CEIS TOR VERGATA Research Papers

Università di Roma Tor Vergata

ISSN 2610-931X

CEIS Tor Vergata

RESEARCH PAPER SERIES

Vol. 22, Issue 2, No. 574 – April 2024

Optimization of the Generalized Covariance Estimator in Noncausal Processes

Gianluca Cubadda, Francesco Giancaterini,

Alain Hecq and Joann Jasiak

Optimization of the Generalized Covariance Estimator in Noncausal Processes

Gianluca Cubadda*, Francesco Giancaterini†, Alain Hecq‡, Joann Jasiak§

Abstract

This paper investigates the performance of routinely used optimization algorithms in application to the Generalized Covariance estimator ($GCov$) for univariate and multivariate mixed causal and noncausal models. The $GCov$ is a semi-parametric estimator with an objective function based on nonlinear autocovariances to identify causal and noncausal orders. When the number and type of nonlinear autocovariances included in the objective function are insufficient/inadequate, or the error density is too close to the Gaussian, identification issues can arise. These issues result in local minima in the objective function, which correspond to parameter values associated with incorrect causal and noncausal orders. Then, depending on the starting point and the optimization algorithm employed, the algorithm can converge to a local minimum. The paper proposes the Simulated Annealing (SA) optimization algorithm as an alternative to conventional numerical optimization methods. The results demonstrate that SA performs well in its application to mixed causal and noncausal models, successfully eliminating the effects of local minima. The proposed approach is illustrated by an empirical study of a bivariate series of commodity prices.

Keywords: Mixed causal and noncausal models, Generalized covariance estimator, Simulated Annealing, Optimization, Commodity prices.

*Centre for Economic and International Studies, Tor Vergata University of Rome, Italy
†Centre for Economic and International Studies, Tor Vergata University of Rome, Italy
‡Department of Quantitative Economics, Maastricht University, The Netherlands
§Department of Economics, York University, Canada

1 Introduction

In recent years, there has been a growing interest in employing mixed causal-noncausal processes for the analysis of economic and financial time series (Lanne and Saikkonen (2013), Hecq, Lieb, and Telg (2016), Gourieroux and Jasiak (2022), Gourieroux and Jasiak (2022), Swensen (2022), and Cavaliere, Nielsen, and Rahbek (2020)). These models have gained popularity due to their ability to incorporate both causal and noncausal components to accurately capture the intricate nonlinear dynamics feature of economic and financial processes. In economics, the integration of both causal and noncausal components is particularly valuable when it comes to modeling rational expectations. Unlike traditional autoregressive models, mixed models offer insight into how economic variables are influenced by their expectations and the mechanisms underlying the formation of these expectations (see Lanne and Saikkonen (2011), Lanne and Saikkonen (2013)). In the financial domain, these models capture nonlinear dynamics, including local trends, commonly referred to as speculative bubbles. Speculative bubbles are explosive financial patterns that frequently emerge in highly volatile markets, such as the cryptocurrency and commodity markets. These bubbles represent periods of excessive prices (rates), driven more by market psychology and speculation than by the underlying fundamentals. During such episodes, asset prices rise to unsustainable levels, often followed by a sudden and sharp decline, leading to adverse economic outcomes. Therefore, the identification and analysis of speculative bubbles are of crucial importance to avoid possible disruptions of economic stability and resource allocation (see Gourieroux and Jasiak (2017), Fries and Zakoian (2019), and Hecq and Voisin (2021)).

The estimation techniques available for mixed causal-noncausal processes fall into two main categories: parametric and semi-parametric estimators. The parametric (approximate) maximum likelihood method yields efficient estimates only when the error distribution is specified correctly. In contrast, semi-parametric methods offer the advantage of robustness to specification errors and do not require a distributional assumption on the errors of the model (Gourieroux and Jasiak (2017), Gourieroux and Jasiak (2022), and Hecq and Velasquez-Gaviria (2022)). As a consequence, the benefit of employing a semi-parametric estimator is evident. Currently, $GCov$ is the only semi-parametric estimator in the time domain available for the estimation of mixed causal-noncausal models. It was introduced by Gourieroux and Jasiak (2017) and subsequently extended to a semi-parametrically efficient estimator in Gourieroux and Jasiak (2022). The $GCov$ is characterized by an objective function based on the autocovariances of linear and nonlinear functions of independent and identically distributed ($i.i.d.$) model errors.

This paper aims to address potential challenges associated with local minimum issues in the objective function of the estimator *GCov* applied to mixed causal and noncausal models and optimized by algorithms that are routinely used. In particular, we show that these challenges may arise from difficulties in distinguishing between causal and noncausal dynamics, often linked to factors such as an insufficient number of nonlinear autocovariances, inappropriate nonlinear transformations, or a close to Gaussian error density. As a result, our findings indicate that traditional optimization algorithms, like the Broyden-Fletcher Goldfarb-Shanno (BFGS) algorithm, may struggle to converge to the global minimum in this context when their starting points are poorly selected (see Dennis Jr and Schnabel (1996), Fletcher (2000), and Byrd, Lu, Nocedal, and Zhu (1995)). This difficulty can lead to inaccurate results, underscoring the limitations of conventional optimization methods when applying *GCov* to processes involving a noncausal component.

To avoid optimization problems caused by a potential local minimum occurrence, we propose combining the estimator *GCov* with the Simulated Annealing (SA) optimization algorithm. SA is a powerful metaheuristic method designed to converge to the global minimum when the objective function contains numerous local minima. Originally proposed by Kirkpatrick, Gelatt Jr, and Vecchi (1983), SA draws inspiration from the solid annealing process to address optimization problems. Over the years, SA has shown remarkable success in solving complex optimization problems in various fields, including computer (VLSI) design, image processing, molecular physics, and chemistry (see, for example, Wong, Leong, and Liu (2012), Carnevali, Coletti, and Patarnello (1987), Jones (1991), and Pannetier, Bassas-Alsina, Rodriguez-Carvajal, and Caignaert (1990)). This paper shows that the SA algorithm significantly improves the optimization of the estimator *GCov* in mixed causal-noncausal autoregressive models, ensuring accurate parameter estimates and correct inference on autoregressive orders.

It is important to note that this paper is focused exclusively on the estimation of causal and noncausal parameters. We do not study inference on estimated parameters nor perform portmanteau-type tests *GCov* (see Gourieroux and Jasiak (2022), Jasiak and Neyazi (2023)), and assume that the models are correctly identified and specified. This paper focuses primarily on the application of the *GCov* estimator to multivariate models. For comparison, we also present some results on univariate processes to illustrate graphically the objective function of the estimator displayed as a function of a single parameter. In this context, alternative optimization strategies can be employed to achieve a successful convergence of *GCov*. For example, a grid search strategy over the set of parameter values can be used to find the estimators that minimize the objective function (see Bec, Nielsen, and Saidi (2020)

for a grid search approach in the parametric framework). However, applying this alternative methodology in the multivariate framework can be challenging due to the large dimensions of the grid arrangement.

The paper is organized as follows. Section 2 discusses mixed causal and noncausal models and introduces the $GCov$ estimator. Section 3 shows that its objective function may exhibit local minima under some conditions, which adversely affect the results of BFGS optimization. Section 4 suggests the use of SA to solve the problem of local minima and to provide optimal starting values. Section 5 investigates a bivariate series of commodity prices. Section 6 concludes.

2 GCov estimator of mixed causal and noncausal processes

This section describes the causal-noncausal models and defines the $GCov$ estimator.

2.1 Model representation

This section reviews univariate and multivariate mixed causal-noncausal models.

