Identification of Nonlinear Time Series Models with Additive Noise

Francesco Cordoni[1] Nicolas Dorémus[2] Alessio Moneta[2]

[1] University of Pisa [2]Scuola Superiore Sant'Anna

25 August 2022

EEA-ESEM

Milano - Bocconi University

Introduction

- Statistical identificaton in a SVAR model have mainly focused on exploiting:
  - Heteroskedasticity (see e.g. Rigobon 2003, Normadin and Phaneuf 2004, Lanne and Saikkonen 2007).
  - Non-Gaussianity (see e.g. Gourieroux et al. 2017, Lanne et al. 2017, Herwartz 2018).
- In this paper, we emphasize another statistical property which can be used for the sake of identification, namely nonlinearity.

Outline

- How nonlinearity may contribute to identification
- Specific identification scheme (RESIT)
- Non Linear impulse response functions
- Simulation results

Statistical identification

- Statistical identification by non-Gaussianity is an application of ICA (Independent Component Analysis), where the other **key assumption is independence**.

- Identification by nonlinearity relies on a similar idea.

Statistical identification

Independence-based identification

Starting point: Linear VAR identification by non-Gaussianity

$$\begin{cases} y_t = \sum_{\ell=1}^{p} A_\ell y_{t-\ell} + u_t, \\ u_t = A_0 \varepsilon_t \end{cases}$$

- If ε_t is a vector of non-Gaussian and independent shocks, A_0 can be identified up to a (generalized) permutation matrix (see Eriksson and Koivunen 2004; Gourieroux et al. 2017).
- Additionaly assume recursiveness, i.e. A_0 essentialy triangular. Then (under mild conditions) A_0 is uniquely identified (see LiNGAM algorithm in Shimizu et al. 2006; Moneta et al. 2013).

Statistical identification

Independence-based identification

For the sake of illustration, consider the following bivariate model:

$$\boldsymbol{u_{t,1}} = \alpha \boldsymbol{u_{t,2}} + \varepsilon_{t,1}, \quad \varepsilon_{t,1} \perp\!\!\!\perp \boldsymbol{u_{t,2}}$$

Then there exists $\beta \in \mathbb{R}$ and a random variable $\varepsilon_{t,2}$ such that

$$\boldsymbol{u_{t,2}} = \beta \boldsymbol{u_{t,1}} + \varepsilon_{2t}, \quad \varepsilon_{t,2} \perp\!\!\!\perp \boldsymbol{u_{t,1}}$$

if and only if $\varepsilon_{t,1}$ and $u_{t,2}$ are Gaussian.
(Proof follows from Darmois-Skitovic theorem, on which ICA is based; see Peters et al. 2017).

- Thus, under non-Gaussianity, in a bivariate setting, $\varepsilon_{t,1} \perp\!\!\!\perp u_{t,2}$ allows detecting the correct causal direction (NB independence test accounting for higher order statistics, not just correlation)
- This can be extended to a multivariate setting
- But also to a **nonlinear** one.

Statistical identification

Nonlinear additive noise model

(Going nonlinear)

- **Definition**: the joint distribution $P(u_t)$ is said to admit an ANM from $u_{t,j}$ to $u_{t,i}$ if we have

$$u_{t,i} = f_i(u_{t,j}) + \varepsilon_{t,i}, \quad \varepsilon_{t,i} \perp\!\!\!\perp u_{t,j}$$

- **Identifiability of ANM**: Under mild assumptions (f_i 3-times differentiabile, strictly positive densities, etc.) if there is ANM from $u_{t,j}$ to $u_{t,i}$, then there is no backward ANM from $u_{t,i}$ to $u_{t,j}$.
  - Except for few cases, most remarkably, the Gaussian linear case (see Hoyer et al. 2008, Theorem 1, Peters et al. 2014, Prop. 23).

Statistical identification

Nonlinear additive noise model

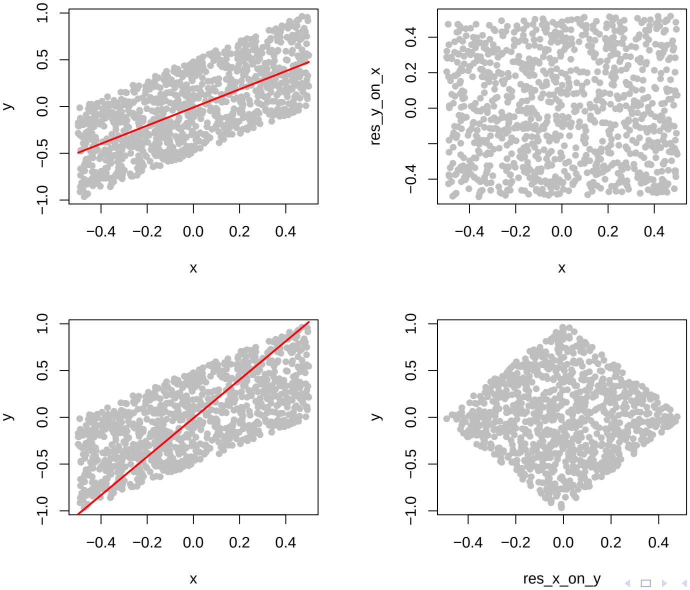
(Going nonlinear)

