

INSTITUTE
OF ECONOMICS



Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

Identification of Vector Autoregressive Models with Nonlinear Contemporaneous Structure

Francesco Cordonì ^a

Nicolas Dorémus ^b

Alessio Moneta ^b

^a Royal Holloway University of London, UK.

^b Institute of Economics and EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy.

2023/07

January 2023

ISSN(ONLINE) 2284-0400

Identification of Vector Autoregressive Models with Nonlinear Contemporaneous Structure

Francesco CORDONI¹

Nicolas DORÉMUS²

Alessio MONETA³

February 20, 2023

Abstract

We propose a statistical identification procedure for structural vector autoregressive (VAR) models that present a nonlinear dependence (at least) at the contemporaneous level. By applying and adapting results from the literature on causal discovery with continuous additive noise models to structural VAR analysis, we show that a large class of structural VAR models is identifiable. We spell out these specific conditions and propose a scheme for the estimation of structural impulse response functions in a nonlinear setting. We assess the performance of this scheme in a simulation experiment. Finally, we apply it in a study on the effects of monetary policy on the economy, which shows the importance of taking into account the potential asymmetry between the impact of an expansionary versus a contractionary monetary policy.

Keywords: Structural VAR models, Impulse Response Functions, Causal Discovery, Nonlinearity, Additive Noise Models.

JEL classification: C32, C52, E52.

¹Department of Economics, Royal Holloway University of London. Email: Francesco.Cordoni@rhul.ac.uk

²Institute of Economics, Scuola Superiore Sant'Anna, Pisa. Email: nicolas.doremus@santannapisa.it

³Institute of Economics, Scuola Superiore Sant'Anna, Pisa. Email: a.moneta@santannapisa.it

The authors acknowledge support from the project “How good is your model? Empirical evaluation and validation of quantitative models in economics” funded by PRIN grant no. 20177FX2A7. We thank Mario Martinoli and Sebastiano Michele Zema for comments on an earlier version of the paper. We are also grateful to the participants of the Italian Workshop of Econometrics and Empirical Economics (Rimini 2022), the Econometric Society European Meeting (Milan 2022), and the Workshop “Model Evaluation and Causal Search” (Pisa 2022).

1 Introduction

Since the seminal work of Reiersøl (1950), it is well known in the econometric literature that linear additive noise models with Gaussian and spherical errors are not identifiable, unless one uses external instruments or a priori restrictions. Structural vector autoregressive (VAR) models, widely used in empirical macroeconomic research since Sims (1980), share the same problem of identification. This has been addressed, in the literature, by relying on constraints derived from economic theory (i.e., short-run, long-run, and sign restrictions, as pioneered, respectively, by Sims (1980), Blanchard and Quah (1989), and Faust (1998)), by using external instruments (from Romer and Romer (1989) to Montiel Olea et al. (2021)), and, finally, by exploiting specific statistical properties of the data that are at odds with the standard model.

This paper contributes to the statistical identification approach. As regards this approach, one finds two strands of the structural VAR literature. One of them exploits the non-Gaussianity of the errors, the other exploits heteroskedasticity. In common, they have the use of higher moments for identification (Montiel Olea et al., 2022; Herwartz et al., 2022). This is possible thanks to the introduction of assumptions that depart from the standard model. In this contribution, we want to show that another swerve from the linear model with spherical-normal errors can be exploited for the identification of structural VAR models, namely nonlinearity.

Identification by non-Gaussianity is achieved by the application of ideas and techniques developed in the literature on Independent Component Analysis (ICA), a research field that emerged at the intersection of statistics with signal processing (Comon, 1994; Hyvärinen et al., 2001). Studies in the machine learning literature have shown that ICA, rather unexpectedly at the beginning, can be used for causal inference, and more specifically, for the identification of linear structural equation models, under the assumption of noise independence and non-Gaussianity, as well as the presence of a recursive (i.e. acyclic) structure among the variables (see, e.g., Shimizu et al., 2006, 2011; Hyvärinen, 2013). Moneta et al. (2013) have applied some of these techniques to structural VAR analysis. The structural VAR-ICA literature has been rapidly growing in the recent years, with contributions by Lanne and Lütkepohl (2010); Lanne et al. (2017); Gouriéroux et al. (2017); Herwartz (2018); Fiorentini and Sentana (2022), among others. It is important to note that the recursiveness assumption has been relaxed by these studies, and that the impact matrix of the shocks is identified up to the post-multiplication of a generalized permutation matrix, which means, in practice, that one is able to label the shocks only after involving ex post some form of economic reasoning, jointly with an inspection, for instance, of the obtained impulse response functions. Recent studies are also relaxing the independence assumption, maintaining non-Gaussianity and use of higher moments (Lanne and Luoto, 2021; Guay, 2021; Mesters and Zwiernik, 2022). Identification by heteroskedasticity was pioneered by Sentana and Fiorentini (2001) and further pursued by Rigobon (2003); Lanne and Lütkepohl (2008); Sims (2020); Brunnermeier et al. (2021); Lewis (2021), among others.¹

¹Identification by heteroskedasticity is a “higher-moment” procedure in the sense of using information from conditional (instead of unconditional) second moments (Montiel Olea et al., 2021), but its underlying model is not, differently from the ICA model, part of the general class of models we consider in this paper.

We claim that nonlinearity, analogously to non-Gaussianity and heteroskedasticity, can be exploited for structural VAR identification. This idea has been much less explored in the econometric literature and we believe that it is important to trigger a novel discussion on this. We put forth this idea by — similarly to what has been done in the VAR-ICA approach — importing techniques from the machine learning literature on causal discovery (see the review by Peters et al., 2017) and by presenting a class of autoregressive processes in which it is possible to recover structural shocks (without any problem of labelling) and their dynamic effects on variables of interests directly from the data. In tune with the causal discovery literature, we frame the problem of identifying structural shocks and their contemporaneous impacts in a structural VAR model as a particular case of the general problem of learning a causal structure (which can be usefully represented by a graph) from an observational joint distribution. Hoyer et al. (2008) and Peters et al. (2014) have shown that if the observational distribution follows a structural equation model with an additive noise structure, then the causal graph becomes identifiable from the distribution under specific assumptions. Peters et al. (2014) also provide practical algorithms to retrieve the causal structure from finite samples.

It is noteworthy that, in the causal discovery framework we introduce, identification by non-Gaussianity and identification by nonlinearity are nested within the general framework of causal discovery with additive noise models. In the first case the key assumptions are, as noted above, shocks' non-Gaussianity plus independence, while in the second case nonlinearity in the shock transmission mechanism plus (again) independence. A related contribution of this paper is to shed light on the connection between these two related identification methods. Furthermore, we spell out the conditions that allow identification where the underlying model is a VAR process with nonlinear dependence at the contemporaneous level. Such a process may or may not show nonlinearities in the autoregressive structure, but the functional autoregressive form is not key for identification. Similarly to the first wave of applications of ICA to VAR analysis (Moneta et al., 2013; Guerini and Moneta, 2017), however, the recursiveness assumption is here required.

We also provide a practical scheme (two algorithms) to recover the structure linking shocks to (reduced-form) VAR innovations and, on the basis of this, to estimate nonlinear structural impulse response functions from time series data. Apart from the role it plays in identification, there are many empirical cases in which the presence of nonlinearity in the processes underlying observed data should not be underestimated. As Ramey (2016) points out, “positive shocks might have different effects from negative shocks, effects might not be proportional to the size of the shock, or the effect of a shock might depend on the state of the economy when the shock hits.”

In the next section we present the model set up and the identification methodology, which includes both theoretical results and practical algorithms. In section 3 we present the results of a simulation study. Section 4 presents an empirical application. Section 5 concludes.

2 Identification methodology

In this section, we present our method to achieve statistical identification of a specific class of VAR models, namely VAR models with generic time dependence and nonlinear but recursive contemporaneous causal structure.

In section 2.1, we introduce this class of models, which we call generalized VAR models with additive noise innovations. This class of VAR models is characterised by a recursive causal structure among innovation terms, which can be represented by a directed acyclic graph (DAG). As established in the literature, identifying such causal structure is key for identification, since it allows transforming the reduced-form model into a structural model with independent shocks.

In section 2.2, we show that the contemporaneous recursive structure among the innovation terms can be recovered by exploiting nonlinearity. In section 2.3, we present an algorithm which infers such structure from estimated reduced-form VAR innovations. In section 2.4, we introduce an algorithm to estimate structural impulse response functions.

2.1 Model setup

We consider here a general class of VAR models (see, e.g., the nonlinear structural VAR models in Kilian and Lütkepohl, 2017, ch. 18), in which a vector of K time series variables y_t depends on its lags with an additive vector of disturbances u_t :

$$y_t = F_t(y_{t-1}, \dots, y_{t-L}) + u_t, \quad (1)$$

where $F_t(\cdot)$ is a generic (possibly time dependent) nonlinear or linear function, and u_t is a zero-mean i.i.d. vector and depends only on contemporaneous structural shocks $\varepsilon_{t,1}, \dots, \varepsilon_{t,K}$:

$$u_t = G(\varepsilon_t) \quad (2)$$

We assume that equation (2), for an appropriate ordering, can be rewritten as:

$$u_{t,k} = \varphi_k(\underline{u}_{t,k}) + \varepsilon_{t,k}, \quad \text{for } k = 1, \dots, K \quad (3)$$

where $\underline{u}_{t,k}$ is a subset of the variables $u_{t,1}, \dots, u_{t,k-1}$, and $\varphi_k(\emptyset) = 0$. In other words, u_t can be arranged in a recursive order. We also assume that $\varepsilon_{t,1}, \dots, \varepsilon_{t,K}$ are cross-sectionally mutual independent and i.i.d. Henceforth, we will refer to the model as formalized in equations (1-3) as the *generalized VAR with additive noise innovation model* (GVAR-ANIM).

Equations (1) - (2) can be seen as a generalization of the standard-linear structural VAR model:

$$y_t = \sum_{\ell=1}^L A_{\ell} y_{t-\ell} + A_0 \varepsilon_t, \quad (4)$$

where

$$u_t = A_0 \varepsilon_t. \quad (5)$$

If $\varepsilon_{t,1}, \dots, \varepsilon_{t,K}$ are mutually independent and non-Gaussian, equation (5) is an ICA model and the matrix A_0 can be identified up to a post-multiplication of a generalized permutation matrix (Eriksson and Koivunen, 2004; Gouriéroux et al., 2017). If, for any permutation matrix P , the matrix $PA_0^{-1}P^T$ is lower triangular (i.e. A_0^{-1} is *essentially triangular*), under this assumption the matrix A_0^{-1} is uniquely identified (Shimizu et al., 2006)². We call the assumption on A_0^{-1} *recursiveness* assumption, since it imposes a recursive causal structure on the contemporaneous variables of the structural VAR model: if this assumption is true, one can re-order the variables entering in y_t in a “Wold causal chain” (Wold, 1960), so that each variable $y_{t,i}$ causes $y_{t,j}$ and no variable $y_{t,j}$ causes $y_{t,i}$ ($i < j$, for i, j in $1, \dots, K$).

If A_0^{-1} is essentially triangular, a convenient way to represent the u_t from equation (5) is with a directed acyclic graph (DAG) (see Spirtes et al., 2000; Pearl, 2009). Let us suppose, only for the sake of illustration, that $PA_0^{-1}P^T$ is lower triangular for $P = I$ and $K = 3$. Then we have this system of structural equations between the innovation terms:

$$\begin{cases} u_{t,1} &= \varepsilon_{t,1} \\ u_{t,2} &= \alpha u_{t,1} + \varepsilon_{t,2} \\ u_{t,3} &= \beta u_{t,1} + \gamma u_{t,2} + \varepsilon_{t,3}, \end{cases} \quad (6)$$

which is represented by the DAG in Figure 1.

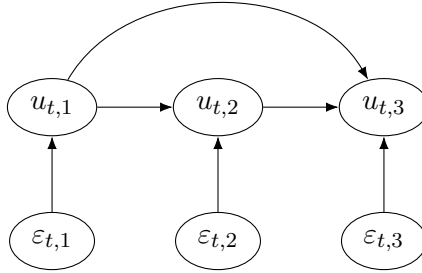


Figure 1: Example of DAG between innovation terms.

If the true DAG were known, one would be able to put (just-identifying or over-identifying) zero restrictions on A_0^{-1} and eventually to recover the independent shocks ε_t (Swanson and Granger, 1997; Demiralp and Hoover, 2003; Moneta, 2008).

Knowing the causal structure is key for identification also in the general nonlinear case of equations (1) - (2). Let us consider the following instance of equation (2), which we have assumed that can be arranged in a recursive order with noises entering in an additive fashion:

$$\begin{cases} u_{t,1} &= \varepsilon_{t,1} \\ u_{t,2} &= \varphi_2(u_{t,1}) + \varepsilon_{t,2} \\ u_{t,3} &= \varphi_3(u_{t,1}, u_{t,2}) + \varepsilon_{t,3} \end{cases} \quad (7)$$

²Shimizu et al. (2006) prove identifiability of linear, non-Gaussian and recursive model (LiNGAM in their terminology) using ICA and, specifically, the Darmois-Skitovich theorem (Comon, 1994, Theorem 11) (see also Peters et al., 2017, Theorem 7.6).

The structure of equation (7) can also be represented by the DAG of Figure 1. Knowing the DAG, or having a method that infers it, allows us to recover the shocks and the manner they impact innovations from observed realizations of u_t . This can be done in a recursive way. Suppose the DAG in Figure 1 is true. Then, following equation (7), one can (i) assign $\varepsilon_{t,1}$ to $u_{t,1}$; (ii) run an adequate (e.g., nonparametric) regression of $u_{t,2}$ on $u_{t,1}$ and get $\varepsilon_{t,2}$ as residual; (iii) regress $u_{t,3}$ on $u_{t,1}$ and $u_{t,2}$ and get $\varepsilon_{t,3}$ as residual. In the next subsections, we will present a method to recover the DAG that represents the causal structure at stake from u_t in the nonlinear case (in the framework of GVAR-ANIM, i.e. in the nonlinear with additive noise case).

2.2 Exploiting nonlinearity

The causal discovery principle that we exploit here was first introduced in the literature by Hoyer et al. (2008). This can be seen as a generalization of the principle underlying causal inference in the linear non-Gaussian case. Let us first show this case, starting from a two-variables model.

Consider a bivariate VAR model, whose vector of reduced-form residuals is $u_t = (u_{t,1}, u_{t,2})$. In this sub-section, to improve readability, we drop the time-subscript t when it is evident from the context and is not relevant, so that, for instance, $u \equiv u_t$ and $u_i \equiv u_{t,i}$. It holds this result, proven by Peters et al. (2017, Theorem 4.2):

Theorem 1³ Consider the following linear model:

$$u_1 = \alpha u_2 + \varepsilon_1, \quad \varepsilon_1 \perp\!\!\!\perp u_2. \quad (8)$$

Then there exists $\beta \in \mathbb{R}$ and a random variable ε_2 such that

$$u_2 = \beta u_1 + \varepsilon_2, \quad \varepsilon_2 \perp\!\!\!\perp u_1 \quad (9)$$

if and only if ε_1 and u_2 are Gaussian (where $\perp\!\!\!\perp$ denotes statistical independence).

As a corollary of this theorem, it follows that in a bivariate linear structural VAR, as the one in equation (4) but with $K = 2$ and non-Gaussian shocks, if ε_1 is independent of u_2 , then equation (8) is the true structural model and the causal structure can be represented by the DAG $u_2 \rightarrow u_1$.

Shimizu et al. (2011) extend this result to a multivariate framework. But let us focus here on another possible extension, namely the generalization of this result for the nonlinear case, under the condition that we preserve the noise-additivity assumption. It turns out that, in the strictly nonlinear case, Gaussianity does not preclude identification any longer. Following the terminology by Peters et al. (2017, Definition 4.4), let us define this property:

Definition (bivariate ANM): The joint distribution $P(u)$ is said to admit a (bivariate) ANM from u_i to u_j if, for any measurable function φ_j and a variable ε_j we have:

$$u_j = \varphi_j(u_i) + \varepsilon_j, \quad \varepsilon_j \perp\!\!\!\perp u_i \quad (10)$$

³The proof of this theorem hinges heavily on the Darmois-Skitovich Theorem (see Peters et al., 2017, Theorem 4.3).