A strictly stationary univariate mixed causal and noncausal process for a zero-mean series y_t, where $t = 1, 2, \ldots,$ is given by:

$$\phi(L)\varphi(L^{-1})y_t = \eta_t, \tag{1}$$

where the backward-looking polynomial, also defined as the causal polynomial, is given by $\phi(L) = 1-\phi_1 L-\cdots-\phi_r L^r$. On the other hand, the forward-looking polynomial, also defined as the noncausal polynomial, is defined as $\varphi(L^{-1}) = 1 - \varphi_1 L^{-1} - \cdots - \varphi_s L^{-s}$. Furthermore, η_t represents a sequence of random variables $i.i.d.$ with a mean of 0 and a variance of σ^2. In (1), both the causal and noncausal polynomials are characterized by roots outside the unit circle:

$$\phi(z) \neq 0 \;\text{ and }\; \varphi(z) \neq 0 \;\text{ for }\; |z| \leq 1.$$

If $\varphi \neq 0$ for some $j \in \{1, \ldots, s\}$, the process in (1) is defined as purely noncausal if $\phi_1 = \phi_2 = \cdots = \phi_r = 0$. The conventional causal autoregression is obtained when $\varphi_1 = \varphi_2 = \cdots = \varphi_s = 0$.

As shown in Lanne and Saikkonen (2011), the mixed causal and noncausal process expressed in (1) has the following alternative model representation (Breidt, Davis, Lh, and

Rosenblatt (1991)):

$$y_t = \sum_{j=1}^{p} \theta_j y_{t-j} + \epsilon_t,$$

with $\theta(z) \neq 0$ for $|z| = 1$. When $p = r + s$, the autoregressive polynomial can be factored as $\theta(z) = \theta^+(z)\theta^-(z)$, where:

$$\theta^+(z) = 1 - \theta_1^+ z - \cdots - \theta_r^+ z^r \neq 0 \quad \text{for} \quad |z| \leq 1,$$

and

$$\theta^-(z) = 1 - \theta_1^- z - \cdots - \theta_s^- z^s \neq 0 \quad \text{for} \quad |z| \geq 1.$$

This alternative model representation, even if characterized by roots inside the unit circle, is not an explosive process because the error term ϵ_t is not an innovation with respect to the past of y_t, since $\epsilon_t = -(1/\varphi_s)\eta_{t-s}$ (Lanne and Saikkonen (2011)).

Due to the presence of the noncausal component, the process in (1) becomes capable of capturing nonlinear dynamics, including local trends (bubbles) and conditional heteroskedasticity (see Breidt, Davis, Lh, and Rosenblatt (1991), Lanne and Saikkonen (2011), Hencic and Gouriéroux (2015), and Gourieroux and Jasiak (2018)).

As shown in Breidt, Davis, Lh, and Rosenblatt (1991), Lanne and Saikkonen (2011), Gourieroux and Zakoian (2017), mixed causal-noncausal processes defined in (1) admits a unique two-sided strictly stationary solution:

$$y_t = \sum_{j=-\infty}^{\infty} \psi_j \eta_{t+j}, \tag{2}$$

with ψ_0 equal to 1. The two-sided MA representation clarifies that the autoregressive process (1) is mixed, that is, causal-noncausal since the current value of the process y_t is affected by past, present, and future shocks. When the process in (1) is purely causal (resp. noncausal), then $\psi_j = 0$ for all $j > 0$ (resp. $j < 0$) and its current value is affected only by current and past shocks (resp. present and future shocks) (see Breidt, Davis, Lh, and Rosenblatt (1991) and Lanne and Saikkonen (2011)).

Findley (1986) points out that the coefficients of a two-sided moving average representation (including present, past, and future errors) can be distinguished from a one-sided moving average representation (including the past and present errors) only if the error term η_t follows a non-Gaussian distribution. The reason is that the Gaussian distributions are entirely characterized by their second-order moments, which display symmetry over time in stationary processes. Therefore, in Gaussian processes, distinguishing between backward and forward

representations is not possible (see Giancaterini, Hecq, and Morana (2022)). As a consequence, any estimator that relies solely on linear second-order properties, such as the OLS, does not possess the capability to discern this feature. Therefore, mixed causal-noncausal processes can always be represented as purely causal AR(p) (resp. purely noncausal) with the same linear sample autocovariance function as the true data-generating process (DGP). In addition, the causal representation of a noncausal process has autoregressive roots equal to the inverses of autoregressive roots of the DGP that lie inside the unit circle. In general, the sample autocovariance functions are identical for mixed causal-noncausal processes and their representations are obtained by replacing the autoregressive coefficients by the coefficients of autoregressive polynomials with roots equal to the inverses of the true ones (see Breidt, Davis, Lh, and Rosenblatt (1991)). However, among all processes that share the same linear sample autocovariance functions as the true DGP, only the correct specification has serially *i.i.d.* errors. For this reason, in addition to the assumption of non-Gaussianity, the correct identification of the noncausal component also requires serially *i.i.d.* model errors (Hecq, Lieb, and Telg (2016)).

In the multivariate level, the autoregressive representation with roots inside and outside the unit circle and the multiplicative specification in (1) do not always overlap. In particular, as underscored by Davis and Song (2020), Swensen (2022), Cubadda, Hecq, and Voisin (2023) and Gourieroux and Jasiak (2022), the multiplicative representation in (1) does not always exist and covers only a subset of mixed causal-noncausal processes at the multivariate level. This makes the following specification:

$$Y_t = \Theta_1 Y_{t-1} - \dots \Theta_p Y_{t-p} + u_t, \tag{3}$$

with $|\Theta| \neq 0$ for $|z| = 1$, more general compared to the multiplicative representation. Therefore, this paper only considers the specification in (3) at the multivariate level.

Let us assume that the process in (3) is strictly stationary and satisfies the following assumptions:

- **Assumption A.1:** The roots of $\det(\Theta(z))$ are of modulus different from 1.
- **Assumption A.2:** Vectors $u_t, t = 1, ..., T$ are serially i.i.d., non-Gaussian and square-integrable with zero mean $E(u_t) = 0$ and variance–covariance matrix $V(u_t) = \Sigma_u$;

In addition, we suppose that $\det(\Theta(z))$ has n_1 roots outside and $n_2 = n - n_1$ inside the unit circle. As in Gourieroux and Jasiak (2017), we consider the semi-parametric specification of the causal-noncausal model and do not impose any distributional assumptions on the errors,

except for serial independence and non-Gaussian distribution in A.2. Like in the univariate framework, the assumptions of *i.i.d.* and non-Gaussian errors $\{u_t\}_{t=1}^T$, are required to hold for correct identification of the noncausal component of the process with autoregressive roots inside the unit circle (see Davis and Song (2020), Gourieroux and Jasiak (2017)).

The existence of a strictly stationary solution of (3), as well as the two-sided moving average representation of Y, is shown in Gourieroux and Jasiak (2017). We review their Representation Theorem for a Vector Autoregressive process of order 1, VAR(1): $Y_t = \Theta Y_{t-1} + u_t$, satisfying A.1 and A.2 where the matrix Θ is of dimension $n \times n$ and has eigenvalues of modulus different from 1, to set up the notation.

Representation Theorem *(Gourieroux and Jasiak (2017))*: Under Assumptions A.1-A.2, a mixed causal-noncausal *n-dimensional* VAR (1) (with $n \geq 1$), admits a decomposition of the autoregressive matrix Θ with an invertible $(n \times n)$ real matrix A of eigenvectors and two square real matrices: J_1 of dimension $(n_1 \times n_1)$ and J_2 of dimension $(n_2 \times n_2)$ containing the eigenvalues of Θ of modulus strictly less (resp. larger) than 1, and such that:

$$Y_t = A_1 Y_{1,t}^* + A_2 Y_{2,t}^* \tag{4}$$

$$Y_{1,t}^* = J_1 Y_{1,t-1}^* + u_{1,t}^*, \quad Y_{2,t}^* = J_2^{-1} Y_{2,t+1}^* - J_2^{-1} u_{2,t+1}^* \tag{5}$$

$$Y_{1,t}^* = A^1 Y_t, \quad Y_{2,t}^* = A^2 Y_t \tag{6}$$

$$u_{1,t}^* = A^1 u_t, \quad Y_{2,t}^* = A^2 u_t \tag{7}$$

where $[A_1, A_2] = A$ and $[A^{1\prime}, A^{2\prime}]' = A^{-1}$.