- **Definition**: the joint distribution $P(u_t)$ is said to admit an ANM from $u_{t,j}$ to $u_{t,i}$ if we have

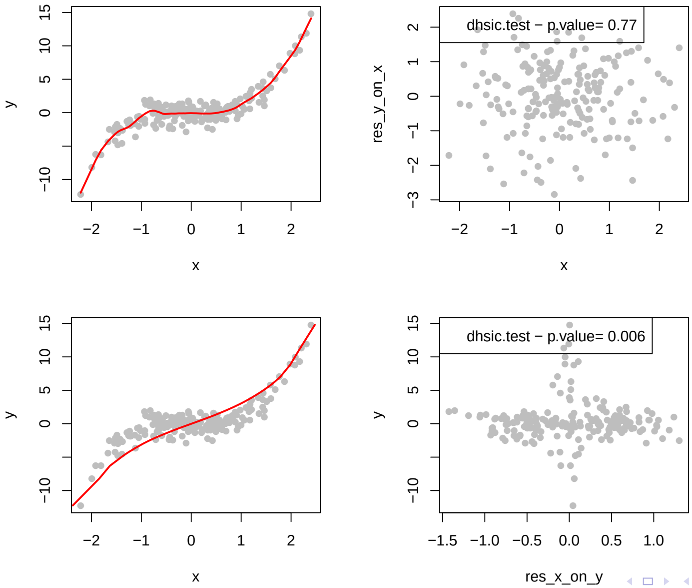
$$u_{t,i} = f_i(u_{t,j}) + \varepsilon_{t,i}, \quad \varepsilon_{t,i} \perp\!\!\!\perp u_{t,j}$$

- **Identifiability of ANM**: Under mild assumptions (f_i 3-times differentiabile, strictly positive densities, etc.) if there is ANM from $u_{t,j}$ to $u_{t,i}$, then there is no backward ANM from $u_{t,i}$ to $u_{t,j}$,
  - Except for few cases, most remarkably, the Gaussian linear case (see Hoyer et al. 2008, Theorem 1, Peters et al. 2014, Prop. 23).

Linear non-Gaussian case

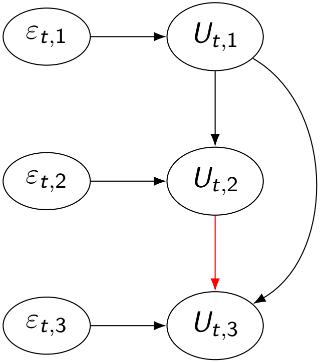


Nonlinear case



Statistical identification

DAG and SEM representations



$$
\begin{cases}
u_{t,1} &= \varepsilon_{t,1} \\
u_{t,2} &= \varphi_2(u_{t,1}) + \varepsilon_{t,2} \\
u_{t,3} &= \varphi_3(u_{t,1}, {\color{red}u_{t,2}}) + \varepsilon_{t,3}
\end{cases}
$$

- DAG and SEM representations of a generic ANM with recursive/topological ordering $u_{t,1}, u_{t,2}, u_{t,3}$.
- In red: elements to remove from each representation to eliminate the direct causation between $u_{t,2}$ and $u_{t,3}$.

Statistical identification

Nonlinear ANM and VAR

(Back to time series modelling)

- Let's consider this class of nonlinear VAR (see e.g. Kilian and Lütkepohl, 2017, ch. 18):

$$\begin{cases} y_t = F_t(y_{t-1}, \ldots, y_{t-p}) + u_t, \\ u_t = G_t(\varepsilon_t) \end{cases}$$

- Assume recursive causal structure among contemporaneous variables, i.e. it can be represented by a DAG (directed acyclic graph) over u_t, so that

$$u_{t,i} = \varphi_i(Pa(u_{t,i})) + \varepsilon_{t,i}$$

where $Pa(u_{t,i})$ is the set of graphical *parents* (i.e. direct causes) of $u_{t,i}$ (NB: only noise additivity is actually required)