Identifiability of a bivariate ANM is based on the following theorem proven by Hoyer et al. (2008, Theorem 1) (see also Peters et al., 2014, 2017):

Theorem 2 Let us assume that $P(u)$ admits an ANM from u_i to u_j and that ε_j and u_i have strictly positive densities $P(\varepsilon_j)$, and $P(u_i)$, with φ_j , $P(\varepsilon_j)$, and $P(u_i)$ three times differentiable. To simplify notation, let $f := \varphi_j$, and $\xi := \log P(u_i)$ and $\nu := \log P(\varepsilon_j)$, skipping the arguments $u_j - f(u_i)$, u_i and u_i for ν , ξ and f and their derivatives, respectively. Consider the following condition (see Peters et al., 2014, Condition 19).

Condition (C1): the triple $(\varphi_j, P(u_i), P(\varepsilon_j))$ does not satisfy the following differential equation for all u_i, u_j with $\nu''(u_j - f(u_i))f'(u_i) \neq 0$:

$$\xi''' = \xi'' \left(-\frac{\nu''' f'}{\nu''} + \frac{f''}{f'} \right) - 2\nu'' f'' f' + \nu' f''' + \frac{\nu' \nu''' f'' f'}{\nu''} - \frac{\nu' (f'')^2}{f'}, \quad (11)$$

If condition C1 is satisfied, then a backward ANM from u_j to u_i , i.e. $u_i = \varphi_i(u_j) + \varepsilon_i$ with $\varepsilon_i \perp\!\!\!\perp u_j$, is not admitted.

As shown by Hoyer et al. (2008), a “generic” triple $(\varphi_j, P(u_i), P(\varepsilon_j))$ is expected to satisfy condition C1. Zhang and Hyvärinen (2009) actually provide an exhaustive list of five settings that admit an ANM in both the forward and backward direction, the most remarkable case being the linear Gaussian case. Notice that if a joint distribution admits a (bivariate) ANM in one direction (say from u_j to u_i), but not in the backward direction, the DAG representing the ANM is identifiable ($u_j \rightarrow u_i$).

Figure 2 shows results from an illustrative simulation (see also Peters et al., 2017), in which data x and y are generated from a structural model $y = \beta x + N_y$, with x and N_y following a uniform distribution, with $x \perp\!\!\!\perp N_y$. In the top left panel, we show the results of a linear regression of y on x , whose residuals are plotted over x in the top right panel. It is easy to notice the lack of dependence between such residuals and x . The bottom left panel show the regression with exactly the same data, but performed in the opposite direction, namely we regress x on y . Residuals of such regression are plotted over y (the covariate in the new regression) in the bottom right panel. It is easy to note a (higher-order) statistical dependence. This asymmetry can also be detected in the nonlinear generic case. Figure 3 shows results from another illustrative simulation in which data x and y are generated from a structural model $y = x^3 + N_y$, with x and N_y following a normal distribution, with $x \perp\!\!\!\perp N_y$. In the top left panel, we show the results of kernel regression of y on x , whose residuals are plotted over x in the top right panel. The lack of dependence between such residuals and x is confirmed by a nonparametric independence test, specifically the Hilbert-Schmidt Independence Criterion proposed by Gretton et al. (2007). The p-value for the null hypothesis of independence between residuals and covariate is reported inside the plot. Bottom left panel show the kernel regression with exactly the same data, but performed in the opposite direction, namely we regress x on y . Residuals of such kernel regression are plotted over y (the covariate in the new regression) in the bottom right panel. The p-value for the null hypothesis of independence between residuals and covariate is reported inside the plot and clearly suggests to reject independence.

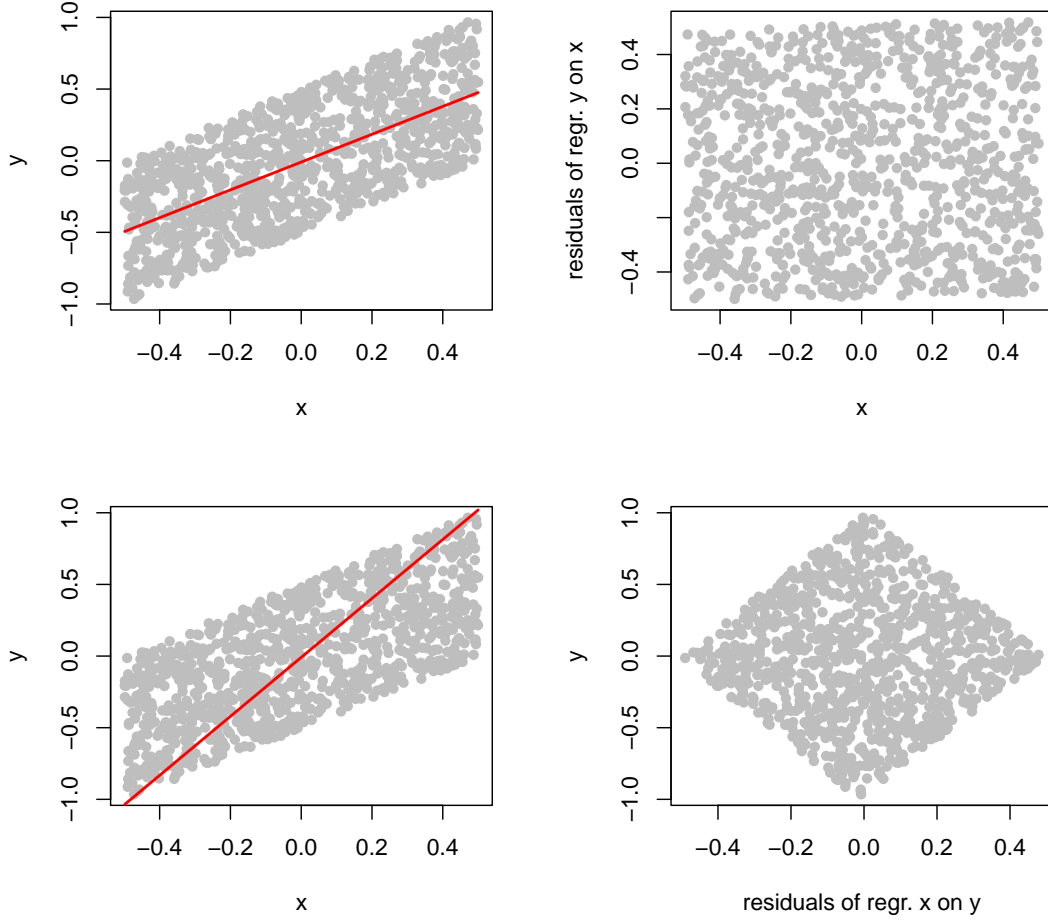


Figure 2: Illustration of the ANM principle with a sample generated from the linear model $y = \beta x + N_y$ with x and N_y independent and drawn from a uniform distribution. Left-side panels: scatter plot of y on x with OLS regression line (red) corresponding to the ‘forward’ (true) specification (top panel) and to the ‘backward’ specification (bottom panel). Right-side panels: corresponding regression residuals are plotted with respect to their regressor values, either x (top panel) or y (bottom panel).

Peters et al. (2014) proved that the identifiability result stated in Theorem 2 can be extended from the bivariate to the multivariate case. First of all, the definition of ANM given above can be straightforwardly extended.

Definition (multivariate ANM). We call a system of K structural recursive equations a (multivariate) ANM if it can be written as:

$$u_k = \varphi_k(Pa(u_k)) + \varepsilon_k \quad \text{for } k \text{ in } 1, \dots, K, \quad (12)$$

where $\varepsilon_1, \dots, \varepsilon_K$ are mutually independent. The set of variables $Pa(u_k)$, called *parents* of u_k , is defined as $Pa(u_k) \subseteq \{u_1, \dots, u_K\} \setminus \{u_k\}$. We denote by \mathcal{G} the DAG representing the structural

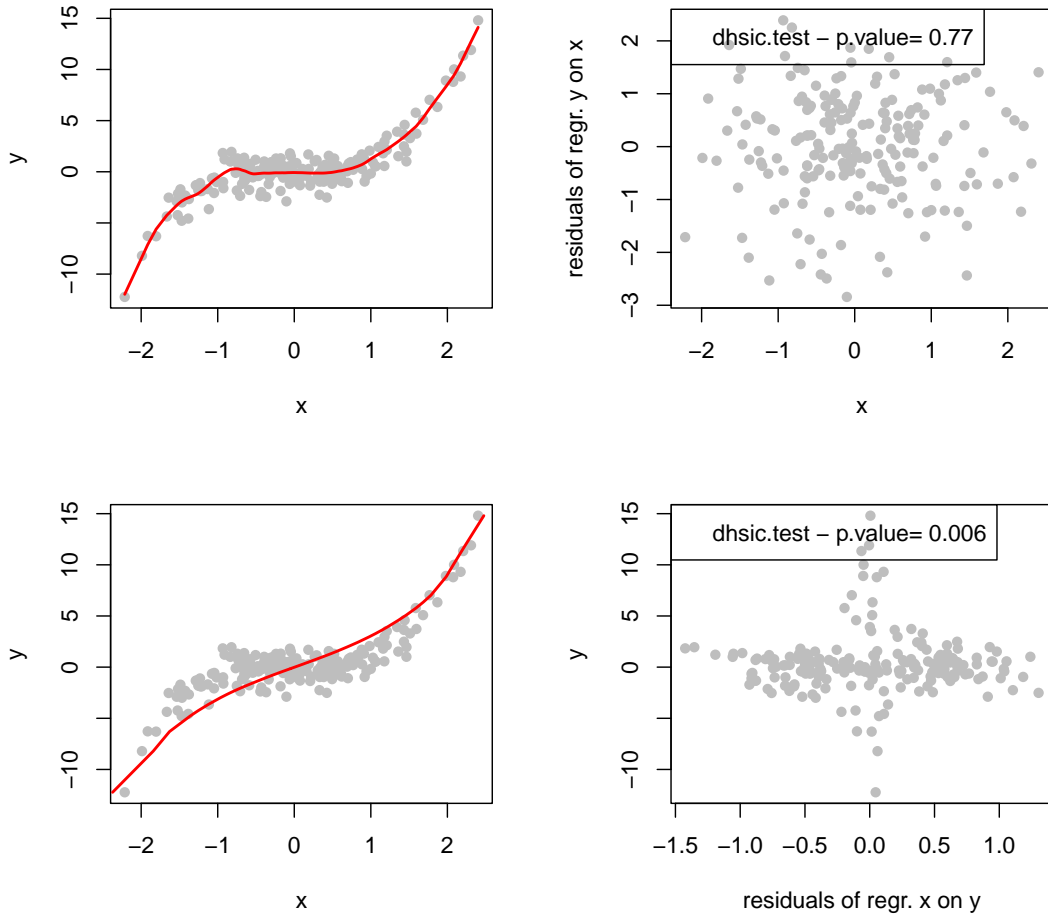


Figure 3: Illustration of the ANM principle with a sample generated from the nonlinear model $y = x^3 + N_y$ with x and N_y independent and drawn from a normal distribution. Left-side panels: scatter plot of y on x with kernel regression line (red) corresponding to the ‘forward’ (true) specification (top panel) and to the ‘backward’ specification (bottom panel). Right-side panels: corresponding regression residuals are plotted with respect to their regressor values, either x (top panel) or y (bottom panel).

relations between the u 's, i.e. $u_i \longrightarrow u_j$ iff $u_i \in Pa(u_j)$.

Notice that \mathcal{G} is a DAG since the system is recursive. Identifiability of the multivariate ANM is based on the following theorem:

Theorem 3. Consider a multivariate ANM

$$u_k = \varphi_k(Pa(u_k)) + \varepsilon_k \quad \text{for } k \text{ in } 1, \dots, K, \quad (13)$$

associated to DAG \mathcal{G} and entailing joint distribution $P(u)$, with φ_k three times differentiable and $P(\varepsilon_k)$ strictly positive for all k . An alternative ANM with DAG \mathcal{G}' , where $\mathcal{G}' \neq \mathcal{G}$, also entailing $P(u)$, is not admitted under the following conditions:

Condition (C2) The functions $\varphi_1(\cdot), \dots, \varphi_K(\cdot)$ are not constant in any of their arguments.

Condition (C3) Let us denote with $ND(u_j)$ the set of graphical “non-descendant” of u_j , i.e. there is no directed path from u_j to any variable in $ND(u_j)$ (see Spirtes et al., 2000). For all $u_j \in u$, and for all $u_i \in Pa(u_j)$, and all $u_S \in u$ such that $Pa(u_j) \setminus u_i \subseteq u_S \subseteq ND(u_j) \setminus \{u_i, u_j\}$, there is a value $\overline{u_S}$ of u_S with $p(\overline{u_S}) > 0$ such that the triple

$$\left(\varphi_j(\overline{Pa(u_j) \setminus \{u_i\}}, u_i), p(u_i | \overline{u_S}), p(\varepsilon_j) \right) \quad (14)$$

satisfies condition C1. In equation (14) the upper bar indicates a specific value that a random variable or vector takes.

The conditions related to Theorem 3 and its proof can be found in (Peters et al., 2014, see section 3.2 and appendix A therein). Theorem 3 allows us to recover the contemporaneous causal structure among the innovation terms in the GVAR-ANIM. In the next subsection, we will present an algorithm to learn such causal structure from estimated reduced-form VAR residuals.

Topological order

As shown above, knowing the causal structure among innovation terms in the form of a DAG allows specifying the structural model, since the set of graphical parents determines the arguments of $\varphi_k(\cdot)$ in equation (12). Structural shocks can be recovered by regression (parametric or non-parametric) methods applied to the same equation.

Let us now introduce the notion of topological order. Given a DAG over vertices v_1, \dots, v_K , it is possible to associate a total order on its vertices such that if v_i is a parent of v_j (i.e. there is a directed edge from v_i to v_j), then v_i comes before v_j in the ordering. Notice that there is only one topological order associated to a fully connected DAG (a DAG in which each pair of vertices is connected). But, in the case of a sparse DAG, there are possibly multiple topological orders associated to it. Also notice that removing an edge from a fully connected DAG does not invalidate the original topological order, but there could be alternative ordering compatible with it.⁴

Suppose that a set of structural recursive equations, associated to a graph \mathcal{G}_0 , has generated some data, and suppose that there are multiple topological orders associated to \mathcal{G}_0 . One can deduce a unique fully connected DAG, which we call \mathcal{G}^{max} , for each of these topological orders. It may be the case $\mathcal{G}_0 \neq \mathcal{G}^{max}$. But \mathcal{G}^{max} differs from \mathcal{G}_0 only in displaying extra edges, i.e. \mathcal{G}_0 is a subgraph of \mathcal{G}^{max} . Therefore, if we write down a set of structural equations (as in equation 12) associated to \mathcal{G}^{max} , this set of equations will differ from the “true” one, i.e. the one associated to \mathcal{G}_0 , only in that it includes extra variables in right-hand side. These extra-variables are not in a parent set in (the structural model associated to) \mathcal{G}_0 . But since the original causal ordering (associated to \mathcal{G}_0) is maintained, no endogeneity issues arise when one uses \mathcal{G}^{max} as a template

⁴Consider the fully connected DAG on v_1, v_2, v_3 associated to the topological order $\langle v_1, v_2, v_3 \rangle$. From removing $v_1 \rightarrow v_3$, we get $v_1 \rightarrow v_2 \rightarrow v_3$, and the original topological order $\langle v_1, v_2, v_3 \rangle$ is still the only one valid. From removing $v_2 \rightarrow v_3$, we get $v_2 \leftarrow v_1 \rightarrow v_3$, and there are two possible topological orders: $\langle v_1, v_2, v_3 \rangle, \langle v_1, v_3, v_2 \rangle$.

for a regression model. Indeed, if one applies regression methods to the set of equations associated to \mathcal{G}^{max} , which may therefore include extra variables (with respect to the “true” structural model) on the right hand side, one will then get consistent estimates, since these variables are at most irrelevant, but do not entail reverse causality problems.

