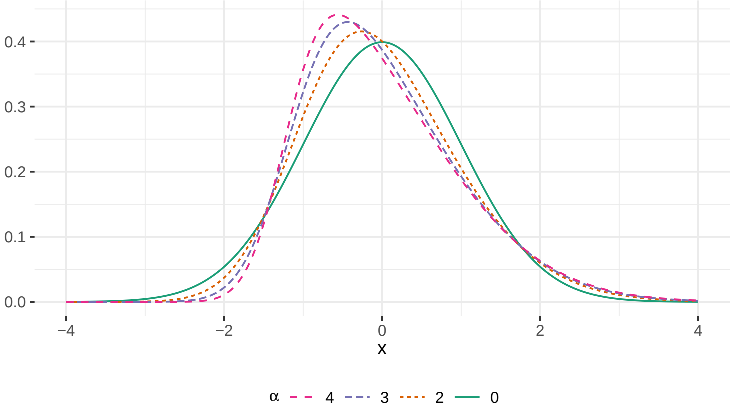
Figure 30: $t'(\nu)$ density



Simulation study 2 | Size table 1 | Size table 2 | Size table 3

Independent components analysis - density shapes

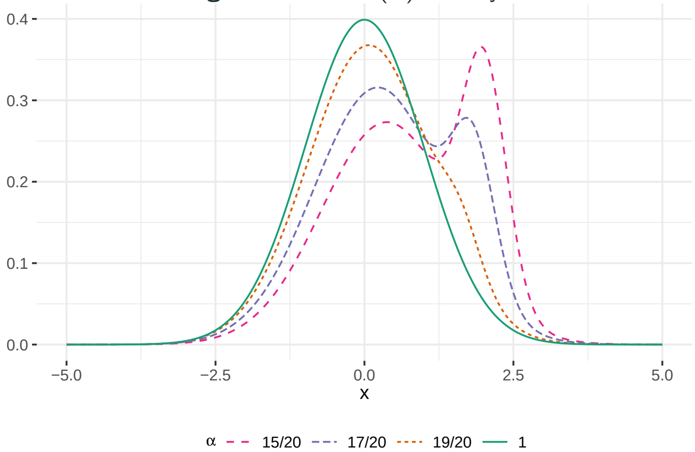
Figure 31: $\mathcal{SN}'(0,1,\alpha)$ density



Simulation study 2 | Size table 1 | Size table 2 | Size table 3

Independent components analysis - density shapes

Figure 32: ABM(α) density



◂ Simulation study 2 ◂ Size table 1 ◂ Size table 2 ◂ Size table 3