Statistical identification

Nonlinear ANM and VAR

(Back to time series modelling)

- Let's consider this class of nonlinear VAR (see e.g. Kilian and Lütkepohl, 2017, ch. 18):

$$\begin{cases} y_t = F_t(y_{t-1}, \ldots, y_{t-p}) + u_t, \\ u_t = G_t(\varepsilon_t) \end{cases}$$

- Assume recursive causal structure among contemporaneous variables, i.e. it can be represented by a DAG (directed acyclic graph) over u_t, so that

$$u_{t,i} = \varphi_i(Pa(u_{t,i})) + \varepsilon_{t,i}$$

where $Pa(u_{t,i})$ is the set of graphical *parents* (i.e. direct causes) of $u_{t,i}$ (NB: only noise additivity is actually required)

Statistical identification

Nonlinear SVAR

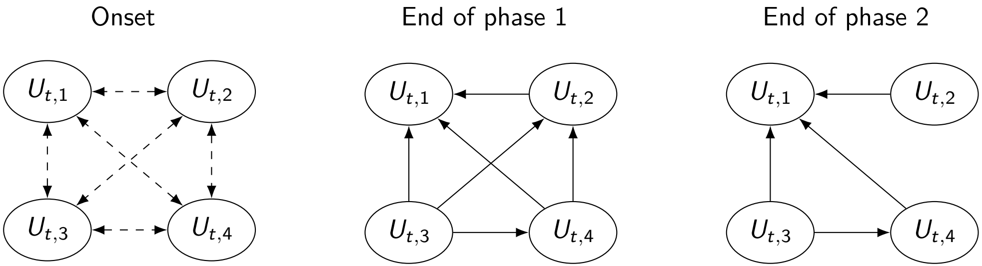
- Then the following nonlinear structural VAR is identifiable:

$$\begin{cases} y_t = F_t(y_{t-1}, \ldots, y_{t-p}) + u_t \\ u_{t,i} = \varphi_i(Pa(u_{t,i})) + \varepsilon_{t,i}, \quad \text{for } i \text{ in } 1, \ldots, k, \end{cases}$$

- Identifiability of the contemporaneous causal order follows from the requirement that in the true structural model, shocks are independent of covariates, i.e. $\varepsilon_{t,i} \perp\!\!\!\perp Pa(u_{t,i})$
- Any order different from the correct one would not satisfy this requirement

RESIT identification scheme

illustration: RESIT phases



RESIT identification scheme

Identification algorithm: Regression with subsequent independence test (RESIT); Peters et al. (2014).
Input: estimated reduced-form VAR residuals u_t.
Output: DAG

- **Phase 1**: determine a fully connected DAG (topological order within u_t).
  - Iterative procedure: in each step we identify a *sink* node. This is done, given a set of variables S, by regressing each variable in S on all the other variables in S and measuring dependence between residuals and covariates (p-value of HSIC independence test). The variable for which the corresponding residual display the weakest dependence on the covariates is denoted as sink and eliminated from S in the next step (until all the variables are ordered).
  - Main idea: in a DAG underlying an ANM for each node $u_{t,i}$ the noise $\varepsilon_{t,i}$ is independent of all non-effects of $u_{t,i}$

RESIT identification scheme

- **Phase 2**: remove superfluous edges.
- Main idea here: one edge from a putative cause x to an effect is superfluous if regressing the effect on the putative causes omitting x one still obtain residuals independent of covariates.
- Thus, if (i) in the DAG output of phase 1, $u_{t,i}$ is a child of $u_{t,j}$ and $u_{t,k}$, and if (ii) $\varepsilon_{t,i} \perp\!\!\!\perp u_{t,k}$, where $\varepsilon_{t,i}$ is obtained in a regression of $u_{t,i}$ on $u_{t,k}$ only (without $u_{t,j}$), then the edge, e.g., $u_{t,j} \rightarrow u_{t,i}$ is cut off.

Impulse Response Functions

Relaxing the linearity assumption implies a need for a more general IRF definition

$\Rightarrow$ Nonlinear impulse responses as differences of conditional expectations:

$$\mathrm{IRF}(h, \delta, \Omega_{t-1}) = E(y_{t+h}|\varepsilon_{t,i} = \delta, \Omega_{t-1}) - E(y_{t+h}|\Omega_{t-1}),$$

- where δ is the (positive or negative) magnitude of the shock $\varepsilon_{t,i}$ one wants to study, Ω_{t-1} is the history of the model data up to time $t-1$, and h is the horizon point up to which the impulse response functions are studied.