In practice, if the aim of the study is to estimate the structural shocks (e.g., in impulse response function analysis), one may aim at recovering a topological order instead of a DAG, if the first task turns out to be a preliminary step of the second and one wants to minimize inference errors in finite samples. Instead, if causal discovery is the goal, then a topological order is not enough. This is why in the next section we will present a search algorithm that can be stopped at the topological order step.

A simple case

While the paper, including the algorithms presented in the next subsections, covers a wide class of models that we have labelled GVAR-ANIM (see equations 1-3), in both the simulation and the empirical analysis, we are going to focus on a simpler case, namely the case of linear lagged relationships with nonlinear contemporaneous relationships between innovations (with additive structural shocks). In terms of equations:

$$\begin{cases} y_t &= \sum_{\ell=1}^L A_{\ell} y_{t-\ell} + u_t, \\ u_{t,k} &= \varphi_k(Pa(u_{t,k})) + \varepsilon_{t,k} \quad \text{for } k \text{ in } 1, \dots, K. \end{cases} \quad (15)$$

We decide to focus on this simple case because the crucial issue in VAR identification (see, e.g. Stock and Watson, 2001) is the solution of the contemporaneous causality problem, which allows to estimate the structural shocks. This holds both in the linear and nonlinear VAR case. Since the method we propose exploits contemporaneous nonlinearity, equation (15) represents the most simple time series model in which we can apply it.

2.3 Algorithm for the contemporaneous structure

Previous sections showed that, under the assumption of additive noise, the (graphical) causal structure of a recursive structural equation model is generically identifiable. There are alternative strategies that exploit this identifying principle using sample data (see Peters et al., 2014 and Peters et al., 2017 for an overview). Here we are presenting the RESIT (REGression with Subsequent Independence Test) algorithm proposed in Peters et al. (2014), which incorporates the principles of identification in the nonlinear additive noise setting directly and straightforwardly. It also has the advantage of segmenting the causal search in, first, a topological order search phase and, subsequently, a DAG search phase. This modular architecture fits well with the possibility that for shock recovery (and impulse response analysis) a topological order can be sufficient for the scope (in finite samples), as noted above. The original RESIT algorithm of Peters et al., 2014 operates in two phases (see Appendix A). We add a preliminary phase, which we call Phase 0, in

which we estimate a reduced-form GVAR-ANIM (see equation 1) and use its estimated residuals as input for the next phase.

Phase 1 aims at ranking the input variables in a topological order. (Notice, from the previous section, that the fully connected DAG associated with a topological order may contain extra edges with respect to the true DAG.) The topological order is elicited in a backward fashion by determining which variable is the latest (i.e. a sink node) among the whole set of variables to be ordered, removing it from the set and repeating the search over the resulting subset until only one element remains, which is given the first position in the ranking. At each iteration, the search is performed by regressing each variable jointly on all the others and measuring the dependence (more on this below) between the regressors and the obtained residual. The regressand yielding the weakest dependence is regarded as the sink node among the current set of candidates; it is removed from the set of candidates and is assigned the first position among the set of variables that have already been removed. The topological order simply corresponds to the reversed order of elimination. The topological order determines, for each variable, a set of potential parents. This set coincides with all the variables ranked higher.

Phase 2 (which we call *pruning phase*) further narrows down the sets of potential parents by performing variable selection (see Appendix A). In other words, it prunes edges from the fully connected DAG which one can draw from the topological order of Phase 1. It essentially performs a variable selection task. It regresses each variable on its putative parents (according to the topological order derived above) but omitting iteratively one of them. If the estimated residual of the corresponding nested model is still found independent of the regressors, the omitted regressor is removed from the set of parents of the regressand. The search stops when no regressor can be removed without creating dependence.

Any regression method and dependence measure can be embedded in the RESIT algorithm, as long as they are adequate with respect to the data-generating process. Not knowing the functional dependence, i.e. $\varphi_k(\cdot)$ in equation (12), an appropriate strategy would be to use a nonparametric kernel estimator (Fan and Gijbels, 2018). In our implementation, in tune with Hoyer et al. (2008), we opt for the Gaussian process approach (see Williams and Rasmussen, 2006). An alternative, more restricted, nonparametric model is the generalized additive noise model (Hastie and Tibshirani, 2017). At the extreme, the functional form can be “restricted” down to linearity, as long as one assumes non-Gaussianity of the noises. In this case, the ANM collapses to the Linear Non-Gaussian Acyclic Model studied by Shimizu et al. (2006) (see also Shimizu et al., 2011).

As regards the dependence measure, any measure that accounts for higher order statistics (rather than linear dependence only) can in principle work. This feature is essential because regression techniques that minimize quadratic errors including linear components yield estimated noises that are orthogonal to the covariates. Peters et al. (2014) propose to use the p -value of a nonparametric independence test as dependence score, so that the dependence is minimized when the p -value, under the null hypothesis of independence, is maximized. This follows the principle of Hodges-Lehmann estimation, which was proposed by Herwartz (2018) for independent component analysis. The specific independence test that we perform in our implementation, following

(Peters et al., 2014), is the kernel independence test proposed by Gretton et al. (2007) with the Hilbert-Schmidt Independence Criterion (HSIC). The asymptotic results presented by these authors guarantee a convergence of the p -value to zero only under dependence.

We mentioned above that the RESIT algorithm can be stopped at the topological order step (i.e. Phase 1) if the aim is to estimate structural shocks and impulse response functions. The second phase should instead be applied if one is interested in uncovering the exact causal structure among the innovation terms. Since both in our simulation study and in the empirical application we focus on structural impulse response analysis, the algorithm shown in the main body of the text skips the edge pruning phase (see Algorithm 1 here below). In Appendix A, one can find the original RESIT algorithm embedded in a more general procedure that deals with time series data. Notice that the motivation to skip Phase 2 of RESIT is only of practical nature, when one deals with finite sample. Contrary to Phase 1, Phase 2 operates on the basis of statistical hypothesis testing (independence test), in which one has to decide a rejection threshold. In finite sample, this decision may lead to mistakenly classify weak dependence as independence and therefore to mistakenly cutting edges between directly causally related variables. For the sake of retrieving structural shocks from reduced-form VAR residuals, one has to run a battery of regressions of the form expressed in equation (13), in which shocks will be the residuals of these regressions. If we omit any variable from the parents' set (i.e. covariates of these regressions) because we have made a mistake in Phase 2, we introduce an omitted variable bias. But if we keep an irrelevant variable in the parents' set, we may add some variance in the estimates but we do not introduce any bias. Given this trade-off, we opt for not using Phase 2 for the sake of estimating impulse response functions.

A simple illustration

Let us suppose that a 3-variable GVAR-ANIM process of the type of equations (1-3) has generated some data. Specifically, suppose that equation (3) corresponds to:

$$\begin{cases} u_{t,1} = \varepsilon_{t,1} \\ u_{t,2} = \varphi_2(u_{t,1}, u_{t,3}) + \varepsilon_{t,2} \\ u_{t,3} = \varepsilon_{t,3} \end{cases} \quad (16)$$

The DAG associated to these equations is $u_{t,1} \longrightarrow u_{t,2} \longleftarrow u_{t,3}$. Now, consider to have estimated $u_{t,1}, u_{t,2}, u_{t,3}$ as reduced-form residuals of a VAR of the form of equation (1), following Phase 0 of Algorithm 1. A mistake-free output of Phase 1 will be either the topological order $\langle u_{t,1}, u_{t,3}, u_{t,2} \rangle$ or $\langle u_{t,3}, u_{t,1}, u_{t,2} \rangle$. A mistake-free output of Phase 2 (see Appendix A) will be $u_{t,1} \longrightarrow u_{t,2} \longleftarrow u_{t,3}$. In order to estimate the structural shocks $\varepsilon_1, \varepsilon_2, \varepsilon_3$ one can run a set of regressions on the basis of equations (16). But suppose one stops at Phase 1. Then one would estimate one of the following regression models instead:

Algorithm 1 VAR + RESIT (Phase 1 is from Peters et al., 2014)

- 1: **Input:** A K -dimensional time series vector $(y_{t,1}, \dots, y_{t,K})'$
 - 2: PHASE 0: Estimate the reduced-form model.
 - 3: Estimate a reduced-form time series model of the class GVAR-ANIM, see eq. (1) and extract residuals $\hat{u}_t = (\hat{u}_{t,1}, \dots, \hat{u}_{t,K})'$. To simplify the notation, let us call $\hat{u}_{t,k} \equiv v_k$, for $k = 1, \dots, K$.
 - 4: $S := 1, \dots, K$, $\pi := [\]$
-

- 5: PHASE 1: Determine topological order and potential parent set
- 6: **repeat**
- 7: **for** $k \in S$ **do**
- 8: Regress v_k on $\{v_i\}_{i \in S \setminus \{k\}}$ and obtain residuals e_k
- 9: Measure dependence between e_k and $\{v_i\}_{i \in S \setminus \{k\}}$
- 10: **end for**
- 11: Let k^* be the k with the weakest dependence
- 12: $S := S \setminus k^*$
- 13: $pa(k^*) := S$
- 14: $\pi := [k^*, \pi]$
- 15: **until** $\#S = 0$
- 16: **Output:** π
- 17: **Output:** $(pa(1), \dots, pa(K))$

Note: with $pa(k)$ we refer to the set of indices associated to the variables in $Pa(v_k)$, for any k .

$$\begin{cases} u_{t,1} = \varepsilon_{t,1} \\ u_{t,2} = \varphi_2(u_{t,1}, u_{t,3}) + \varepsilon_{t,2} \\ u_{t,3} = \varphi_3(u_{t,1}) + \varepsilon_{t,3} \end{cases} \quad (17)
 \qquad
 \begin{cases} u_{t,1} = \varphi_1(u_{t,3}) + \varepsilon_{t,1} \\ u_{t,2} = \varphi_2(u_{t,1}, u_{t,3}) + \varepsilon_{t,2} \\ u_{t,3} = \varepsilon_{t,3} \end{cases} \quad (18)$$

Both systems of equations (17) and (18) differ from (16) in including irrelevant variables in the regression systems, but not involving any reverse causality issue.

2.4 Algorithm for the structural impulse response functions

The advantage of identifying a structural model is that one can estimate the dynamic effects of structural shocks. But, under the generic GVAR-ANIM framework, this cannot be done by using the conventional tools of impulse response analysis. Indeed in such a framework the variables' responses to a specific shock at a certain time depend on the system's past history, sign and magnitude of all contemporaneous shocks and subsequent ones. Following Koop et al. (1996) and Kilian and Lütkepohl (2017, ch.18), we define the *structural* (nonlinear) impulse response function (IRF) as:

$$\text{IRF}(h, \delta, \Omega_{t-1}) = E(y_{t+h} | \varepsilon_{t,i} = \delta, \Omega_{t-1}) - E(y_{t+h} | \Omega_{t-1}), \quad (19)$$

where δ is the (positive or negative) magnitude of the shock $\varepsilon_{t,i}$, whose effects 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 function is studied. To estimate equation (19), we can use the Monte Carlo integration approach suggested by Kilian and Lütkepohl (2017).

The idea of this approach, whose detailed scheme is presented in Algorithm 2, is to take pairs of simulation runs: $\tilde{Y}^{\delta,n}$ and \tilde{Y}^n (for $n = 1, \dots, N$), in which $\tilde{Y}^{\delta,n}$ denotes the “treated” time series and \tilde{Y}^n the “control” one, where N is the number of Monte Carlo runs. Both elements share a common history Ω_{t-1} , but diverge at date t , in which a structural shock of interest is set to δ in one run (treatment), while no such restriction applies to the other (control). Apart from this restriction, all shocks are free to fluctuate in both groups of series from time t onwards. The difference is then taken for each pair. Finally, estimated impulse response functions are calculated by taking average of these simulated differences.

Algorithm 2 requires as input estimates (from the observed data) of the functional relationship $\hat{\varphi}_k(\cdot)$ between innovations in $u_{t,k} = \varphi_k(Pa(u_{t,k})) + \varepsilon_{t,k}$ (for k in $1, \dots, K$); estimates of the lagged functional relationship $\hat{f}_k(\cdot)$ in $y_{k,t} = f_k(y_{t-1}, \dots, y_{t-p}) + u_{k,t}$ (for k in $1, \dots, K$);⁵ estimates of the structural shocks $\varepsilon_{t,1}, \dots, \varepsilon_{t,K}$. These can be estimated through the following procedure. For each $k = 1, \dots, K$, if $Pa(u_{t,k}) = \emptyset$ then set $\hat{\varepsilon}_{t,k} := \hat{u}_{t,k}$; else (non-parametrically) regress $\hat{u}_{t,k}$ on $\{\hat{u}_{t,i}\}_{i \in pa(k)}$ and set $\hat{\varepsilon}_{t,k} := \hat{u}_{t,k} - \hat{\varphi}_k(\{\hat{u}_{t,i}\}_{i \in pa(k)})$.⁶

Each simulated⁷ path ($\tilde{Y}^{\delta,n}$ and \tilde{Y}^n) is computed iteratively at each time period, starting from the time point of the “treatment”. At each step, new values are generated summing (i) lagged effects $f_k(y_{t-1}, \dots, y_{t-L})$; with (ii) innovation terms $u_{t,k} = \varphi_k(Pa(u_{t,k})) + \varepsilon_{t,k}$, for every k . New values of $\varepsilon_{t,k}$ are independently drawn from their marginal empirical distribution at each time horizon (except for the one set at δ at time t in the treatment run). Since the innovation model is recursive, each $\tilde{u}_{t,k}$ is computed iteratively at each time step by adding its structural shock $\varepsilon_{t,k}$ to the contemporaneous effect of its parents ($Pa(u_{t,k})$). Parents’ sets can be assigned in two alternative ways: (i) by simply taking the parents’ sets as given by the full version of RESIT as it appears in Appendix A; or (ii) in a more conservative fashion (see discussion above), by considering the sole topological order (see Algorithm 1 in section 2.3) and by regarding any variable $u_{t,k}$ as a parent of those it precedes in that ordering. Either way, the functional forms $\hat{\varphi}_k(\cdot)$ of the contemporaneous effects are estimated via (nonparametric) regression of each variable over the parent set considered.

3 Simulation analysis

The simulation study documented in this section aims at evaluating the performance of our approach in estimating structural impulse response functions. More specifically, we simulate a sim-

⁵For simplicity, we dropped the time index from the lagged effects $f_{k,t}(\cdot)$. A time-varying function can be incorporated into the algorithm we are discussing in this subsection by updating the $f_{k,t}(\cdot)$ at each step.

⁶With $pa(k)$ we refer to the set of indices associated to the variables in $Pa(u_{t,k})$, for any k .

⁷Simulated variables are marked with a tilde sign (\sim).