In equation (5), the processes $Y_{1,t}^*$ and $Y_{2,t}^*$ are the purely causal and purely noncausal components of the process Y_t, respectively. Any mixed causal-noncausal VAR(p) model (3), with $p \geq 2$, can be written as mixed causal-noncausal VAR(1) by using the companion form as follows (Gourieroux and Jasiak (2017)):

$$X_t = \Psi X_t - 1 + \xi_t,$$

where $X_t = [Y_t, Y_{t-1}, \ldots, Y_{t-p+1}]'$, $\xi_t = [u_t, 0, 0, \ldots, 0]$, and:

$$\Psi = B \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} B^{-1},$$

with B and J containing the eigenvectors and eigenvalues of matrix Ψ, respectively. As a consequence, for $p \geq 2$, we have:

$$X_t = B_1 X_{1,t}^* + B_2 X_{2,t}^*$$

$$X^*_{1,t} = J_1 X^*_{1,t-1} + \xi^*_{1,t}, \quad X^*_{2,t} = J_2^{-1} X^*_{2,t+1} - J_2^{-1}\xi^*_{2,t+1}$$
$$X^*_{1,t} = B^1 Y_t, \quad X^*_{2,t} = B^2 Y_t$$
$$\xi^*_{1,t} = B^1 u_t, \quad X^*_{2,t} = B^2 u_t$$

where $[B_1, B_2] = B$ and $[B^{1\prime}, B^{2\prime}]' = B^{-1}$. Consequently, when $p \geq 2$, the causal and noncausal components are functions of the current and lagged values of Y_t, since $X^*_{1,t} = B^1 X_t = \sum_{h=0}^{p-1} B^1_h Y_{t-h}$ and $X^*_{2,t} = B^2 X_t = \sum_{h=0}^{p-1} B^2_h Y_{t-h}$.

2.2 GCov estimator

Section 2.1 showed that identification of mixed causal-noncausal processes from second-order moments only is not possible. It also noted that among the processes that share the same linear autocovariance function, only the true model has serially *i.i.d.* errors. Although distinguishing the true model based on the linear second-order moments only is not feasible, the second-order moments and second cross-moments of nonlinear functions of *i.i.d.* non-Gaussian errors can identify the process (Chan, Ho, and Tong (2006)). This concept underlies the semi-parametric estimators *GCov* introduced by Gourieroux and Jasiak (2017) and Gourieroux and Jasiak (2022) and denoted *GCov*17 and *GCov*22, respectively. The estimator *GCov*22 minimizes a portmanteau-type objective function involving the autocovariances of nonlinear transformations of model errors viewed as functions of model parameters. For example, the *GCov*22 estimator of the parameter $\theta = vec(\Theta')$ of the strictly stationary n-dimensional mixed causal-noncausal VAR(1) process with $u_t = Y_t - \Theta Y_{t-1}$ minimizes the following portmanteau statistic:

$$\hat{\theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{h=1}^{H} Tr\big[\hat{\Gamma}_a(h;\theta)\hat{\Gamma}_a(0;\theta)^{-1}\hat{\Gamma}_a(h;\theta)'\hat{\Gamma}_a(0;\theta)^{-1}\big], \tag{8}$$

where H is the highest selected lag, $\hat{\Gamma}_a(h;\theta)$ is the sample autocovariance between $a(u_t)$ and $a(u_{t-h})$, with $a(u_t) = \big[a_1(u_t)', \ldots, a_K(u_t)'\big]$, and $a_j(u_t)$ is an element by element function, for $j = 1, \ldots, K$. K indicates the number of linear and nonlinear transformations included in the estimator *GCov*22 and Tr denotes the trace of a matrix. The choice of an informative set of transformations (a_j) depends on the specific series under investigation. Gourieroux and Jasiak (2017) and Gourieroux and Jasiak (2022) explain that this problem is analogous to selecting moments in the Generalized Method of Moments (GMM) estimation or instruments in the Instrumental Variable (IV) estimation. For example, in financial applications that aim to capture the absence of a leverage effect, one can select both linear and quadratic functions. For a bivariate process with $n = 2$, we may consider the following set of four functions

$(K = 4)$: $a_1(u_t) = u_{1,t}$, $a_2(u_t) = u_{2,t}$, $a_3(u_t) = u_{1,t}^2$, and $a_4(u_t) = u_{2,t}^2$. This implies that a_1 and a_2 are linear functions of errors in a causal-noncausal process, while, a_3 transforms the error term of the first variable, $u_{1,t}$, squaring it for each $t = 1, \ldots, T$, where T represents the total number of observations. Similarly, the function a_4 emulates the behavior of a_3, except that it applies the squaring operation to $u_{2,t}$ for each $t = 1, \ldots, T$. Alternatively, we can consider the signs of returns and their squares to separate the volatility dynamics from the bounce effect of the bid-ask: $a_1(u_t) = \text{sign}(u_1)$, $a_2(u_t) = \text{sign}(u_2)$, $a_3(u_t) = u_1^2$, and $a_4(u_t) = u_2^2$. It is important to note that if $a(u_t)$ includes only linear transformations of the error term, then $\hat{\Gamma}(h)$, with $h = 1, \ldots, H$, would provide only information on the second-order linear moments of the process, rendering the estimator unable to identify and estimate the correct specification. Therefore, the nonlinear transformations enable us to estimate the true process. Under the regularity conditions given in Gourieroux and Jasiak (2022), the semi-parametric estimator in (8) is consistent and asymptotically normally distributed when the fourth moments of $a(u_t)$ are finite. Additionally, the $GCov22$ estimator in (8) is semi-parametrically efficient. The matrix on the r.h.s of (8) is diagonalizable, with the sum of its eigenvalues being the sum of the squares of the canonical correlations between $a(u_t)$ and $a(u_{t-h})$, for $h = 1, \ldots, H$.

For comparison, we could estimate the parameters of the n-dimensional VAR(1) by the $GCov17$, which minimizes:

$$\hat{\theta} = \underset{\Theta}{\text{argmin}} \sum_{h=1}^{H} Tr\left[\hat{\Gamma}_a(h;\theta) diag(\hat{\Gamma}_a(0;\theta))^{-1} \hat{\Gamma}_a(h;\theta)' diag(\hat{\Gamma}_a(0;\theta))^{-1}\right], \tag{9}$$

where $diag(\hat{\Gamma}_a(0;\theta))$ is the matrix containing solely the diagonal elements of $\hat{\Gamma}_a(0)$. Therefore, the only difference between (8) and (9) is that $Gcov17$ takes into account only the diagonal elements of the matrix $\hat{\Gamma}_a(0)$. This feature makes $GCov17$ particularly appealing in the high-dimensional framework, when the matrix $\hat{\Gamma}(0)$ is of large dimension, potentially leading to a numerically more stable computation. Like the estimators defined in (8), $GCov17$ is consistent and is normally distributed asymptotically when the fourth-order moments of $a(u_t)$ are finite. In general, estimator (9) is not semi-parametrically efficient, except when the weights in the objective function are $\hat{\Gamma}(0)$, instead of $diag\hat{\Gamma}(0)$ and the estimators in (8) and (9) coincide (see Gourieroux and Jasiak (2017)).