- Computed via Monte Carlo integration approach suggested in Kilian and Lutekopohl (2017, Ch. 18)

1. Statistical identification
2. RESIT identification scheme
3. IRF
4. Simulations
   - Causal learning
   - IRFs
5. Conclusions

Simulations

Simulations targeted at two distinct aspects

1. **Causal learning** (or causal discovery)
   - Assess the ability of RESIT to retrieve the true causal structure of the innovations
   - Comparisons with other causal discovery methods
2. Added value of **accounting for short-run non-linearities** for IRFs
   Contrast our IRF identification method with a standard benchmark (Choleski decomposition)

Simulations

Reduced form models for simulation

In all simulation exercises, we opted for simple **linear** VAR(1) models

$$y_t = A_1 y_{t-1} + u_t$$

Motivations:

- Emphasis on modelling contemporaneous structure
- Better use a simple/neutral time series model to highlight effects of contemporaneous linearities

Simulations

Causal learning

Goal: Assess the performance of the ANM principle (*near*-impossibility of backward model) in identifying the contemporaneous causal structure (sufficient for identification), formalized by a DAG.

Simulation model:

- Innovation structure characterized by a sparse (randomly drawn) DAG
- Corresponding dependencies embodied as ANM of two possible classes:
  - Linear effect and non-Gaussian noise (**NG**)
  - Non-linear effect randomly generated from a GAM-GP model (**NL**)

Simulations

Causal learning

- 5 **identification algorithms** (RESIT, LiNGAM, PC, CPC, RAND)
- Performance assessed by calculating distance between true DAG and inferred DAG via '**structural Hamming distance**' (SHD) and '**structural intervention distance**' (SID) (see Peters et al. 2014).

Simulations

Causal learning

Table: Average **SHD** between the estimated and the true DAG, varying the number of observations T For a selected model (columns) the average and standard errors (in round brackets) over 500 simulations are reported for each employed method (rows).

	T = 250		T = 500		T = 1000	
$k = 8$	VAR+NL	VAR+NG	VAR+NL	VAR+NG	VAR+NL	VAR+NG
RESIT	**6.14**	6.15	**5.82**	4.91	**6.57**	3.3
	(0.149)	(0.189)	(0.167)	(0.18)	(0.203)	(0.155)
LINGAM	8.71	**2.53**	9.67	**2.14**	10.53	**1.5**
	(0.143)	(0.094)	(0.16)	(0.098)	(0.178)	(0.094)
PC	6.79	5.57	6.73	5.16	6.75	4.23
	(0.124)	(0.115)	(0.127)	(0.106)	(0.133)	(0.104)
CPC	7.66	6.07	7.55	5.48	7.47	4.53
	(0.134)	(0.125)	(0.131)	(0.117)	(0.137)	(0.112)
RANDOM	16.21	16.05	15.94	15.51	16.21	16.2
	(0.24)	(0.237)	(0.244)	(0.234)	(0.233)	(0.231)

Simulations

Causal learning

Table: Average **SID** between the estimated and the true DAG, varying the number of observations T For a selected model (columns) the average and standard errors (in round brackets) over 500 simulations are reported for each employed method (rows).

	T = 250		T = 500		T = 1000	
$k = 8$	VAR+NL	VAR+NG	VAR+NL	VAR+NG	VAR+NL	VAR+NG
RESIT	**10.65**	13.3	**7.63**	9.71	**6.18**	5.4
	(0.372)	(0.5)	(0.346)	(0.415)	(0.327)	(0.31)
LINGAM	24.29	**7.15**	24.39	**5.76**	23.77	**3.5**
	(0.478)	(0.307)	(0.474)	(0.302)	(0.46)	(0.253)
PC	21.61	23.69	20.54	22.22	19.38	18.36
	(0.512)	(0.576)	(0.512)	(0.541)	(0.502)	(0.518)
CPC	26.9	26.04	25.67	23.44	23.92	19.64
	(0.582)	(0.628)	(0.575)	(0.595)	(0.541)	(0.578)
RANDOM	20.38	21.12	20.04	21.16	19.71	20.16
	(0.499)	(0.494)	(0.469)	(0.479)	(0.48)	(0.496)

1. Statistical identification
2. RESIT identification scheme
3. IRF
4. Simulations
   - Causal learning
   - IRFs
5. Conclusions