Algorithm 2 Structural IRFs at time t^*

1: Setting Values:

- t^* : date of the shock to be studied
- H : max horizon
- k^* : index of the shocked variable
- δ : shock magnitude
- N : number of Monte Carlo runs

2: Input data

- history (Ω_{t^*-1}) at date t^*
- estimated structural shock matrix \widehat{E}
- estimated $\widehat{f}_k(\cdot)$, $\widehat{\varphi}_k(\cdot)$ (for each k)
- topological order: $\pi = (\pi_1, \dots, \pi_K)$; parents set: $(pa(1), \dots, pa(K))$

3: PHASE 1: Simulate N couples of time paths and get the difference
4: for $n \in \{1, \dots, N\}$ do

5: Create a $(t^* + H) \times K$ empty matrix \widetilde{Y}^n . Set 1 : $(t^* - 1)$ rows := Ω_{t^*-1}

6: $\widetilde{Y}^{\delta, n} := \widetilde{Y}^n$

7: Create a $(H + 1) \times K$ matrix \widetilde{E} , randomly sampling from \widehat{E} (columnwise)

Note: denote any (t, k) entry of any matrix E with $E_{t,k}$

8: Create a $(H + 1) \times K$ matrix \widetilde{E}^δ , as in line 7, but set $\widetilde{E}_{t^*, k^*}^\delta := \delta$

9: Create a $(H + 1) \times K$ matrices \widetilde{U} and \widetilde{U}^δ (initially empty)

10: Create a $(H + 1) \times K$ matrices $\widetilde{I}^{n, k^*, \delta}$ (initially empty)

11: for $h \in \{0, \dots, H\}$ do
12: for $k \in \{\pi_1 \dots \pi_K\}$ do

13: $\widetilde{U}_{h,k}^\delta := \widehat{\varphi}_k(\{\widetilde{U}_{h,i}^\delta\}_{i \in pa(k)}) + \widetilde{E}_{h,k}^\delta$ (with $\varphi_k(\emptyset) = 0$)

14: $\widetilde{Y}_{t^*+h,k}^{\delta, n} := \widehat{f}_k(\widetilde{Y}_{t^*+h-1}^{\delta, n}, \dots, \widetilde{Y}_{t^*+h-L}^{\delta, n}) + \widetilde{U}_{h,k}^\delta$

15: $\widetilde{U}_{h,k} := \widehat{\varphi}_k(\{\widetilde{U}_{h,i}\}_{i \in pa(k)}) + \widetilde{E}_{h,k}$

16: $\widetilde{Y}_{t^*+h,k}^n := \widehat{f}_k(\widetilde{Y}_{t^*+h-1}^n, \dots, \widetilde{Y}_{t^*+h-L}^n) + \widetilde{U}_{h,k}$

17: $\widetilde{I}_{h,k}^{n, k^*, \delta} := \widetilde{Y}_{t^*+h,k}^{\delta, n} - \widetilde{Y}_{t^*+h,k}^n$

18: end for
19: end for
20: end for
21: PHASE 2: Average values

22: $\widetilde{I}^{k^*, \delta, t^*} := \frac{1}{N} \sum_{n=1}^N \widetilde{I}^{n, k^*, \delta, t^*}$

23: **Output:** $\widetilde{I}^{k^*, \delta, t^*}$

ple VAR with a nonlinear contemporaneous structure (under different specifications), we then estimate the structural IRFs using Algorithm 2, but under different identification approaches, and, we systematically compare them. Our simulation results, as reported below, demonstrate that, in case of nonlinearities in the data generating process, our procedure is able to satisfactorily estimate the structural IRFs and that a linear approximation may lead astray even under correctness of the recursive order.

3.1 Data generating processes

We simulate a 3-variables VAR model with one lag, which can be written as

$$y_t = A_1 y_{t-1} + u_t, \quad (20)$$

where A_1 , to control for the persistence of the process, is a lower triangular matrix with all elements below and on the main diagonal equal to a . We vary the persistence of the process by selecting a among 0.1, 0.5, 0.9. The u_t terms are generated following three types of causal structures, which we call the “common cause”, “causal chain” and “v-structure”, which incorporate nonlinear functional relationships. The common-cause structure reads as:

$$\begin{cases} u_{t,1} = \varepsilon_{t,1} \\ u_{t,2} = \text{sign}(u_{t,1})|u_{t,1}|^\alpha + \varepsilon_{t,2} \\ u_{t,3} = \sin(\text{sign}(u_{t,1})|u_{t,1}|^\beta) + \varepsilon_{t,3}, \end{cases} \quad (21)$$

whereas the causal-chain structure corresponds to

$$\begin{cases} u_{t,1} = \varepsilon_{t,1} \\ u_{t,2} = \text{sign}(u_{t,1})|u_{t,1}|^\alpha + \varepsilon_{t,2} \\ u_{t,3} = \sin(\text{sign}(u_{t,2})|u_{t,2}|^\beta) + \varepsilon_{t,3}, \end{cases} \quad (22)$$

and the v-structure is:

$$\begin{cases} u_{t,1} = \varepsilon_{t,1} \\ u_{t,2} = \varepsilon_{t,2} \\ u_{t,3} = \text{sign}(u_{t,1})|u_{t,1}|^\alpha + \sin(\text{sign}(u_{t,2})|u_{t,2}|^\beta) + \varepsilon_{t,3}. \end{cases} \quad (23)$$

The parameters α and β are independently drawn from a uniform distribution with support $[1, 4]$, the shocks $\varepsilon_{t,1}, \varepsilon_{t,2}, \varepsilon_{t,3}$ are i.i.d. and normally distributed with variance-covariance matrix equal to the identity matrix. The nonlinear functions entering in equations (21-23) were selected in accordance with the analysis of Hoyer et al. (2008) and Peters et al. (2014). We generate samples $\{y_t\}_{t=1}^T$ of size $T = 250, 500, 1000$.

3.2 Identification schemes

For each selected T , a , type of structure, we generate 200 artificial (3-variate) time series, from which we estimate IRFs. In each simulation run, we randomly draw values for α , β and shocks. These are constructed following Algorithm 2 (sticking to the definition of IRF given in equation 19), according to five different schemes, which we call: (i) “theoretical” IRFs, (ii) Cholesky-based (linear) IRFs, (iii) true-graph-based nonlinear IRFs, (iv) true-topological-order-based nonlinear IRFs, and (v) Algorithm-1-based nonlinear IRFs.

Scheme (i) builds IRFs by applying Algorithm 2 to each artificial time series, except that instead of the estimated structural shock matrix (see “Input data” in the pseudo code) we use the true DGP’s one, instead of the estimated functional forms we use the autoregressive matrix A_1 , and instead of the true nonlinear function and $\varphi_k(\cdot)$, and we use the correct set of parents (and topological order) implied by the true causal structure. We calculate IRFs using this scheme because we want to have a ground truth against which we can assess our procedure. In practice, scheme (i) is the application of Algorithm 2 to the true model, bypassing the model estimation step. For this reason, we call the produced IRFs “theoretical”.

In contrast, the remaining four procedures are based on estimates of the model. Scheme (ii) is meant to provide a linear benchmark to the nonlinear approaches. Here structural shocks are computed via Cholesky factorization of the reduced-form residuals’ covariance matrix, where the recursive ordering of the innovation structure is the true one, instead of being determined in a data driven fashion as Algorithm 1 does. Moreover, for the sake of comparison with the other schemes, we do not apply here the standard linear approach that derives IRFs from MA coefficients, but we feed Algorithm 2 with the estimated linear coefficients. We label IRFs from this scheme “CHOL” (from Cholesky).

Scheme (iii) is similar to scheme (i) except that the functional (autoregressive and contemporaneous) dependencies are estimated from the data, on the basis of the true causal graph (which automatically implies a correct topological order). Functional forms are estimated via OLS as regards the autoregressive components and through Gaussian process regression as regards the contemporaneous structure. In practice, in this scheme we are bypassing the data-driven causal (order) search by using the correct specification of the model. The aim here is to isolate the performance of Algorithm 2, avoiding potentially compounding mistakes of Algorithm 1. We label IRFs from this scheme “TDAG” (i.e. true DAG).

Scheme (iv) is a slight variant of scheme (iii): we feed Algorithm 2 with a correct topological order and a fully connected DAG deduced from it, which may have extra edges with respect to the true DAG. Thus, we also isolate here the performance of Algorithm 2 (as in (iii)), but we allow the introduction of potentially irrelevant variables in its inputs, to see if they make a difference in terms of performance. Notice that Algorithm 1, which is involved in the next scheme, may also introduce these extra variables, so that scheme (iv) is a closer benchmark to scheme (v). Furthermore, scheme (iv) is a nonlinear counterpart of the Cholesky scheme (ii) because they are both based on the true topological order, without any further restrictions (i.e., they may include

irrelevant covariates). We label IRFs from this scheme “TTOP” (i.e. true topological order).

Scheme (v) implements the data-driven approach proposed in Section 2 so that its assessment is the focus of this analysis, the previous schemes ultimately being benchmarks for this one. Here, from each generated time-series sample we estimate the topological order, from which we deduce a fully connected DAG and a corresponding set of parents, by Algorithm 1 and subsequently we apply Algorithm 2. As input for the functional forms we use OLS for the autoregressive part (as in schemes (ii-iii-iv)) and Gaussian process regression for the contemporaneous structure (as in schemes (iii-iv)). Recall from Section 2.4 that the structural shocks (further input of Algorithm 2) are recovered from the residuals of the regression model at the contemporaneous level. We label IRFs from this scheme “ALG1” (i.e. based on Algorithm 1).

3.3 Simulation results

Let us call $\text{IRF}_{k,h}^\bullet$ the IRF referring to the responses of variable $y_{t,k}$ to shock $\varepsilon_{t,1}$, at horizon h , using scheme \bullet , where $k \in \{2, 3\}$, $h \in \{0, 1, 2, 3, 4\}$, $\bullet \in \{\text{CHOL}, \text{TDAG}, \text{TTOP}, \text{ALG1}\}$. For a given data generating process, we calculate the following average mean squared error:

$$\widehat{\text{AMSE}}^\bullet = \frac{1}{10} \frac{1}{200} \sum_{k=2,3} \sum_{h=0}^4 \sum_{l=1}^{200} (\widehat{\text{IRF}}_{k,h}^\bullet - \widehat{\text{IRF}}_{k,h}^{\text{theor.}})^2, \quad (24)$$

where l indicates a simulation run. We focus on the response of the first shock to other variables since the first variable is always exogenous in all the causal structures we consider.

Table 1 shows estimated AMSE, averaged across 200 simulation experiments, for different values of T and a , along with standard errors, for the causal-chain structure. Analogous results for the common-cause (Table 2) and v-structure (Table 3) are reported in Appendix B since they are qualitatively similar.

Looking at the top panels (of all these tables), we note that the ALG1 scheme (v) outperforms the Cholesky one (scheme ii) whatever the structure, sample size and persistence of the model in time are. Recall that scheme (v) has the advantage, over scheme (ii), of allowing for nonlinearity in the estimation of impulse response functions (since it uses Gaussian process regressions via Algorithm 2), when the data are generated by a nonlinear DGP, as in this case. But it undertakes the risky task of inferring the causal order from data.⁸ On the other hand, scheme (ii) runs in this case under a misspecified model (the linear one). But it has the advantage of overriding the search for a topological order from the data: it is given a correct one.

⁸Peters et al. (2014) report simulation results about the accuracy of the RESIT algorithm, showing that the performance (in a nonlinear setting) in terms of distance between the estimated structure and the correct DAG is good both in outperforming other established causal discovery algorithms, but also in absolute terms. We have replicated this analysis by applying RESIT to VAR residuals and we found similar results. We also studied the accuracy of Algorithm 1 in returning a correct topological order. Both these results are available upon request. The accuracy rate is satisfactory in most cases, but we also found that small samples (i.e. $T < 500$) can be problematic for some causal structure (specifically for the common cause). Note that, however, this problem seems to have only a limited impact on the estimates of the structural IRFs (which is the main object of this study), as shown in the subsequent analysis reported in Figures 4 and 5.

Table 1: Average MSE, across 200 simulations, between IRFs estimated using different schemes (i.e. CHOL, TDAG, TTOP, ALG1) and the theoretical IRFs. MSE are calculated over the responses of $y_{t,2}$ and $y_{t,3}$ to the first shock (of one standard-deviation magnitude) and over the first five periods. The contemporaneous structure is the “causal chain”. We consider different values of T (sample size) and a (persistence parameter). Standard errors are reported in round brackets.

| | | CHOL | | | ALG1 | | |
|-----|---|----------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | T | 250 | 500 | 1000 | 250 | 500 | 1000 |
| a | | | | | | | |
| 0.1 | | 0.69 (9.60E-02) | 0.518 (6.60E-02) | 0.654 (8.30E-02) | 0.062 (9.00E-03) | 0.049 (8.00E-03) | 0.035 (4.00E-03) |
| 0.5 | | 1.428 (1.05E-01) | 1.023 (7.10E-02) | 1.303 (8.90E-02) | 0.188 (1.20E-02) | 0.124 (1.00E-02) | 0.085 (5.00E-03) |
| 0.9 | | 12.636 (8.93E-01) | 9.253 (5.82E-01) | 11.76 (7.27E-01) | 1.506 (1.09E-01) | 1.274 (1.08E-01) | 0.916 (5.80E-02) |
| | | TDAG | | | TTOP | | |
| | T | 250 | 500 | 1000 | 250 | 500 | 1000 |
| a | | | | | | | |
| 0.1 | | 0.046 (6.00E-03) | 0.032 (7.00E-03) | 0.038 (8.00E-03) | 0.046 (5.00E-03) | 0.033 (6.00E-03) | 0.034 (7.00E-03) |
| 0.5 | | 0.142 (8.00E-03) | 0.086 (9.00E-03) | 0.091 (8.00E-03) | 0.135 (7.00E-03) | 0.086 (7.00E-03) | 0.081 (8.00E-03) |
| 0.9 | | 1.256 (9.20E-02) | 0.839 (1.06E-01) | 0.99 (8.80E-02) | 1.095 (7.40E-02) | 0.776 (6.60E-02) | 0.848 (7.70E-02) |

The TTOP scheme (iv) fills the gap between RESIT and Cholesky in being a nonlinear approach — like the former — and hacking the order search — like the latter. As expected, TTOP systematically beats RESIT in terms of AMSE, but respective scores notably remain within the same order of magnitude (whereas AMSEs for Cholesky are greater by one order of magnitude). Comparisons between TDAG and TTOP do not clearly discriminate between the two. It seems that within our sample range the less parsimonious modeling strategy (TTOP) often allows to better mimic the DGP than what can be achieved by faithfully sticking to the actual DGP specification (TDAG). This finding, admittedly unexpected, supports our decision to use Algorithm 1 (i.e. a truncated version of the original RESIT by Peters et al. (2014)).

From Table 1, we also see that the AMSE is higher the greater is a , no matter which scheme is used. This should not be a surprise. Given our reduced-form specification (see equation 20), initial prediction errors of the shocks’ diffusion at horizon 0 are bound to have stronger echoes at later horizons with more persistent models (i.e. the higher the a). This has an impact on our performance measure, which averages out the first five horizons. As regards of how AMSE

changes with respect to the sample size T , we notice that with CHOL AMSE does not show any downward trend over T , while with ALG1 AMSE shows a clear tendency to decrease, at least over the values of T we consider.

Figures 4 and 5 show structural IRFs in the causal chain setting computed via Monte Carlo integration (Algorithm 2) under the theoretical scheme (i), the CHOL scheme (ii), and the ALG1 scheme (v), plotted from horizon 0 to 10. Dashed lines give the 68% confidence intervals of those averages. Specifically, Figure 4 shows the dynamic responses of variable 2 ($y_{t,2}$) to setting structural shock $\varepsilon_{t,1}$ to one at time $t^* = 2$, while Figure 5 displays its counterpart for variable 3 ($y_{t,3}$). For the sake of conciseness and readability, we do not plot IRFs obtained with TDAG and TTOP since they are very similar to each other and tend to overlap with those obtained by ALG1. For the same reasons, we display simulation results only for the smallest samples ($T = 250$, left-hand panels) and largest ($T = 1000$, right-hand panels) while the gradient of persistence parameters are displayed vertically ($a = 0.1$ in top panels, $a = 0.5$ middle and $a = 0.9$ bottom). Analogous figures for the common-cause (Figures 11 and 12) and v-structure (figures 13 and 14) settings can be found in Appendix B. From these plots, one can notice that the ALG1-IRFs are in most cases, even at horizon equal to zero, indistinguishable from the theoretical ones. Moreover, the failure of the CHOL scheme in recovering contemporaneous shocks' impacts is quite evident. Since CHOL is built on the correct (imposed a priori) topological order, this suggests that knowing this order is not enough to accurately estimate structural IRFs, but it is also crucial to allow for nonlinearities in the regression method, if present. The nonparametric regression we use (GP regression) seems to successfully capture the nonlinear contemporaneous effects.