3 BFGS Optimization of Gcov22

This Section examines the behavior of the objective function of the $GCov22$ estimator for different types and numbers K of nonlinear error transformations and illustrates the per-

formance of the BFGS optimization algorithm as implemented in Broyden (1970), Fletcher (1970), Goldfarb (1970), and Shanno (1970).

Our investigation concerns both multivariate and univariate models. As mentioned in Section 1, our analysis of univariate models is intended solely to offer a clear visual representation of the objective function in a two-dimensional Cartesian plane. For this purpose, we consider a purely noncausal AR(1) process. In the multivariate framework, we consider the mixed causal-noncausal VAR(1) process.

The results presented focus exclusively on the $Gcov22$ estimator. This decision is made because both $Gcov22$ and $Gcov17$ perform similarly in terms of accuracy in the univariate and multivariate models considered. Results for $GCov17$ are available upon request.

3.1 The univariate framework

We consider a univariate ($n = 1$) purely noncausal autoregressive process of order 1:

$$y_t = \theta y_{t+1} + \eta_t,\ t = 1, ..., T. \tag{10}$$

where the autoregressive coefficient $|\theta| < 1$ and η_t is a strong ($i.i.d.$) white noise with the t-Student(ν) distribution. This process is a strictly stationary noncausal process and characterized by a root outside the unit circle (Lanne and Saikkonen (2011)). As highlighted in Section 2, the noncausal process in (10) can be written as a causal process $y_t = 1/\theta y_{t-1} + \epsilon_t$ where $E(\epsilon_t\ y_{t-1}) \neq 0$. In other words, the error process ϵ_t is not an innovation process, and the coefficient on y_{t-1} is greater than 1 in absolute value. Consequently, a local minimum at $1/\theta$ can emerge, and we aim to investigate this phenomenon in this section.

To display the objective function of the estimator $GCov22$ of θ and to analyze the possible existence of local minima, we calculate the values of the objective function in (8). This computation is carried out based on the simulated process given above, using $\theta = (0.66, 0.9)$, $\nu = (4, 10)$, $T = 500$, and $H = 10$. This choice of autoregressive coefficients of the DGP allows us to analyze the objective function when the coefficient is close to the unit circle ($\theta = 0.9$) or farther away from it ($\theta = 0.66$). The choice of parameter ν in the error density is intended to examine cases where the process is well identified ($\nu = 4$) and situations where identification issues may arise due to the proximity of the error density to the Gaussian framework ($\nu = 10$).

In empirical investigations, we do not have prior knowledge of the most suitable functions a_j and the appropriate value of K for our dataset. Indeed, as mentioned in Section 2, the choice depends on the specific process under investigation. Therefore, we explore various combinations of a_k and K:

- T0: $a_1(\eta_t)$=η_t;
- T1: $a_1(\eta_t)$=η_t, $a_2(\eta_t)$=η_t^2, $a_3(\eta_t)$=η_t^3, $a_4(\eta_t)$=η_t^4;
- T2: $a_1(\eta_t)$=η_t, $a_2(\eta_t)$=$log(\eta_t^2)$;
- T3: $a_1(\eta_t)$=$sign(\eta_t)$, $a_2(\eta_t)$=η_t^2;
- T4: $a_1(\eta_t)$=$sign(\eta_t)$, $a_2(\eta_t)$=$log(\eta_t^2)$.

We consider $H = 10$ in (8) since this number of lags can capture the correct dynamics in most cases, as reported in Gourieroux and Jasiak (2017) and Gourieroux and Jasiak (2022).

For illustration, the objective function (8) is calculated on a set of the values of the autoregressive coefficients, covering a range of -1 to 5 with a step size of 0.01. For each coefficient within this interval, we calculate and plot the value of the objective function. The objective functions displayed in Figures 1-(a,c) and 2-(a,c) confirm that when only the linear transformations of the error term (T0) are used, the objective function of $GCov22$ exhibits two global minima, associated with the true parameter value (θ) and its causal explosive counterpart ($1/\theta$). Consequently, $GCov22$, in this context, cannot differentiate between the true noncausal and the causal autoregressive processes. The minimum coordinate in the neighborhood of the true parameter value is the maximizer of a Gaussian maximum likelihood function or the OLS estimator of a regression of y_t on its lead or lag.

Next, we consider both linear and nonlinear transformations (T1-T4). The bimodality issue in the objective function of $GCov22$ is alleviated, but not completely resolved. Figures 1-(a,c) show that when the true autoregressive coefficient is 0.66, the objective function has a global minimum at 0.66 but there can remain a local minimum at the incorrect parameter value $1.5 = \theta^{-1}$. Furthermore, Figures 2-(a,c) show that when $\theta = 0.9$ is closer to the unit root, then in addition to the global minimum at 0.9 there is a local minimum corresponding to $1.1 = \theta^{-1}$. In the latter case, it is noticeable that the distance between the local and global minima decreases, making the bimodality issue more problematic. These results remain valid, regardless of the specific choices of a_k and K (similar results were obtained for various a_k and K beyond T0-T4, not presented here but available upon request). Whether the problem with the local minimum is related to the choice of a_j, K, or a combination of both, it is crucial to recognize that in empirical investigations where the optimal selection of these inputs is unknown, the optimization of the objective function can be challenging in the presence of local minima.

Let us illustrate the performance of the BFGS optimization algorithm under these conditions. The BFGS algorithm is a well-known deterministic optimization technique that

approximates the inverse gradient of the objective function to locate the minimum. It starts from the given starting value and iteratively refines this estimate using gradient information and an approximation of the inverse Hessian matrix. We also optimize the objective function related to $GCov22$ with other commonly used numerical optimization algorithms, such as the Nelder-Mead method, the conjugate gradient method, and BFGS with limited memory. We do not report these findings here, but they are available upon request. In Figures 1-(b,d) and 2-(b,d), we present the empirical density function of the estimator $GCov22$, which has been optimized using the BFGS algorithm. The empirical density is derived from Monte Carlo simulations of the noncausal process in equation (8) with $N = 1000$ replications and parameters $\theta = (0.66, 0.9)$, $\nu = (4, 10)$, $T = 500$, and $H = 10$. Specifically, we illustrate scenarios where the starting value of the optimization algorithm corresponds to the local minimum $(0.66^{-1}, 0.9^{-1})$, respectively. Figures 1-(b,d) and 2-(b,d) depict the performance of the BFGS optimization algorithm under these conditions: an erroneous choice of an initial value close to the local minimum may hinder the BFGS optimization algorithm from escaping it and converging towards the global minimum. For results obtained when the true value $\theta = (0.66, 1.5)$ is selected as the initial value, refer to Table 1.

Furthermore, the results indicate that when the BFGS optimizer is applied to $GCov22$, it performs worse when the coefficient is farther away from the unit circle ($\theta = 0.66$). Table 1 shows that a lower percentage of correctly estimated (and identified) models is achieved for $\theta = 0.9$ compared to $\theta = 0.66$. The reason is that when the autoregressive coefficient approaches the unit circle, the distance between the local and global minima decreases, making it more challenging for the optimization algorithm to differentiate between the global minimum and the associated correct coefficient estimator.

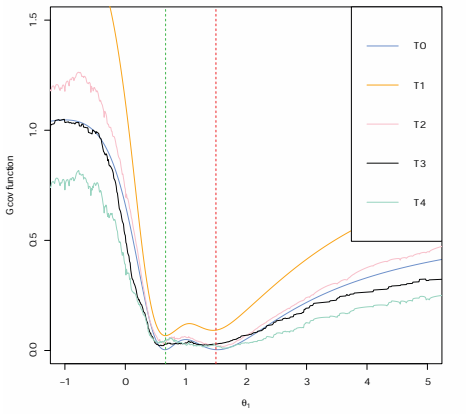
The results suggest that if the stationarity restriction on θ is ignored, local minima issues can arise due to the domain of the estimator $GCov22$ being divided into two sets within this specific DGP: set 1 with coefficients $\theta < 1$ and set 2 with coefficients $\theta > 1$. As a consequence, if the optimization problem starts in set 2, that is, where a local minimum occurs, it is likely that the numerical optimization algorithm will get trapped in that set and converge to the local minimum instead of the global one. Finally, it should be noted that the results illustrated in Figures 1-2 and Table 1 are obtained from simulated length processes $T = 500$. For $T = (1000, 1500)$, the results are slightly better and the local minima disappear asymptotically. These results are not reported but are available upon request.