Simulations

IRFs

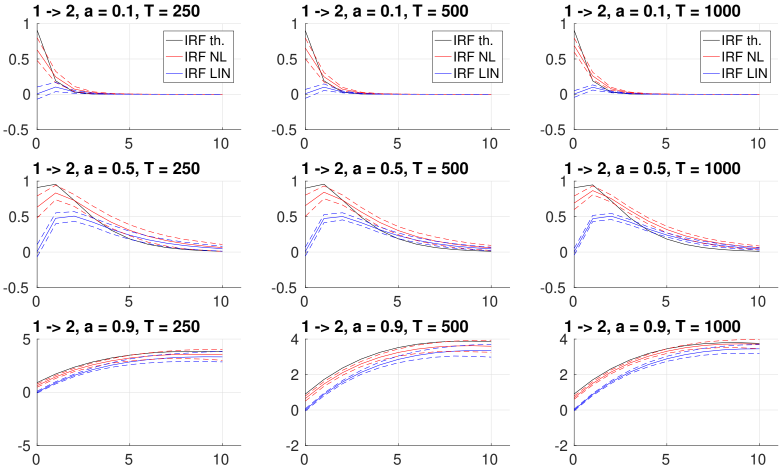
A **nonlinear** causal model for contemporaneous effects:

$$\begin{cases} u_{t,1} &= \varepsilon_{t,1} \\ u_{t,2} &= |u_{t,1}|^{\alpha} + \varepsilon_{t,2} \\ u_{t,3} &= sin(|u_{t,1}|^{\beta}) + \varepsilon_{t,3} \end{cases}$$

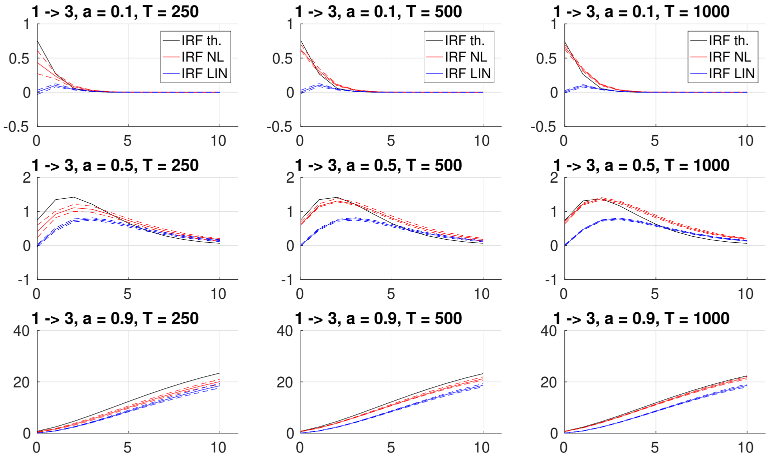
where:

- α, β are drawn from a uniform random variable with support $(1, 4)$, independently from each other
- $\varepsilon_{t,i}$ are the i.i.d. zero-mean uncorrelated Gaussian structural shocks

NB: $u_{t,3} \leftarrow u_{t,1} \rightarrow u_{t,2}$



Theoretical, linear and non-linear IRFs by varying the parameter a and number of observations T. Red (Blue) lines exhibits the average (among 200 simulations) for the non-linear (linear) IRFs response for the **response of variable 2** to a unitary shocks of $\varepsilon_{t,1}$, confidence interval at 68% are reported in dotted lines.



Theoretical, linear and non-linear IRFs by varying the parameter a and number of observations T. Red (Blue) lines exhibits the average (among 200 simulations) for the non-linear (linear) IRFs response for the **response of variable 3** to a unitary shocks of $\varepsilon_{t,1}$, confidence interval at 68% are reported in dotted lines.

1. Statistical identification
2. RESIT identification scheme
3. IRF
4. Simulations
   - Causal learning
   - IRFs
5. Conclusions

Conclusions

- Identification method:
  - Nonlinearity instrumental for identification in ANM, with similarities and differences with non-Gaussianity in the linear case.
  - RESIT scheme based on independence test between residuals and covariates.
- Simulations:
  - Showed the merits of RESIT approach applied to reduced form residuals
  - Highlighted the discrepancies Linear and non-linear IRF specifications
- Limitations and future research:
  - IRF estimation approach meant to cover the whole class of model

    → Probably better solutions in more specific settings

Thank you for listening

Conclusions

Bonus Slide

Table: Average (%) contemporaneous causal structural estimated by RESIT by varying T. Computed by averaging on the persistence parameters a over 200 simulations.

Var.	T 250	Pa(1)	Pa(2)	Pa(3)	T 500	Pa(1)	Pa(2)	Pa(3)
1		0	48.83	55.83		0	74.17	69.33
2		0.33	0	11.00		0.83	0	13.50
3		1.50	15.50	0		0.33	8.17	0

Var.	T 1000	Pa(1)	Pa(2)	Pa(3)
1		0	91.17	91.67
2		0.50	0	5.33
3		0.83	5.83	0