4 Empirical application

Our empirical application studies the effects of macroeconomic shocks on nominal interest rate, inflation and output, using U.S. data. The study allows for nonlinearity at the contemporaneous level. We compare our findings with the results from a linear recursive SVAR model, highlighting the importance of taking into account nonlinearity. This fact allows us to depart from a typical feature of linear SVAR models, namely the fact that the effect of a positive shock (e.g., a contractionary monetary policy shock) is by construction symmetric to the effect of a negative shock (e.g., an expansionary monetary policy shock). Asymmetric effects of shocks have been studied, among others, by Lo and Piger (2005), Höppner et al. (2008), Kilian (2014), Hussain and Malik (2016).

We estimate a linear VAR model with three variables: inflation (π_t), a measure of macroeconomic activity (o_t), and a nominal short-term interest rate (r_t). We take the data from the Federal Reserve Economic Database (FRED). The data are collected at the quarterly frequency in U.S. and cover the period from 1953-Q3 to 2019-Q4 (270 observations). Inflation is computed as the change in the logarithm of the GDP deflator (FRED mnemonic GDPDEF). As a measure of economic activity, we consider the output gap, computed as the deviation of the logarithm of real GDP (FRED mnemonic GDPC1) from the logarithm of potential GDP (FRED mnemonic GDP-

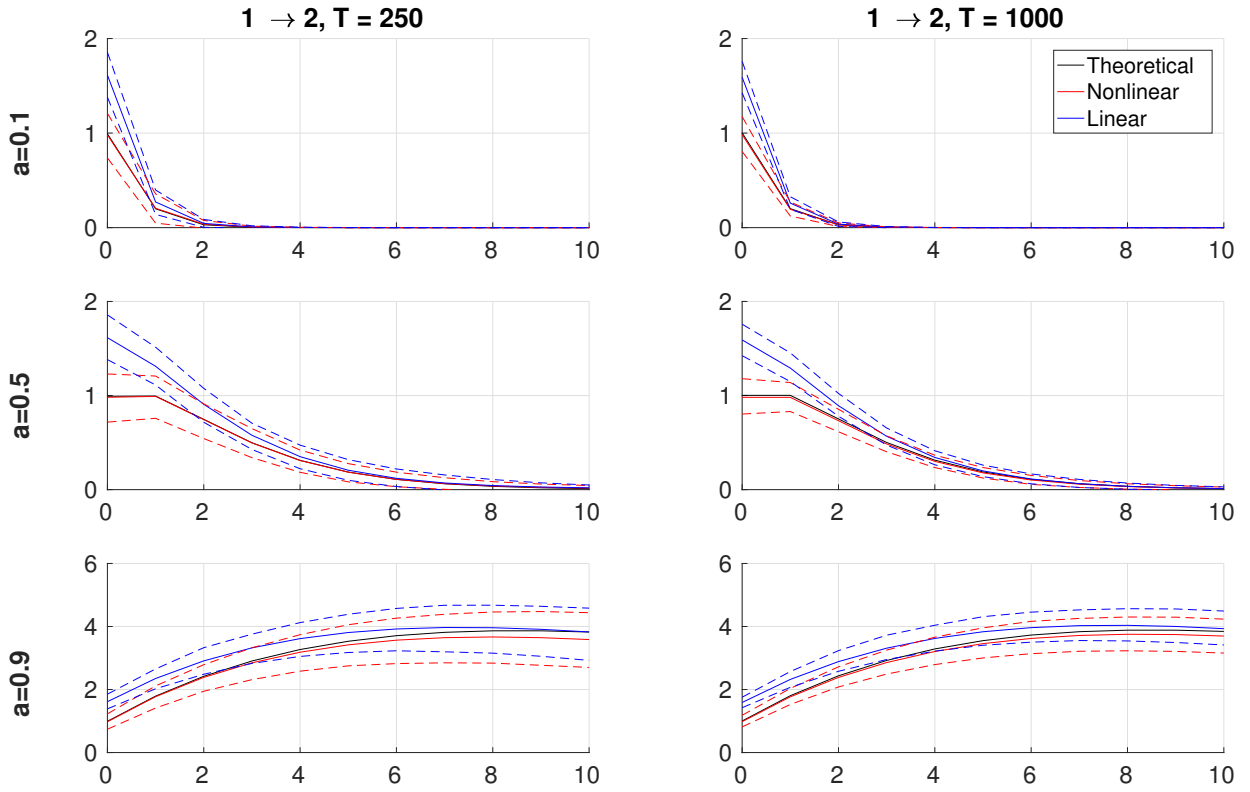


Figure 4: Theoretical, linear (CHOL), and nonlinear (ALG1) IRFs for different values of a and T . Red (blue) lines exhibit the average, across 200 simulations, for the nonlinear (linear) IRFs of variable 2 ($y_{t,2}$) to a unitary shocks $\varepsilon_{t,1}$. DGP with “causal chain”. Confidence interval at 68% are reported in dotted lines.

POT), which we take from the U.S. Congressional Budget Office’s estimate. Finally, the nominal short-term interest rate is set to the Federal Funds Effective Rate (FRED mnemonic FEDFUNDS). The three time series are plotted in Figure 6. A six-lags VAR specification is selected according to the Akaike information criterion.

We identify a SVAR model of the type of equation (15). As mentioned in Section 2, identification can be obtained by knowledge of the causal structure among innovation terms. As noted, for the sake of estimating structural impulse response functions, knowledge of the topological order is sufficient in our framework, given the recursiveness assumption. Therefore, as in the simulation analysis, we apply Algorithm 1 to the data (i.e., first step of RESIT to the reduced-form VAR residuals). Having obtained the topological order and a set of (potential) graphical parents from Algorithm 1, we feed Algorithm 2 with these two pieces of information. Thus, we can estimate the structural IRFs, setting the number of simulations for the Monte Carlo integration $N = 100$. Both in Algorithm 1 and Algorithm 2, estimates of $\varphi_k(\cdot)$ (i.e., functional dependence between innovation terms) are obtained through Gaussian process regressions.

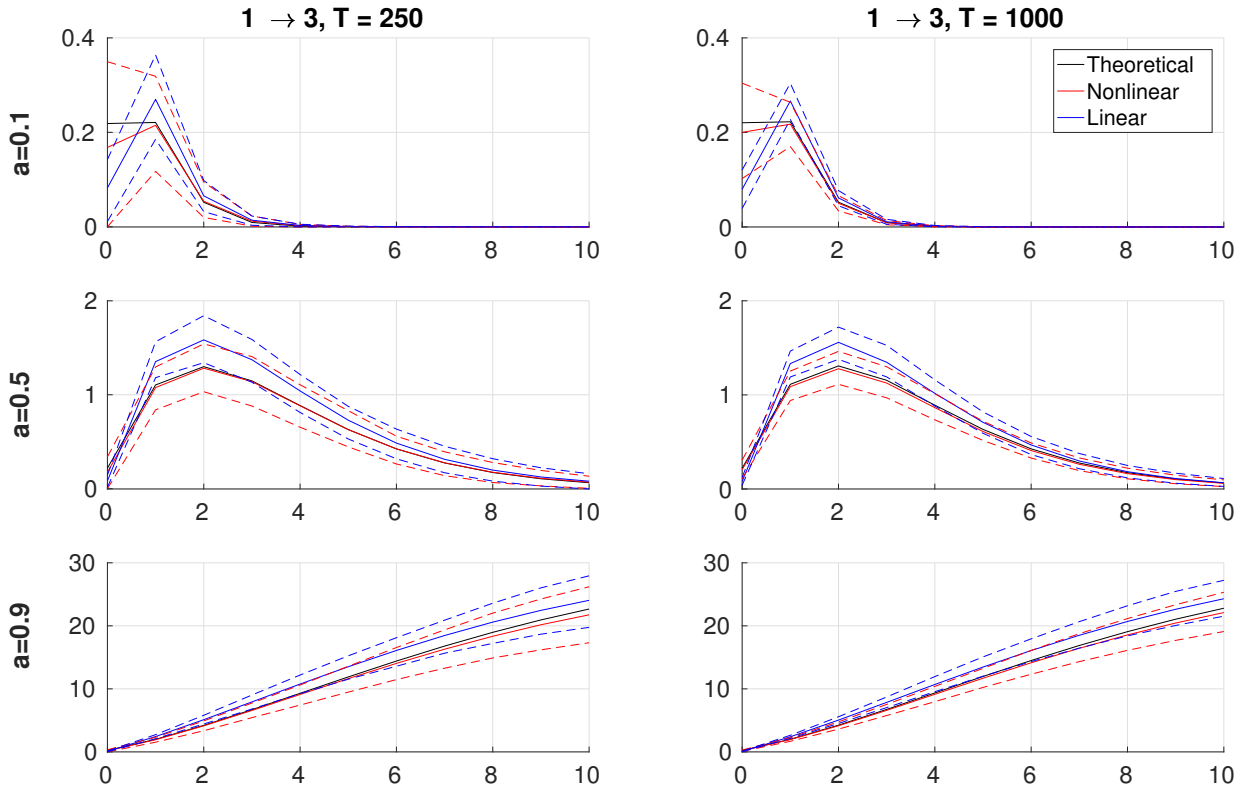


Figure 5: Theoretical, linear (CHOL), and nonlinear (ALG1) IRFs for different values of a and T . Red (blue) lines exhibit the average, across 200 simulations, for the nonlinear (linear) IRFs of variable 3 ($y_{t,3}$) to a unitary shocks $\varepsilon_{t,1}$. DGP with “causal chain”. Confidence interval at 68% are reported in dotted lines.

To address uncertainty in the estimates of the structural IRFs, we perform a bootstrap analysis with 200 iterations. For each bootstrap iteration, we estimate a new VAR model and, maintaining the topological order that we got from the application of Algorithm 1 to the original data, we compute new contemporaneous effects and structural shocks.⁹ On the basis of this, we calculate structural IRFs following Algorithm 2.

The topological order that we obtain by applying Algorithm 1 to our time series is $\langle r_t, o_t, \pi_t \rangle$, which implies that the Federal Funds Rate is a potential contemporaneous cause of the other two variables, and that the output gap is a potential contemporaneous cause of inflation. This goes against the conventional wisdom, according to which central banks set their policy instruments by continuously monitoring inflation and output, while the effects of their interventions do not have immediate effects on output and inflation. On the other hand, the DSGE model estimated by Smets and Wouters (2007) implies that output and inflation should respond immediately to the monetary policy shock (see Figure 6 of their paper) and that the productivity shock affects immediately

⁹We maintain the same topological order at each bootstrap iteration because we do not want to account for uncertainty about inference of the causal structure here. To do that, one can adapt the stability selection procedure by Meinshausen and Bühlmann (2010).

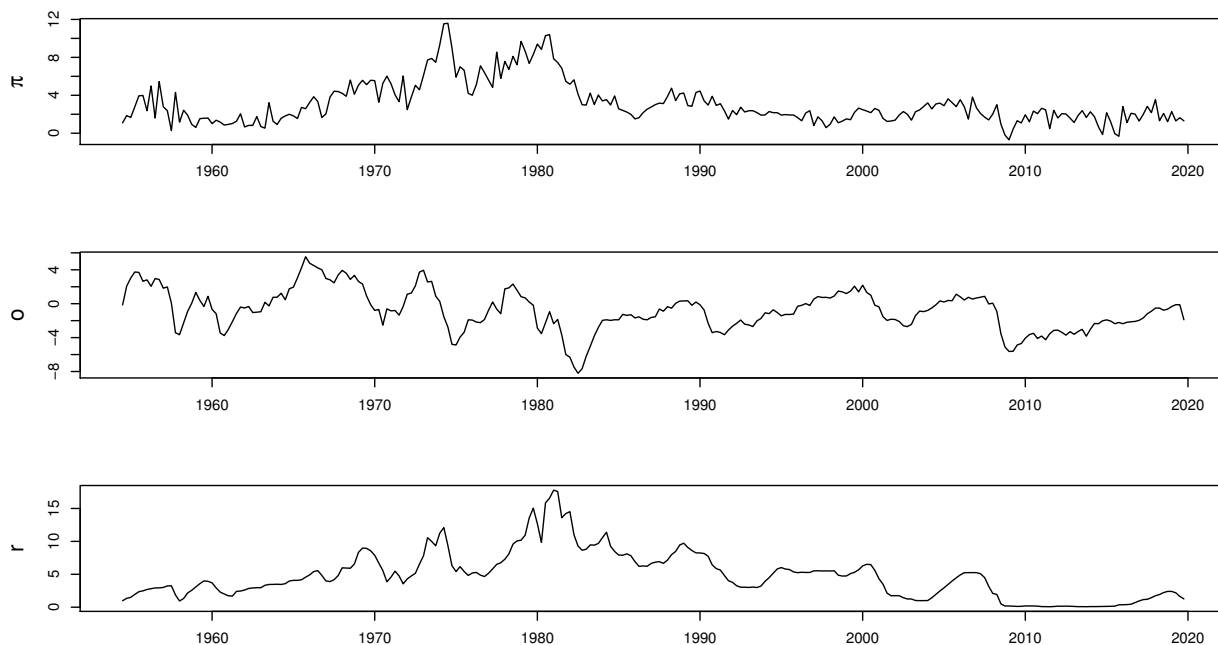


Figure 6: Employed time series. From the top, inflation (π), output gap (o) and nominal interest rate (r). Data sample from 1953-Q3 to 2019-Q4.

inflation (Figure 7 therein).

For each nonlinear IRF we analyze, we consider the response of a variable $y_{(t+h),k}$ to a shock $\varepsilon_{t,j}$ of magnitude equal to one standard deviation of the variable $y_{t,k}$. We do it for $h = 0, \dots, 20$. We compare these nonlinear IRFs with linear IRFs estimated using the following scheme: (i) we impose the Cholesky order derived from the topological order estimated by Algorithm 1 (i.e. $\langle r_t, o_t, \pi_t \rangle$); (ii) hence we get the structural shocks and the coefficients of the impact matrix; (iii) using the latter terms as input, we apply Algorithm 2 and get IRFs.

Figure 7 shows the median effects of a positive interest rate shock (i.e. a contractionary monetary policy shock) on the three variables (i.e. interest rate, output gap, inflation), derived from bootstrapped nonlinear (left column) and linear (right column) IRFs. Confidence intervals refer to the interquartile range (i.e. the 25th and 75th percentiles). Analogously, Figure 8 shows the median effects of a positive output shock (i.e. a productivity shock) on the three variables (i.e. interest rate, output gap, inflation), derived from bootstrapped nonlinear (left column) and linear (right column) IRFs, with confidence intervals. Again, Figure 9 shows the median effects of an inflation shock on the three variables, derived from bootstrapped nonlinear (left column) and linear (right column) IRFs, with confidence intervals.

Looking at the bottom right panel in Figure 7, the linear analysis of IRFs suggests a positive and significant response of inflation to a contractionary monetary policy shock. This finding is referred to in the literature as “price puzzle” (Sims, 1992; Ramey, 2016), because it is at odds with economic theory and background knowledge. The nonlinear analysis of IRFs, however, shows

a negative median contemporaneous effect of the interest rate shock on inflation, even if not for all values of the interquartile range (bottom left panel, same Figure). After some fluctuations, the negative effect on inflation becomes evident in the long-run horizon, leaving substantially no evidence for the price puzzle.

Let us now consider the effect of economic activity in responses to a monetary policy shock. Looking at the second-row panels of Figure 7, we observe, from the nonlinear IRF analysis (left), a negative response of output to a contractionary policy shock while the same effect is (oddly) positive in the first year from the linear IRF analysis (right). Starting from the second year, both in the nonlinear and linear case, the response of output is negative (same panels). As mentioned, the nonlinear analysis gives us the possibility of estimating responses that are not necessarily symmetric under the change in the shock's sign. Thus, let us consider the responses of a negative interest rate shock (i.e. an expansionary monetary policy shock). These are displayed in Figure 10, in which the left hand side panels (showing nonlinear IRFs) are not symmetric to the left hand side panels of Figure 7, while we observe a perfect symmetry between the right hand sides of the two figures (showing linear IRFs).