We conclude that careful selection of an appropriate starting value for an optimization algorithm is essential to ensure accurate parameter estimates. This is even more important in multivariate analysis, as we will show in Section 3.2. Indeed, the larger number of parameters

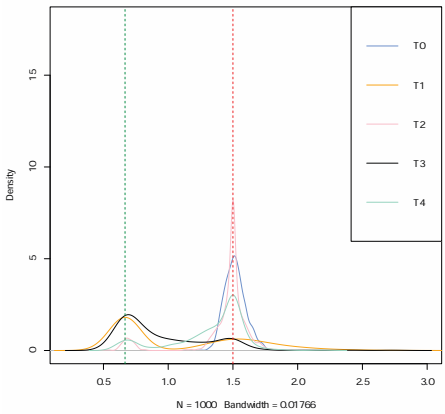
in multivariate analysis adds a layer of complexity, making the task of achieving convergence towards the global minimum a more intricate challenge.

Figure 1: **BFGS optimization of *Gcov*22 in noncausal AR(1) with $\theta = 0.66$, $T = 500$ and OLS starting value**

$\nu = 4$

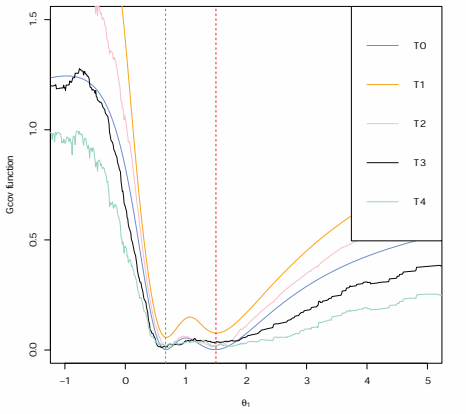


(a) Visualization of the *Gcov*22 objective function. The green and red vertical lines represent the true and inverse values of θ, respectively.

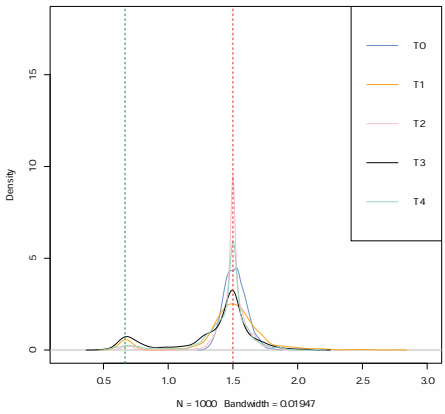


(b) Empirical Density of BFGS optimized $\hat{\theta}$: Monte Carlo with $N = 1000$ replications and $T = 500$, 0.66^{-1} starting value

$\nu = 10$



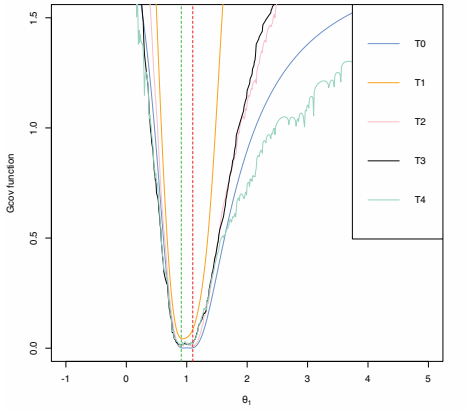
(c) Visualization of the *Gcov*22 objective function. The green and red vertical lines represent the true and inverse values of θ, respectively.



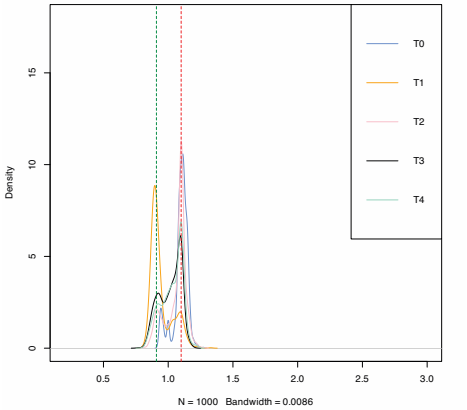
(d) Empirical Density of BFGS optimized $\hat{\theta}$: Monte Carlo with $N = 1000$ replications and $T = 500$, 0.66^{-1} starting value.

Figure 2: **BFGS optimization of *Gcov*22 in noncausal AR(1) with $\theta = 0.9$, $T = 500$ and OLS starting value**

$\nu = 4$

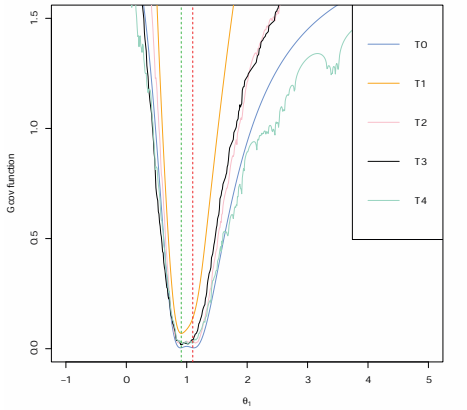


(a) Visualization of the *Gcov*22 objective function. The green and red vertical lines represent the true and inverse values of θ, respectively.

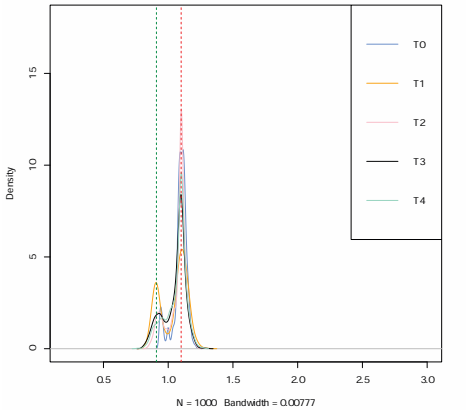


(b) Empirical Density of BFGS optimized $\hat{\theta}$: Monte Carlo with $N = 1000$ replications and $T = 500$, 0.9^{-1} starting value.

$\nu = 10$



(c) Visualization of the *Gcov*22 objective function. The red and green vertical lines represent the true and inverse values of θ, respectively.



(d) Empirical Density of BFGS optimized $\hat{\theta}$: Monte Carlo with $N = 1000$ replications and $T = 500$, 0.9^{-1} starting value.