Next, we examine the responses of the variables to the output shock, as displayed in Figure 8. The responses of both the interest rate and output gap are substantially the same for the linear and nonlinear model, essentially due to the recursive order in which the interest rate precedes output (first and second row, same figure). On the other hand, we observe that there is a difference between the nonlinear and the linear model, as concerns the immediate and short-run response of inflation. The inflation response in the nonlinear model is unambiguously positive since the impact and it remains so in the medium-long run (bottom left panel), while its response in the linear model is around zero at the impact and in the first couple of periods (bottom right panel). Thus, we may attribute the response of interest rate to the output shock (mentioned above), as the delayed reaction of monetary policy to the increase of inflation.

Finally, Figure 9 shows the responses of the three variables to the inflation shock. We notice that there are no substantial differences between the two models, since in the contemporaneous causal order inflation is the sink node. In both the nonlinear and linear case, the responses of interest rate to inflation is increasing over time and positive, the response of output is positive in the first year, and then negative after the second year, and the response of inflation is decisively positive at the impact, and slowing down over time.

In the light of these findings, the nonlinear responses, in comparison with the linear ones, seem to be more convincing, since they solve the price puzzle, allow for asymmetries between responses from contractionary and expansionary monetary shocks, and show an impact of output shock on inflation which seems to mediate the interest rate response.

There are two important remarks that we can draw from this empirical exercise, one on monetary economic policy, one methodological. From a policy perspective, results from the nonlinear-contemporaneous model confirm that tightening monetary policy is successful in reducing inflation, but only after several periods (at least one year) and with major costs. These costs are a

reduction in output, which is immediate and sharp in the first year and takes several periods (some years) to vanish. The hope, sometimes expressed by policy makers, of reducing inflation without triggering recessions is feeble, on the basis of our evidence. Expansionary monetary policy, on the other hand, is successful in increasing output, but its positive effects are visible only after one year, and accompanied by fluctuating effects on inflation (in the short term).

From a methodological point of view, this empirical analysis illustrates the possibility of using a data-driven/statistical approach to identification that yields results which can shed new lights on well-accepted stylized facts. It also shows the advantage of relaxing the nonlinear assumption. In this application, the only difference between the nonlinear and linear model lies in the functional form of the contemporaneous shocks' transmission mechanism and, therefore, all the discrepancies between the two sets of IRFs can be attributed to it. Our findings show that such discrepancies have important consequences for the policy-maker's point of view.

5 Conclusions

There is a departure from the linear model with Gaussian disturbances that can be exploited for identification, namely nonlinearity, as long as the disturbances are additive. We have shown that this feature can be exploited to identify a structural VAR model that contains nonlinearities in the cross-sectional structure. The identification criterion we have exploited in this paper is based on the following simple idea: in case of nonlinearity it is admitted an additive noise model (i.e. a model in which disturbances are independent of covariates) from the cause to the effect, but, in the generic case (see conditions C1-C3 in section 2.2), not from the effect to the cause. Thus, by iterative regressions and subsequent independence tests, it is possible to determine the contemporaneous causal structure underlying a VAR model, under the assumption of recursiveness and mutual independence of the shocks. On the basis of this, we can recover structural impulse response functions, which we define as difference of conditional expectations and estimate through Monte Carlo integration. It also turns out that, for the sake of estimating impulse response functions, the inferred causal structure can contain some redundant links, as long as the topological order is correct. Simulation results have shown that the proposed search procedure and our scheme to estimate impulse response functions perform correctly under data generating processes satisfying the theoretical conditions. Moreover, with the same type of data, they outperform methods based on linearity assumptions, even under the correct zero restrictions. Our empirical analysis has shown that taking into account nonlinearity makes a substantial difference in terms of predicting the effect of macroeconomic shocks to the economy, particularly because it allows for asymmetric effects between an expansionary and a contractionary monetary policy.

The idea of exploiting nonlinearity for identification has been put into practice in this paper by recovering a topological order through Algorithm 1, but one can, of course, conceive different ways to implement it. Other procedures worth exploring are those in the family of the score-based approach to causal discovery (see, e.g., Peters et al., 2014: section 4.2, and Bühlmann et al., 2014). A limitation of our procedure can be seen in the recursiveness assumption. This has been relaxed,

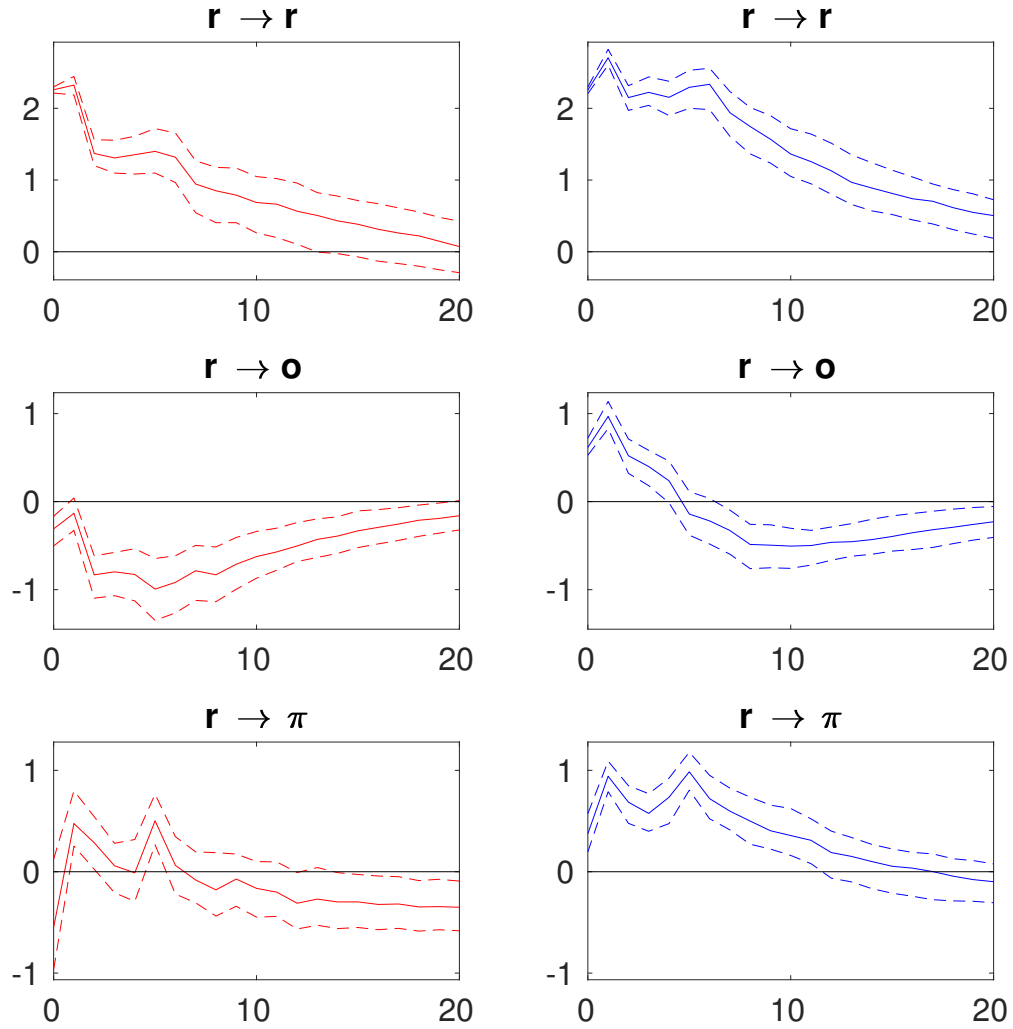


Figure 7: Structural IRFs to an interest rate shock of one-standard deviation magnitude. Left panels show nonlinear structural IRFs obtained from Algorithm 2 with GP regressions. Right panels show linear structural IRFs obtained from Algorithm 2 but fed with coefficients derived from Cholesky factorization. Solid lines refer to the median effects while dotted lines refer to the interquartile range computed over 200 bootstrap simulations.

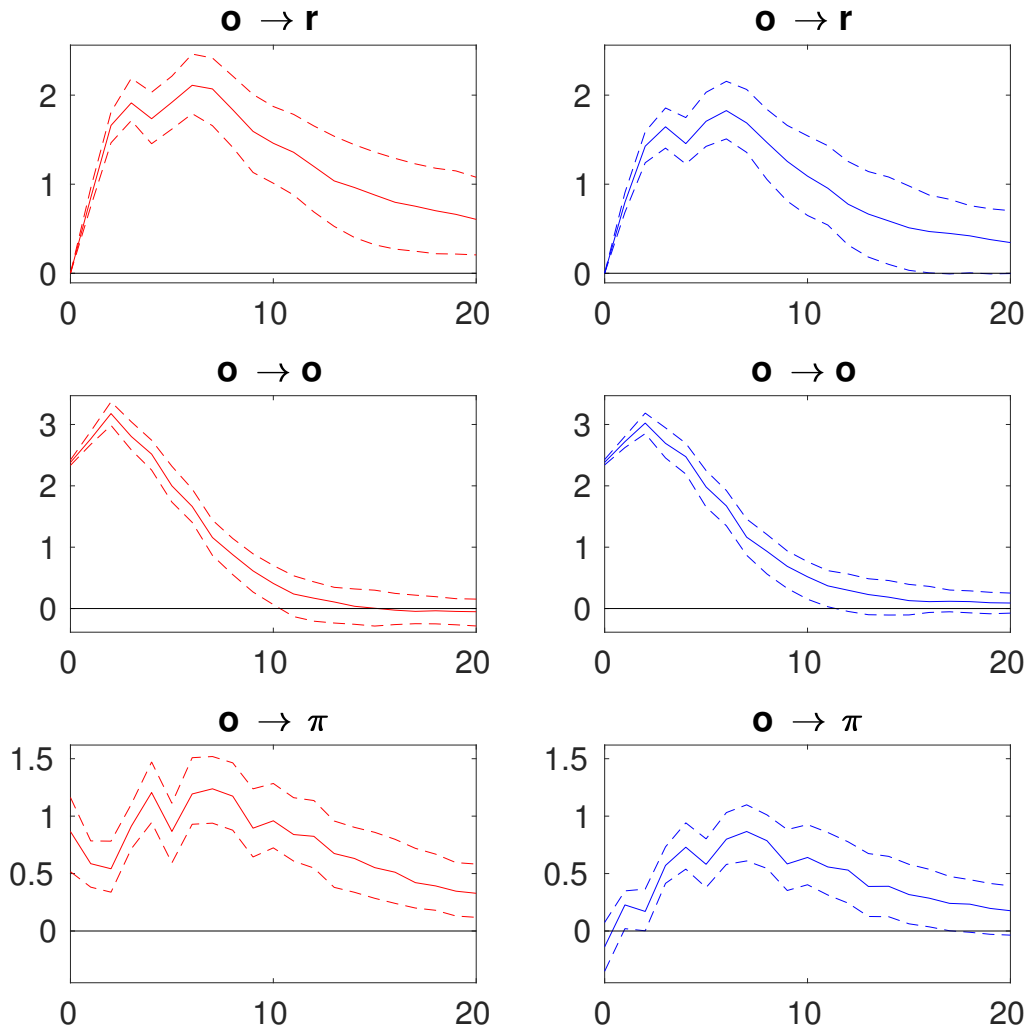


Figure 8: Structural IRFs to an output gap shock of one-standard deviation magnitude. Left panels show nonlinear structural IRFs obtained from Algorithm 2 with GP regressions. Right panels show linear structural IRFs obtained from Algorithm 2 but fed with coefficients derived from Cholesky factorization. Solid lines refer to the median effects while dotted lines refer to the interquartile range computed over 200 bootstrap simulations.

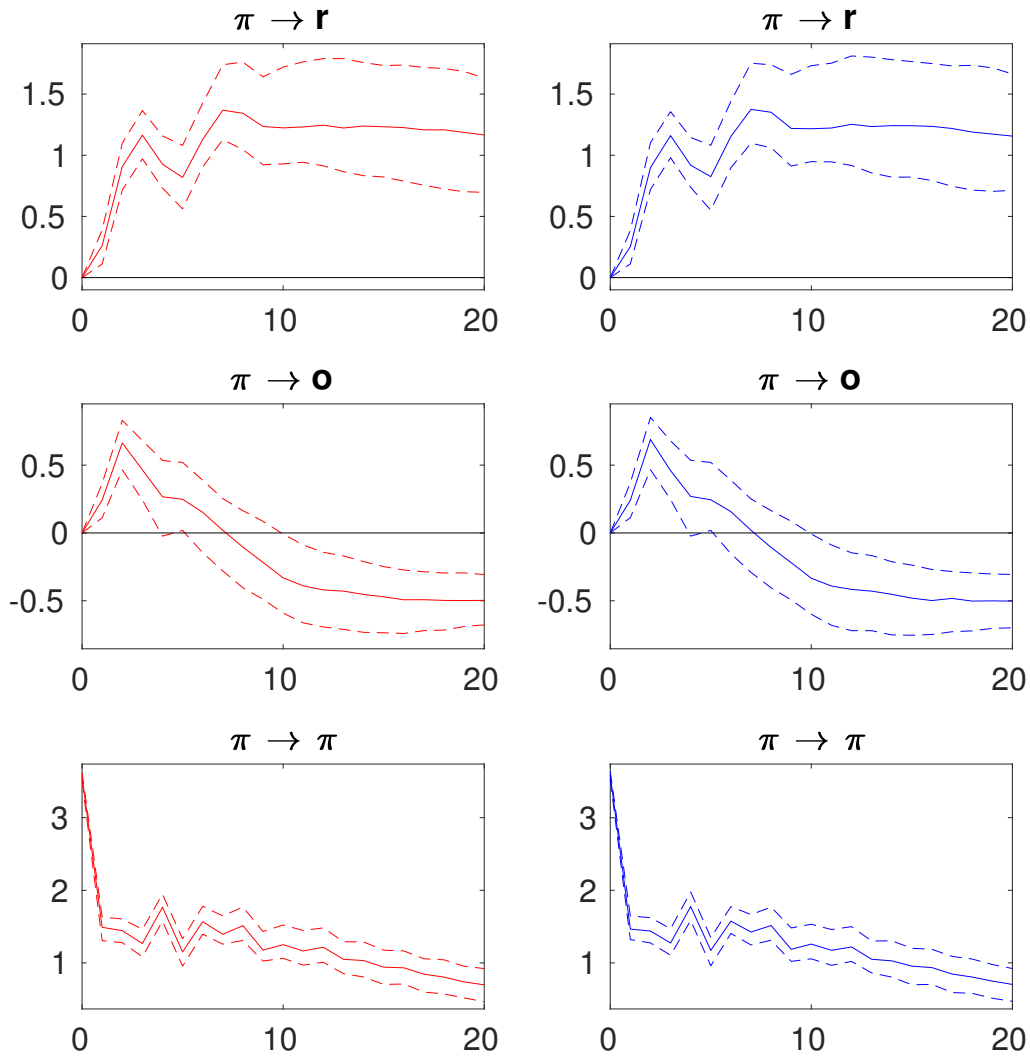


Figure 9: Structural IRFs to an inflation shock of one-standard deviation magnitude. Left panels show nonlinear structural IRFs obtained from Algorithm 2 with GP regressions. Right panels show linear structural IRFs obtained from Algorithm 2 but fed with coefficients derived from Cholesky factorization. Solid lines refer to the median effects while dotted lines refer to the interquartile range computed over 200 bootstrap simulations.

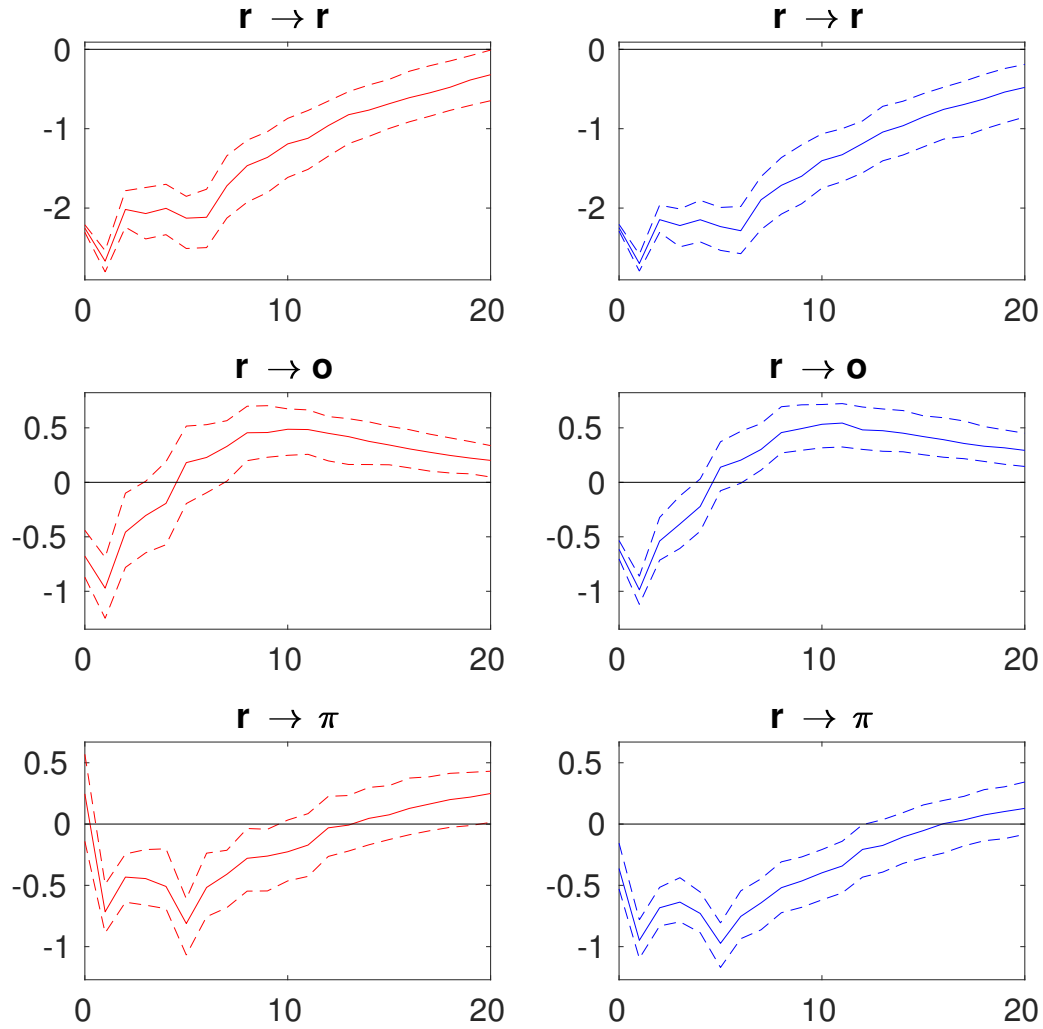


Figure 10: Structural IRFs to a negative interest rate shock of one-standard deviation magnitude. Left panels show nonlinear structural IRFs obtained from Algorithm 2 with GP regressions. Right panels show linear structural IRFs obtained from Algorithm 2 but fed with coefficients derived from Cholesky factorization. Solid lines refer to the median effects while dotted lines refer to the interquartile range computed over 200 bootstrap simulations.

in the bivariate case, by Mooij et al. (2011), but further research is needed for the case with more than two variables and, more in general, for its adaptation to the structural VAR framework. Our hope is that the potentiality that nonlinearity has shown for identification will foster the adoption of this idea within the field of econometrics.

References

- Blanchard, O. J. and D. Quah (1989). The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review* 79(4), 655–673.
- Brunnermeier, M., D. Palia, K. A. Sastry, and C. A. Sims (2021). Feedbacks: financial markets and economic activity. *American Economic Review* 111(6), 1845–79.
- Bühlmann, P., J. Peters, and J. Ernest (2014). CAM: Causal additive models, high-dimensional order search and penalized regression. *The Annals of Statistics*, 2526–2556.
- Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing* 36(3), 287–314.
- Demiralp, S. and K. D. Hoover (2003). Searching for the causal structure of a vector autoregression. *Oxford Bulletin of Economics and statistics* 65, 745–767.
- Eriksson, J. and V. Koivunen (2004). Identifiability, separability, and uniqueness of linear ICA models. *IEEE Signal Processing Letters* 11(7), 601–604.
- Fan, J. and I. Gijbels (2018). *Local polynomial modelling and its applications*. Routledge.
- Faust, J. (1998). The robustness of identified VAR conclusions about money. In *Carnegie-Rochester conference series on public policy*, Volume 49, pp. 207–244. Elsevier.
- Fiorentini, G. and E. Sentana (2022). Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions. *Journal of Econometrics*.
- Gouriéroux, C., A. Monfort, and J.-P. Renne (2017). Statistical inference for independent component analysis: Application to structural VAR models. *Journal of Econometrics* 196(1), 111–126.
- Gretton, A., K. Fukumizu, C. Teo, L. Song, B. Schölkopf, and A. Smola (2007). A kernel statistical test of independence. *Advances in neural information processing systems* 20.
- Guay, A. (2021). Identification of structural vector autoregressions through higher unconditional moments. *Journal of Econometrics* 225(1), 27–46.
- Guerini, M. and A. Moneta (2017). A method for agent-based models validation. *Journal of Economic Dynamics and Control* 82, 125–141.
- Hastie, T. J. and R. J. Tibshirani (2017). *Generalized additive models*. Routledge.

- Herwartz, H. (2018). Hodges–Lehmann detection of structural shocks—an analysis of macroeconomic dynamics in the Euro area. *Oxford Bulletin of Economics and Statistics*.
- Herwartz, H., A. Lange, and S. Maxand (2022). Data-driven identification in SVARs—when and how can statistical characteristics be used to unravel causal relationships? *Economic Inquiry* 60(2), 668–693.
- Höppner, F., C. Melzer, and T. Neumann (2008). Changing effects of monetary policy in the us—evidence from a time-varying coefficient var. *Applied Economics* 40(18), 2353–2360.
- Hoyer, P., D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf (2008). Nonlinear causal discovery with additive noise models. *Advances in neural information processing systems* 21, 689–696.
- Hussain, S. M. and S. Malik (2016). Asymmetric effects of exogenous tax changes. *Journal of Economic Dynamics and Control* 69, 268–300.
- Hyvärinen, A. (2013). Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371(1984), 20110534.
- Hyvärinen, A., J. Karhunen, and E. Oja (2001). *Independent component analysis*. John Wiley & Sons.
- Kilian, L. (2014). Oil price shocks: Causes and consequences. *Annu. Rev. Resour. Econ.* 6(1), 133–154.
- Kilian, L. and H. Lütkepohl (2017). *Structural vector autoregressive analysis*. Cambridge University Press.
- Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics* 74(1), 119–147.
- Lanne, M. and J. Luoto (2021). GMM estimation of non-gaussian structural vector autoregression. *Journal of Business & Economic Statistics* 39(1), 69–81.
- Lanne, M. and H. Lütkepohl (2008). Identifying monetary policy shocks via changes in volatility. *Journal of Money, Credit and Banking* 40(6), 1131–1149.
- Lanne, M. and H. Lütkepohl (2010). Structural vector autoregressions with nonnormal residuals. *Journal of Business & Economic Statistics* 28(1), 159–168.
- Lanne, M., M. Meitz, and P. Saikkonen (2017). Identification and estimation of non-Gaussian structural vector autoregressions. *Journal of Econometrics* 196(2), 288–304.
- Lewis, D. J. (2021). Identifying shocks via time-varying volatility. *The Review of Economic Studies*.
- Lo, M. C. and J. Piger (2005). Is the response of output to monetary policy asymmetric? evidence from a regime-switching coefficients model. *Journal of Money, credit and Banking*, 865–886.

- Meinshausen, N. and P. Bühlmann (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(4), 417–473.
- Mesters, G. and P. Zwiernik (2022). Non-independent components analysis. *arXiv preprint arXiv:2206.13668*.
- Moneta, A. (2008). Graphical causal models and VARs: an empirical assessment of the real business cycles hypothesis. *Empirical Economics* 35(2), 275–300.
- Moneta, A., D. Entner, P. O. Hoyer, and A. Coad (2013). Causal inference by independent component analysis: Theory and applications. *Oxford Bulletin of Economics and Statistics* 75(5), 705–730.
- Montiel Olea, J. L., M. Plagborg-Møller, and E. Qian (2022). SVAR identification from higher moments: Has the simultaneous causality problem been solved? *AEA Papers and Proceedings* 112, 481–85.
- Montiel Olea, J. L., J. H. Stock, M. W. Watson, et al. (2021). Inference in structural vector autoregressions identified with an external instrument. *Journal of Econometrics* 225(1), 74–87.
- Mooij, J. M., D. Janzing, T. Heskes, and B. Schölkopf (2011). On causal discovery with cyclic additive noise models. *Advances in neural information processing systems* 24.
- Pearl, J. (2009). *Causality. Models, Reasoning, and Inference*. Cambridge University Press.
- Peters, J., D. Janzing, and B. Schölkopf (2017). *Elements of causal inference: foundations and learning algorithms*. The MIT Press.
- Peters, J., J. Mooij, D. Janzing, and B. Schölkopf (2014). Causal discovery with continuous additive noise models. *Journal of Machine Learning Research* 15(1), 2009–2053.
- Ramey, V. A. (2016). Macroeconomic shocks and their propagation. *Handbook of Macroeconomics* 2, 71–162.
- Reiersøl, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica*, 375–389.
- Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics* 85(4), 777–792.
- Romer, C. D. and D. H. Romer (1989). Does monetary policy matter? a new test in the spirit of friedman and schwartz. *NBER macroeconomics annual* 4, 121–170.
- Sentana, E. and G. Fiorentini (2001). Identification, estimation and testing of conditionally heteroskedastic factor models. *Journal of Econometrics* 102(2), 143–164.
- Shimizu, S., P. O. Hoyer, A. Hyvärinen, and A. Kerminen (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research* 7(Oct), 2003–2030.

- Shimizu, S., T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer, and K. Bollen (2011). Directlingam: A direct method for learning a linear non-gaussian structural equation model. *The Journal of Machine Learning Research* 12, 1225–1248.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 1–48.
- Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European economic review* 36(5), 975–1000.
- Sims, C. A. (2020). SVAR identification through heteroskedasticity with misspecified regimes. *Princeton University*.
- Smets, F. and R. Wouters (2007). Shocks and frictions in us business cycles: A Bayesian DSGE approach. *American economic review* 97(3), 586–606.
- Spirtes, P., C. N. Glymour, R. Scheines, and D. Heckerman (2000). *Causation, prediction, and search*. MIT press.
- Stock, J. H. and M. W. Watson (2001). Vector autoregressions. *Journal of Economic perspectives* 15(4), 101–115.
- Swanson, N. R. and C. W. Granger (1997). Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. *Journal of the American Statistical Association* 92(437), 357–367.
- Williams, C. K. and C. E. Rasmussen (2006). *Gaussian processes for machine learning*, Volume 2. MIT press Cambridge, MA.
- Wold, H. O. (1960). A generalization of causal chain models (part iii of a triptych on causal chain systems). *Econometrica: Journal of the Econometric Society*, 443–463.
- Zhang, K. and A. Hyvärinen (2009). On the identifiability of the post-nonlinear causal model. *Proceedings of the 25th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 647–655.

Appendix

A Complete RESIT algorithm

Algorithm 1^{full} VAR + RESIT (Phases 1-2 are from Peters et al., 2014)

- 1: **Input:** A K -dimensional time series vector $(y_{t,1}, \dots, y_{t,K})'$
 - 2: PHASE 0: Estimate the reduced-form model.
 - 3: Estimate a reduced-form time series model of the class GVAR-ANIM, see eq. (1) and extract residuals $\hat{u}_t = (\hat{u}_{t,1}, \dots, \hat{u}_{t,K})'$. To simplify the notation, let us call $\hat{u}_{t,k} \equiv v_k$, for $k = 1, \dots, K$.
 - 4: $S := 1, \dots, K, \pi := [\]$
-

- 5: PHASE 1: Determine topological order.
 - 6: **repeat**
 - 7: **for** $k \in S$ **do**
 - 8: Regress v_k on $\{v_i\}_{i \in S \setminus \{k\}}$ and obtain residuals e_k
 - 9: Measure dependence between e_k and $\{v_i\}_{i \in S \setminus \{k\}}$
 - 10: **end for**
 - 11: Let k^* be the k with the weakest dependence
 - 12: $S := S \setminus k^*$
 - 13: $pa(k^*) := S$
 - 14: $\pi := [k^*, \pi]$
 - 15: **until** $\#S = 0$
 - 16: **Output:** π
-

- 17: PHASE 2 (*pruning*): Remove superfluous edges.
- 18: **for** $k \in \{2, \dots, K\}$ **do**
- 19: **for** $p \in pa(\pi(k))$ **do**
- 20: Regress $v_{\pi(k)}$ on $\{v_i\}_{i \in pa(\pi(k)) \setminus \{p\}}$.
- 21: **if** residuals are independent of $\{v_i\}_{i \in \{\pi(1), \dots, \pi(k-1)\}}$ **then**
- 22: $pa(\pi(k)) := pa(\pi(k)) \setminus \{p\}$
- 23: **end if**
- 24: **end for**
- 25: **end for**
- 26: **Output:** $(pa(1), \dots, pa(K))$

Note: with $pa(k)$ we refer to the set of indices associated to the variables in $Pa(v_k)$, for any k .

B Further simulation results: common-cause and v- structure

Table 2: Average MSE, across 200 simulations, between IRFs estimated using different schemes (i.e. CHOL, TDAG, TTOP, ALG1) and the theoretical IRFs. MSE are calculated over the responses of $y_{t,2}$ and $y_{t,3}$ to the first shock (of one standard-deviation magnitude) and over the first five periods. The contemporaneous structure is the “common cause”. We consider different values of T (sample size) and a (persistence parameter). Standard errors are reported in round brackets.

| | | CHOL | | | ALG1 | | |
|-----|---|----------------------|---------------------|----------------------|---------------------|---------------------|---------------------|
| | T | 250 | 500 | 1000 | 250 | 500 | 1000 |
| a | | | | | | | |
| 0.1 | | 0.722 (9.60E-02) | 0.547 (6.60E-02) | 0.682 (8.30E-02) | 0.069 (9.00E-03) | 0.048 (7.00E-03) | 0.033 (4.00E-03) |
| 0.5 | | 1.403 (1.04E-01) | 1.01 (7.00E-02) | 1.28 (8.80E-02) | 0.185 (1.20E-02) | 0.12 (9.00E-03) | 0.077 (4.00E-03) |
| 0.9 | | 11.926 (8.63E-01) | 8.682 (5.59E-01) | 11.075 (7.02E-01) | 1.394 (9.20E-02) | 1.154 (8.80E-02) | 0.888 (5.90E-02) |
| | | TDAG | | | TTOP | | |
| | T | 250 | 500 | 1000 | 250 | 500 | 1000 |
| a | | | | | | | |
| 0.1 | | 0.056 (6.00E-03) | 0.038 (7.00E-03) | 0.042 (8.00E-03) | 0.05 (5.00E-03) | 0.035 (6.00E-03) | 0.035 (7.00E-03) |
| 0.5 | | 0.155 (8.00E-03) | 0.093 (9.00E-03) | 0.094 (8.00E-03) | 0.14 (7.00E-03) | 0.089 (6.00E-03) | 0.081 (8.00E-03) |
| 0.9 | | 1.304 (9.40E-02) | 0.86 (1.06E-01) | 0.982 (8.50E-02) | 1.114 (7.40E-02) | 0.797 (6.70E-02) | 0.843 (7.80E-02) |

Table 3: Average MSE, across 200 simulations, between IRFs estimated using different schemes (i.e. CHOL, TDAG, TTOP, ALG1) and the theoretical IRFs. MSE are calculated over the responses of $y_{t,2}$ and $y_{t,3}$ to the first shock (of one standard-deviation magnitude) and over the first five periods. The contemporaneous structure is the “v-structure”. We consider different values of T (sample size) and a (persistence parameter). Standard errors are reported in round brackets.