3.2 The multivariate framework

Let us now investigate the performance of the BFGS-optimized estimator *GCov*22 in multivariate causal-noncausal models. We consider a 3-dimensional mixed causal-noncausal

VAR(1):

$$Y_t = \Theta Y_{t-1} + u_t, \tag{11}$$

where u_t is serially *i.i.d.* with a multivariate t-Student distribution characterized by $\nu = 4$ degrees of freedom and a diagonal variance-covariance matrix Σ_u. The autoregressive matrix is Θ, and referring to Representation Theorem, we consider:

$$A = \begin{bmatrix} 0.8 & 0.7 & 1.2 \\ 1 & 0.9 & 0.8 \\ 0.6 & 0.7 & 0.65 \end{bmatrix}, \quad J = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 2.2 \end{bmatrix},$$

such that:

$$\Theta = AJA^{-1} = \begin{bmatrix} 3.97 & -3.73 & 1.3 \\ 2.29 & -2.38 & 1.41 \\ 1.87 & -2.16 & 1.40 \end{bmatrix}. \tag{12}$$

Hence, the considered process is characterized by two eigenvalues inside the unit circle related to the causal component (j_1 and j_2) that are collected in matrix J_1:

$$J_1 = \begin{bmatrix} j_1 & 0 \\ 0 & j_2 \end{bmatrix} = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.5 \end{bmatrix},$$

and an eigenvalue outside the unit circle (related to the noncausal component): $j_3 = 2.2$. Despite the assumption that (11) exhibits an eigenvalue outside the unit circle, this process is not explosive in light of the Representation Theorem presented in Section 2. In particular, according to equations (4), (5), and (6), the DGP of process (11) is expressed as the following linear combination of its causal and noncausal components:

$$\begin{aligned} Y_t &= A_1 Y_{1,t}^* + A_2 Y_{2,t}^* \\ &= A_1 J_1 A^1 Y_{t-1} + A_2 j_3^{-1} A^2 Y_{t+1} + A_1 A^1 u_t + A_2 j_3^{-1} A^2 u_{t+1}. \end{aligned}$$

Appendix A displays the path and autocorrelation function of a simulated process Y_t.

Next, we explore the performance of the BFGS optimizer of the $GCov22$ estimator in multivariate VAR(1) processes. We calculate the empirical density function of the estimator of matrix Θ by implementing a Monte Carlo experiment with $N = 1000$ replications of the VAR(1) for $T = 500$ observation.

We examine the "worst case scenario" when the OLS estimate of Θ (Θ_{OLS}) is used as the starting value of the BFGS algorithm for the optimization of $GCov22$. The OLS estimator

of a multivariate causal-noncausal model is inconsistent and Θ_{OLS} is potentially associated with local minima in the objective function, since they are characterized by eigenvalues j_1, j_2, and j_3^{-1} (Gourieroux and Jasiak (2017)). Then, in the presence of such a local minimum, the algorithm would have difficulty converge to the global minimum.

Figure 3 and Table 2 show the results obtained from the nonlinear transformations T1-T4. We do not illustrate the linear transformation T0 since, as shown in the previous sections, it is not capable of capturing the true DGP in mixed causal-noncausal processes. It should be noted that in multivariate processes, the number K increases since we implement the transformations of the errors of each series of components. As a consequence, T1 is now characterized by $K = 12$, while all other transformations are characterized by $K = 6$. The results show that, like in the univariate framework, the objective function of $GCov22$ can have local minima at the parameter values of incorrect autoregressive matrices with eigenvalues replaced by their reciprocals. More specifically, Figure 3 illustrates that the empirical density function of $\hat{\Theta}$ is centered on the starting value Θ_{OLS} of a matrix characterized by the eigenvalues j_1, j_2, and j_3^{-1} rather than the population matrix Θ expressed in (12). Then, the BFGS algorithm remains trapped around the starting value. Table 2 summarizes our findings and shows that, as a consequence, the process is predominantly identified erroneously as purely causal by $GCov22$. The transformation $T1$ performs better than the other transformations in identifying the true process (Table 2), although the density function of $\hat{\Theta}$ still focuses mainly around Θ_{OLS} rather than on the true autoregressive matrix.

For comparison, Figure 4 displays the density function of $\hat{\Theta}$ when the BFGS algorithm starts at the starting value equal to an autoregressive matrix with all eigenvalues outside the unit circle, i.e, j_1^{-1}, j_2^{-1}, and j_3. To obtain this starting value, we estimate $Y_t = \Theta Y_{t+1} + u_t$ using OLS and then find the inverse of the estimated matrix, denoted by $\tilde{\Theta}$. The empirical density function of $\hat{\Theta}$ is obtained from a MC experiment with $N = 1000$ replications. Using $\tilde{\Theta}$ as the starting value makes us mistakenly identify the process as purely noncausal most of the time and produces an empirical density function centered on $\tilde{\Theta}$, instead of matrix (12). The BFGS algorithm is again trapped around the inconsistent estimate that serves as the starting value.

Let us now discuss the case when the BFGS algorithm is initiated at the true autoregressive matrix as the starting value. In Figure 5, the empirical density function of the BFGS optimized $GCov22$ estimator is shown. The results highlight that in this scenario the conventional BFGS optimization algorithm successfully converges to the global minimum, producing an empirical density function centered on (12). As a result, the model is correctly identified most of the time, regardless of the nonlinear transformations T1-T4 employed (see

Table 2).

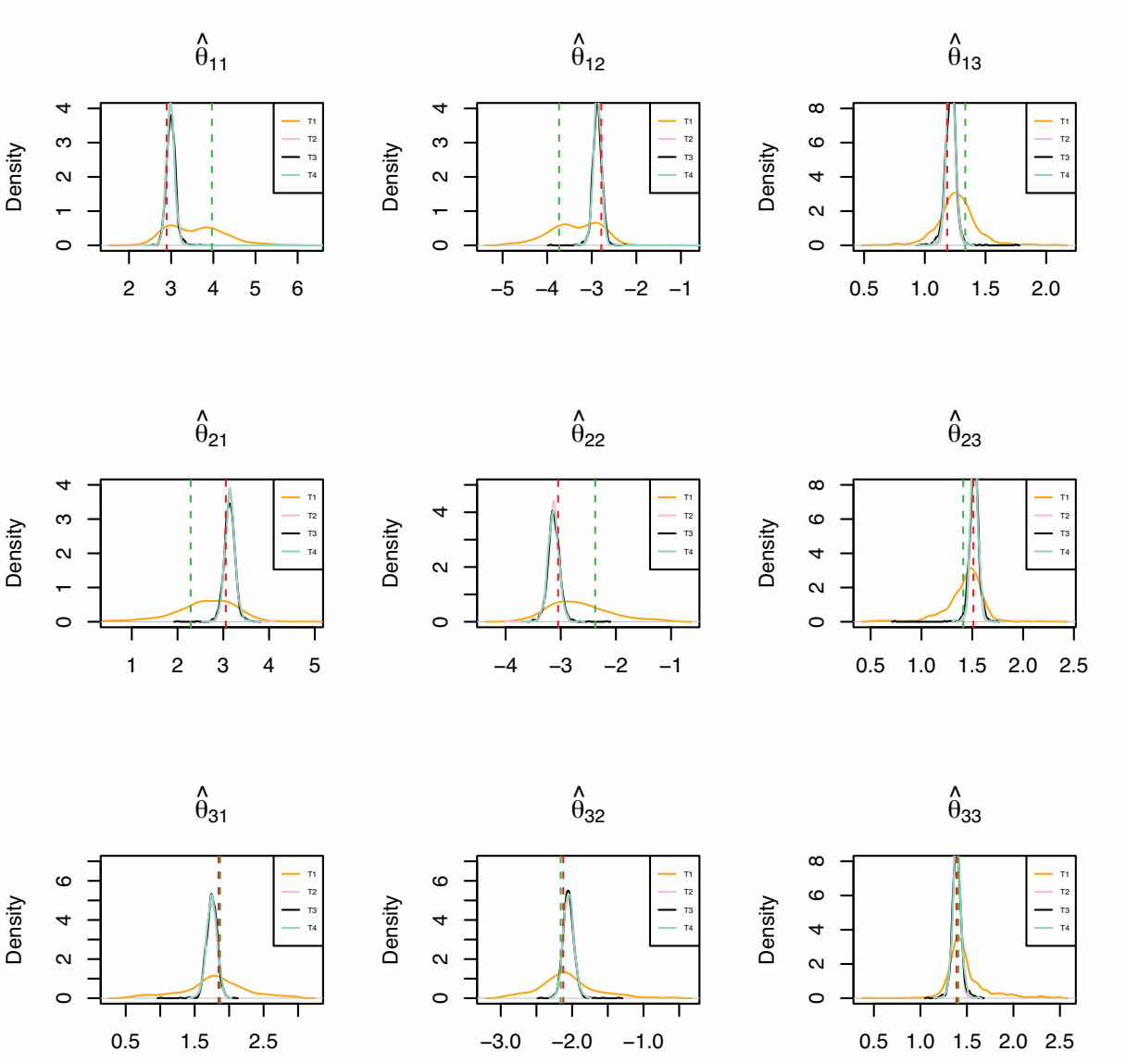
In the multivariate framework, we also explore autoregressive matrices with eigenvalues close to the unit circle and error distributions close to the Gaussian. Our findings about the $GCov22$ objective function resemble those of the previous section: higher degrees of freedom in the t-Student error distribution lead to more identification problems and more pronounced local minimum challenges. In addition, when the eigenvalues are near the unit circle, the distance between the local and global minima is reduced, deteriorating the convergence of the optimization algorithms. Since these results resemble those of Section 3.1, their presentation is omitted but is available upon request.