| | | CHOL | | | ALG1 | | |
|-----|---|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | T | 250 | 500 | 1000 | 250 | 500 | 1000 |
| a | | | | | | | |
| 0.1 | | 0.68 (9.60E-02) | 0.509 (6.60E-02) | 0.643 (8.30E-02) | 0.054 (8.00E-03) | 0.035 (4.00E-03) | 0.03 (4.00E-03) |
| 0.5 | | 0.936 (9.90E-02) | 0.678 (6.80E-02) | 0.861 (8.50E-02) | 0.118 (1.10E-02) | 0.064 (5.00E-03) | 0.048 (4.00E-03) |
| 0.9 | | 2.622 (1.69E-01) | 1.892 (1.08E-01) | 2.345 (1.32E-01) | 0.396 (2.40E-02) | 0.319 (1.80E-02) | 0.254 (1.30E-02) |
| | | TDAG | | | TTOP | | |
| | T | 250 | 500 | 1000 | 250 | 500 | 1000 |
| a | | | | | | | |
| 0.1 | | 0.04 (5.00E-03) | 0.027 (5.00E-03) | 0.032 (7.00E-03) | 0.039 (5.00E-03) | 0.028 (5.00E-03) | 0.027 (5.00E-03) |
| 0.5 | | 0.088 (7.00E-03) | 0.049 (6.00E-03) | 0.053 (7.00E-03) | 0.089 (7.00E-03) | 0.053 (6.00E-03) | 0.046 (5.00E-03) |
| 0.9 | | 0.313 (1.90E-02) | 0.23 (1.40E-02) | 0.3 (1.90E-02) | 0.345 (2.00E-02) | 0.237 (1.50E-02) | 0.246 (1.40E-02) |

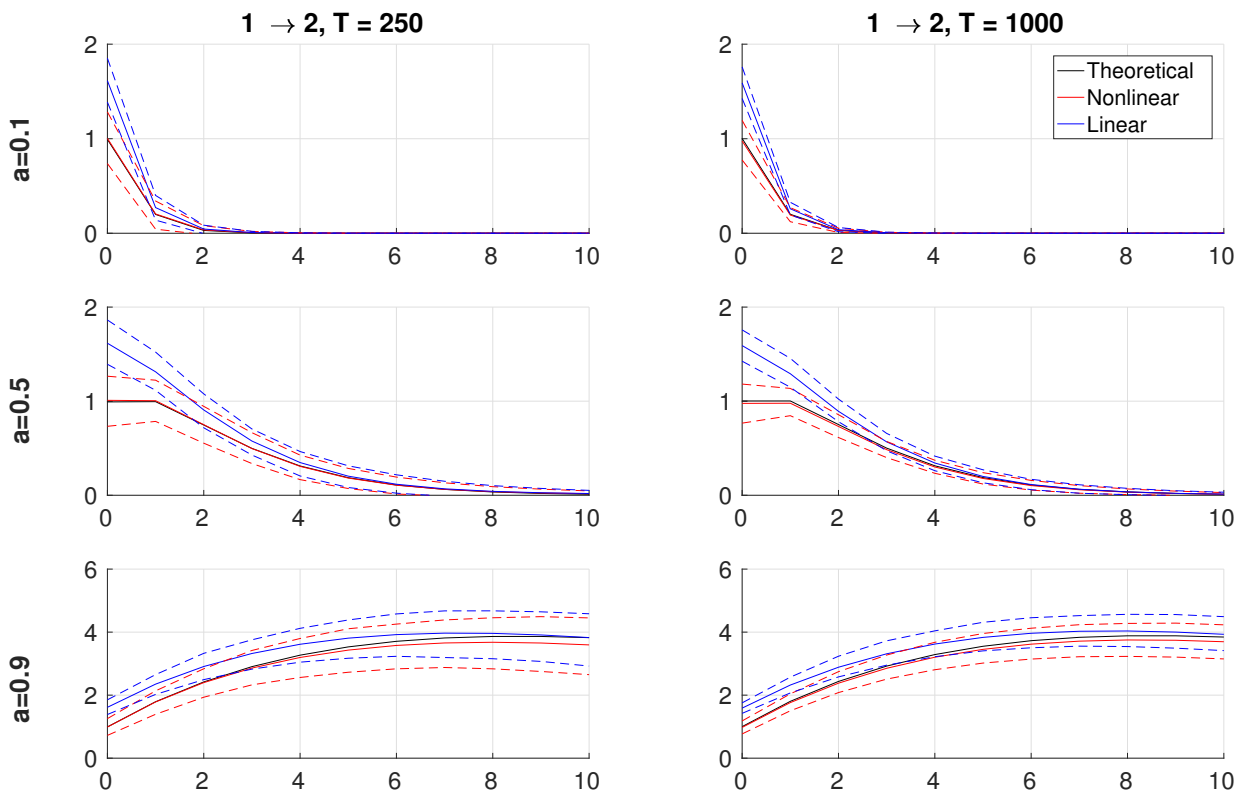


Figure 11: Theoretical, linear (CHOL), and nonlinear (ALG1) IRFs for different values of a and T . Red (Blue) lines exhibits the average (among 200 simulations) for the nonlinear (linear) IRFs response for the response of variable 2 to a unitary shocks of $\varepsilon_{t,1}$ when the “common cause” is selected as the contemporaneous causal structure. Confidence interval at 68% are reported in dotted lines.

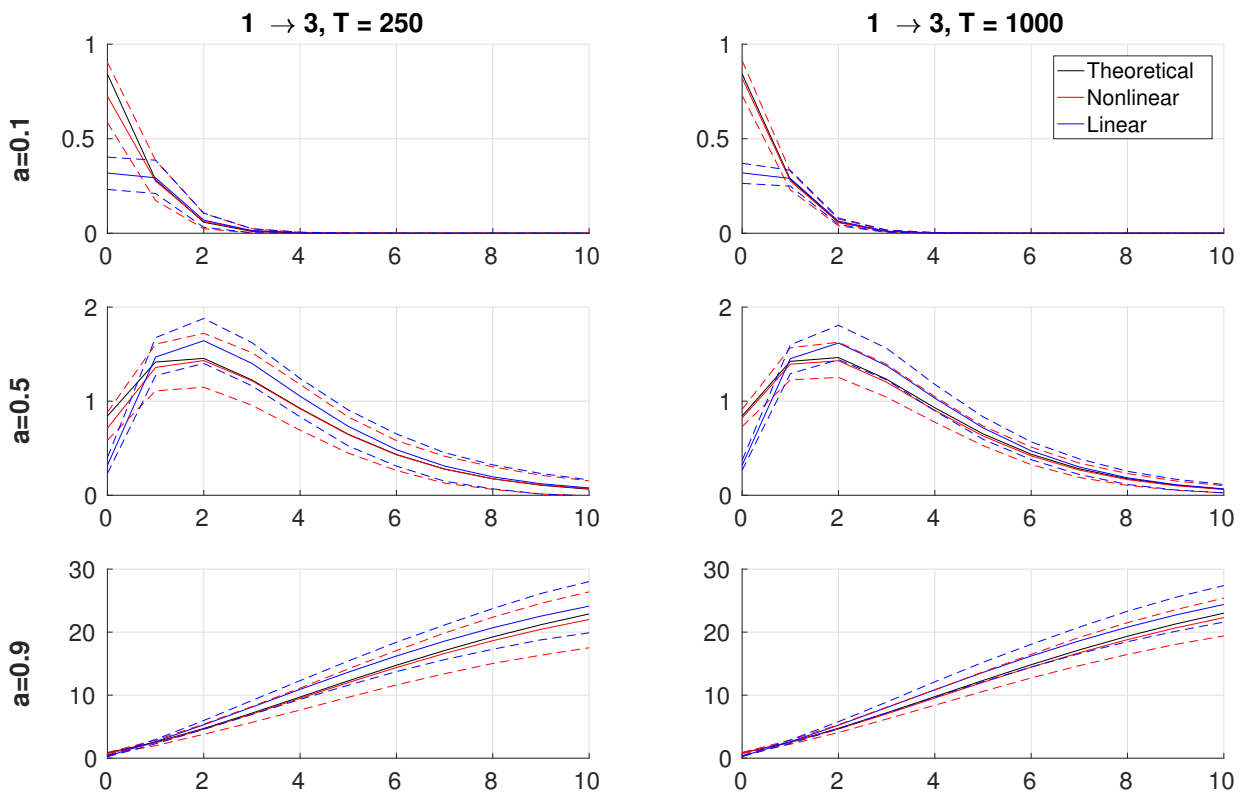


Figure 12: Theoretical, linear (CHOL), and nonlinear (ALG1) IRFs for different values of a and T . Red (Blue) lines exhibits the average (among 200 simulations) for the nonlinear (linear) IRFs response for the response of variable 3 to a unitary shocks of $\varepsilon_{t,1}$ when the “common cause” is selected as the contemporaneous causal structure. Confidence interval at 68% are reported in dotted lines.

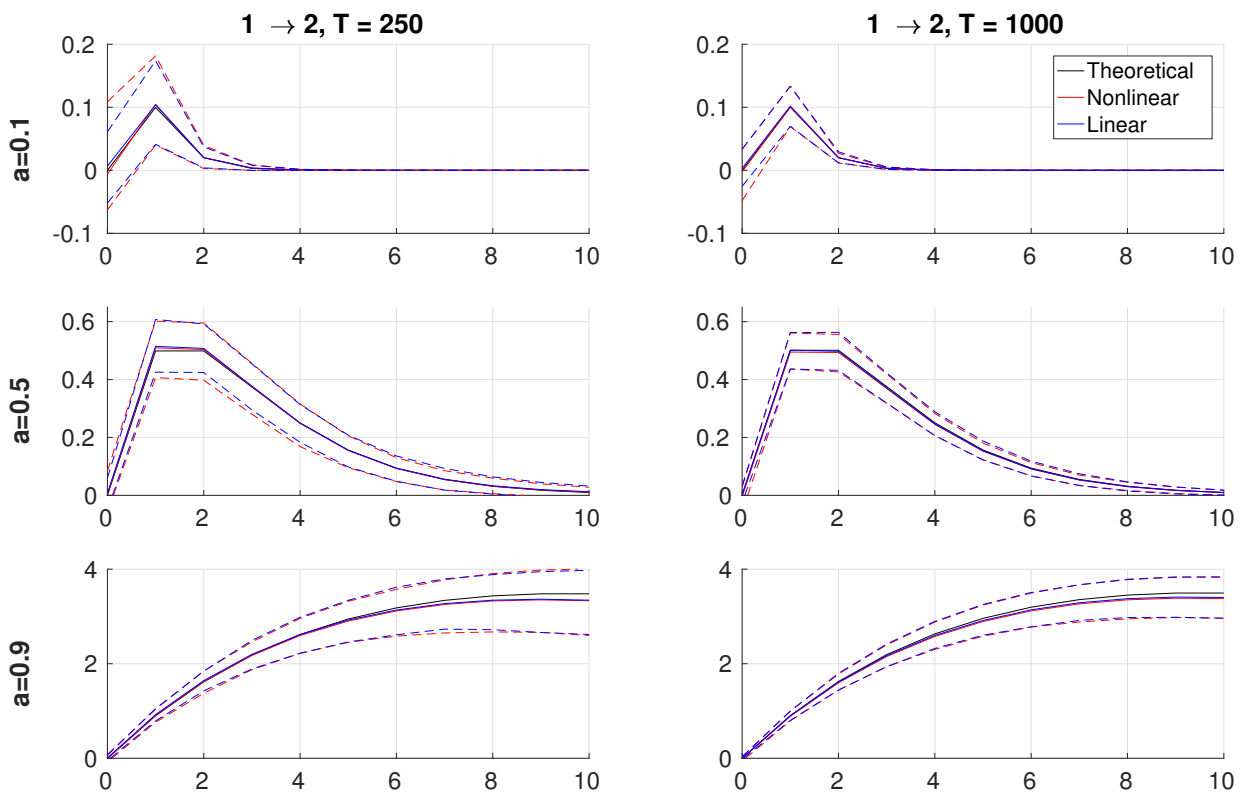


Figure 13: Theoretical, linear (CHOL), and nonlinear (ALG1) IRFs for different values of a and T . Red (Blue) lines exhibits the average (among 200 simulations) for the nonlinear (linear) IRFs response for the response of variable 2 to a unitary shocks of $\varepsilon_{t,1}$ when the “v-structure” is selected as the contemporaneous causal structure. Confidence interval at 68% are reported in dotted lines.

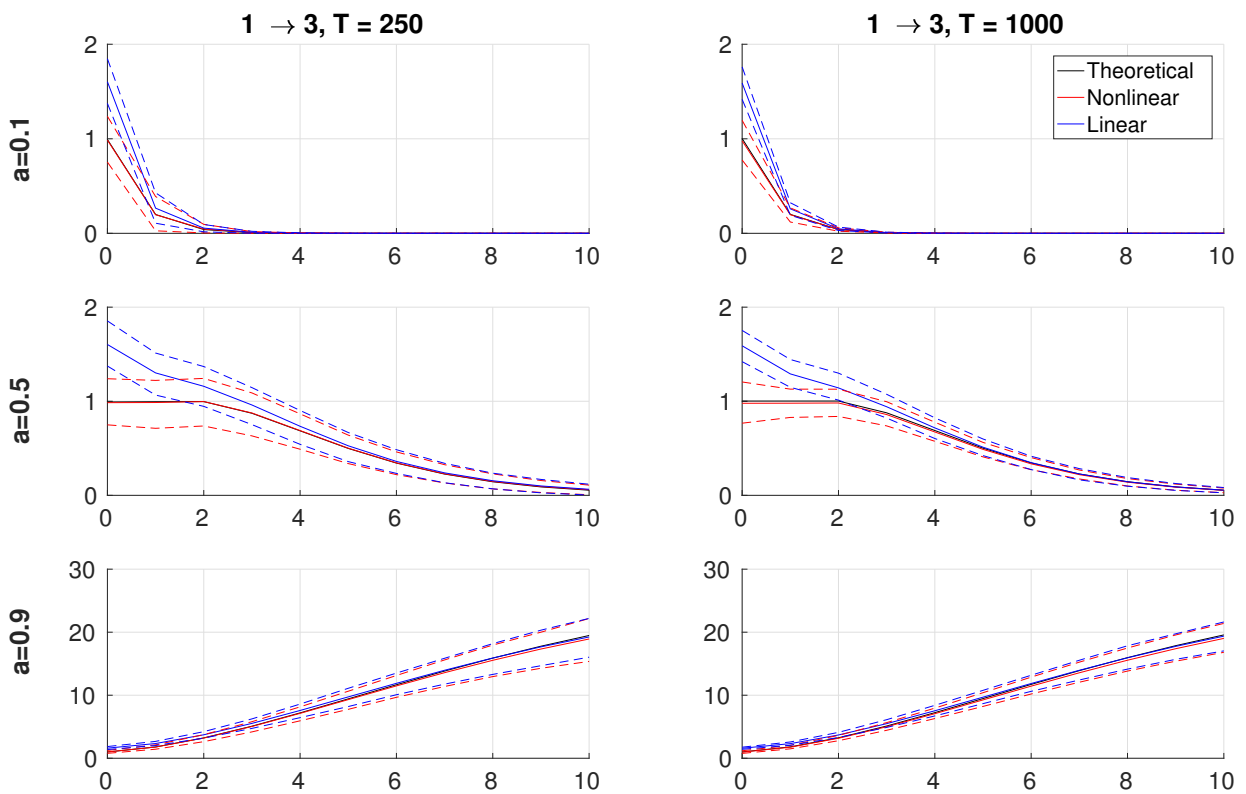


Figure 14: Theoretical, linear (CHOL), and nonlinear (ALG1) IRFs for different values of a and T . Red (Blue) lines exhibits the average (among 200 simulations) for the nonlinear (linear) IRFs response for the response of variable 3 to a unitary shocks of $\varepsilon_{t,1}$ when the “v-structure” is selected as the contemporaneous causal structure. Confidence interval at 68% are reported in dotted lines.

C Codes and data

The R and Matlab codes and the data set to replicate the results reported in this article are available at the following repository:

https://github.com/f-cordoni/NLAM_time_series.git