The above results extend the findings of the univariate framework as follows: the domain of the objective function of the $GCov22$ estimator consists of four sets, characterized by the matrices producing specific roots:

- Set 1: Characterized by all those autoregressive matrices that provide eigenvalues inside the unit circle (resp. roots outside the unit circle);
- Set 2: Characterized by all those autoregressive matrices that provide two eigenvalues inside the unit circle and one eigenvalue inside the unit circle;
- Set 3: Characterized by all those autoregressive matrices that provide one eigenvalue inside the unit circle and two eigenvalues outside the unit circle;
- Set 4: Characterized by all those autoregressive matrices that provide eigenvalues outside the unit circle.

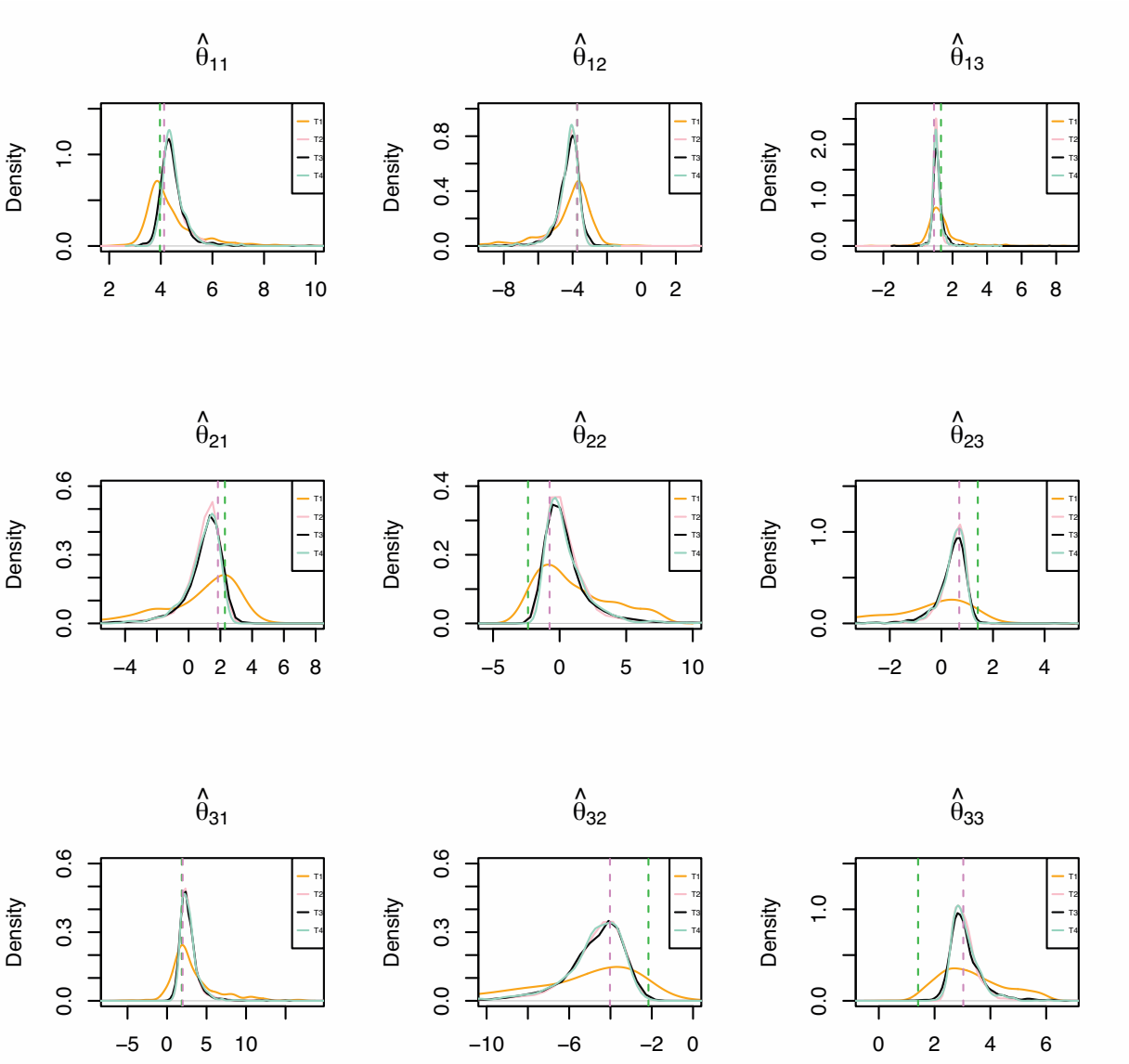
Each set contains a matrix that minimizes the value of the objective function based on the autocovariances of linear functions of model errors. However, when nonlinear transformations of errors are considered, there is a single global minimum associated with the true values of the autoregressive matrix (12), and potentially local minima associated with the parameters of incorrect autoregressive matrices. Therefore, in our case, if the optimization algorithm starts within a set, particularly in proximity to a local minimum (Sets 1-3-4), conventional optimization algorithms are likely to become trapped in that set and converge to the local minimum instead of the global one. On the other hand, successful convergence is always achieved when the starting value is selected from the same set as the global minimum (Set 2). Therefore, the choice of the starting point for the optimization algorithm and the optimization algorithm itself are two crucial steps to avoid identification issues, potentially preventing incorrect identification and estimation of the investigated process.

Figure 3: **Density function of BFGS optimized $\hat{\Theta}$, OLS starting value Θ_{OLS}**



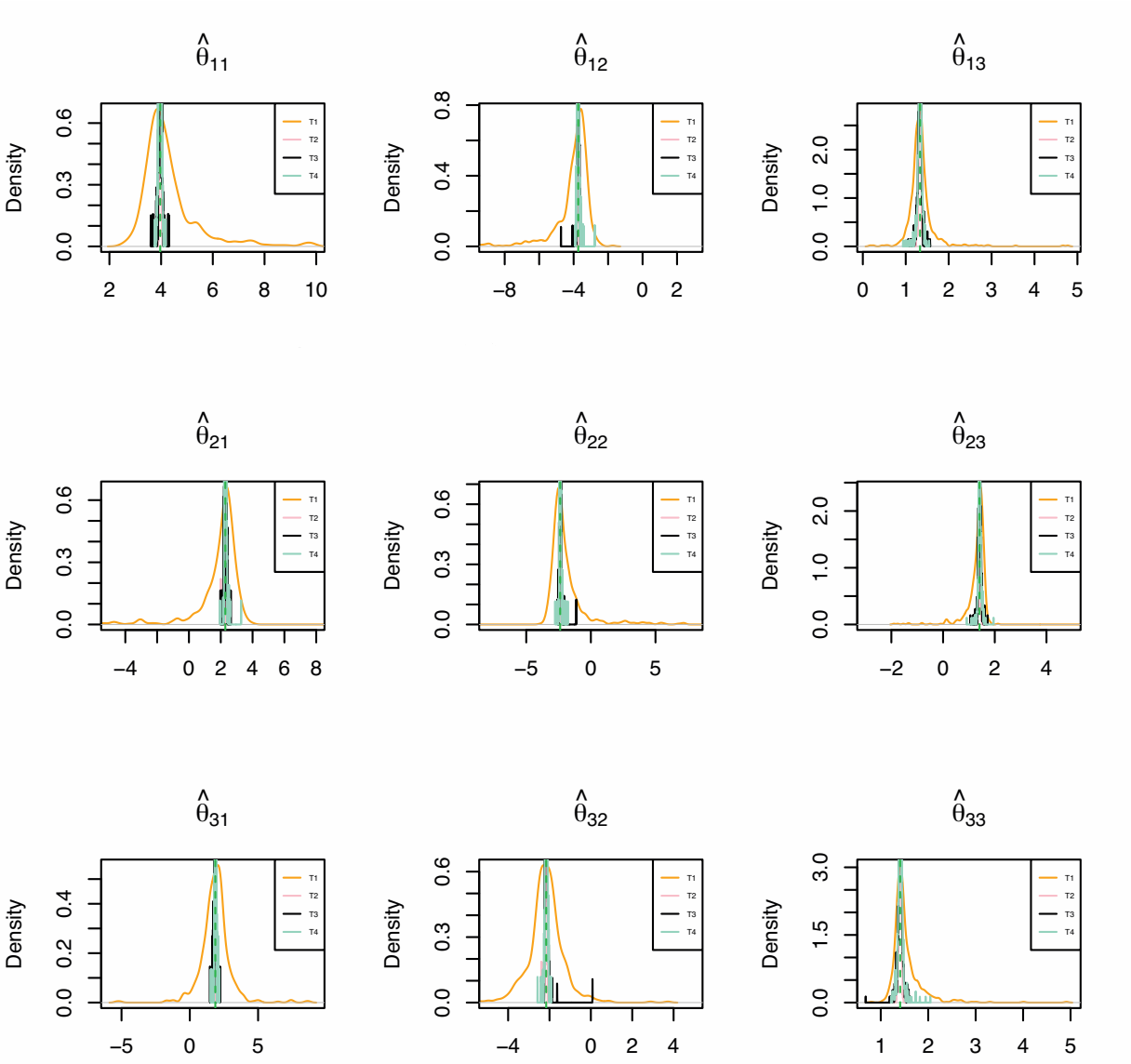
The empirical density function of BFGS optimized $\hat{\Theta}$ estimating the true autoregressive matrix (12), marked with vertical green dashed lines. The starting value for the BFGS optimization algorithm is set at the OLS estimate of the causal counterpart of the matrix (12), that is, Θ_{OLS}, shown with red dashed lines, $T = 500$.

Figure 4: **Density function of BFGS optimized $\hat{\Theta}$: OLS starting value $\tilde{\Theta}$**



The empirical density function of BFGS optimized $\hat{\Theta}$ estimating the true autoregressive matrix (12), marked with vertical green dashed lines. The starting value for the BFGS optimization algorithm is set at the OLS estimate of the noncausal counterpart of the matrix (12), $\tilde{\Theta}$, shown with violet dashed lines, $T = 500$.

Figure 5: **Density function of BFGS optimized $\hat{\Theta}$: starting value true Θ in (12)**



The empirical density function of BFGS optimized $\hat{\Theta}$ estimating the true autoregressive matrix (12), marked with vertical green dashed lines. Here, the matrix in (12) serves a dual purpose as both the population matrix and the starting point for the optimization algorithm

4 Simulated Annealing

In the previous section, we stressed the importance of selecting a starting value of the optimization algorithm that belongs to the same set as the global minimum. Indeed, selecting a

matrix with n_1 and n_2 equal to the true orders, as a starting value, helps achieve a successful convergence of the BFGS optimization algorithm. However, in empirical investigations, determining a priori the number of roots that lie within and outside the unit circle of the population matrix can be challenging. Therefore, in this section, we investigate the performance of the SA optimization algorithm (Kirkpatrick, Gelatt Jr, and Vecchi (1983), Černỳ (1985) and Goffe, Ferrier, and Rogers (1994)) applied to the optimization of the estimator $GCov22$. Our particular focus will be on the nonlinear transformation T1, which, as demonstrated in Section 3 outperforms T2-T4, in addition to being the most commonly employed nonlinear transformation (see Gourieroux and Jasiak (2022) and Gourieroux and Jasiak (2022)).

4.1 The algorithm

SA is an optimization method inspired by the annealing process used in metallurgy. In metallurgy, materials are gradually cooled to eliminate imperfections and achieve a more stable state. The algorithm starts at a high temperature (T^o) and gradually cools over time to reduce the probability of getting stuck at a local minimum. Therefore, in optimization problems, T^o is a parameter that controls the search space exploration during optimization. When T^o is high, the algorithm is more likely to accept worse solutions than the current one, allowing it to escape local optima and explore new areas of the search space. As T^o decreases, the algorithm is less likely to accept suboptimal solutions and converge toward the global optimum. However, if the cooling rate is too high, the algorithm may not be able to escape local minima (see Corana, Marchesi, Martini, and Ridella (1987), Goffe, Ferrier, and Rogers (1992), Goffe, Ferrier, and Rogers (1994), and Goffe (1996)).

Let us now explain how the SA algorithm works when applied to the estimator $GCov22$. We consider a mixed causal-noncausal VAR(1) in (11), and we use f to denote the objective function of $GCov17$ or $GCov22$. Furthermore, we denote the maximum and minimum temperature values of T^o by T^o_{MAX} and T^o_{MIN}, respectively.

To initiate the optimization process, a function evaluation is performed at the randomly selected starting point, denoted Θ^S. Subsequently, a new matrix Θ (Θ') is computed. Specifically, it is determined by adjusting the $ij\text{-}th$ element of the matrix Θ' (θ'_{ij}) using the following equation:

$$\theta'_{ij} = \theta^S_{ij} + m_{ij} \quad \forall i, j = 1, \ldots, n. \tag{13}$$

Here, θ^S_{ij} is the $ij\text{-}th$ element of matrix Θ_S, and m_{ij} is randomly selected from a uniform distribution within the interval $[m_{MIN}, m_{MAX]}$. The value $f(\Theta')$ is then calculated and compared with $f(\Theta^S)$. If $f(\Theta') < f(\Theta^S)$, Θ' is accepted and the algorithm goes downhill.

In the opposite scenario, when $f(\Theta') > f(\Theta^S)$, the potential acceptance of Θ' is determined using the Metropolis criterion. According to this criterion, we compute the variable p^o as follows:

$$p^o = e^{-\frac{(f(\Theta')-f(\Theta^S))}{T^o}}, \tag{14}$$

we then compare it with p^*, that is, a number randomly selected from the range [0, 1]. If $p^o < p^*$, Θ' is rejected, and the algorithm remains at the current point in the function. On the contrary, if $p^o > p^*$, we accept Θ' and move downward. Equation (14) illustrates why a lower value of T^o decreases the probability of making an upward move. To find the optimal solution, the procedure is repeated M times for each T^o, starting from T^o_{MAX} and gradually reducing it at a rate of r, for a total of Q times, until it reaches T^o_{MIN}.

Unlike conventional optimization algorithms, SA can escape local minima (see Corana, Marchesi, Martini, and Ridella (1987), Aarts, Korst, and Michiels (2005)). However, as a drawback, the parameters associated with the SA method, such as θ_{MIN}, θ_{MAX}, T^o_{MAX}, r, Q,, and M, are typically treated as black-box functions and are dependent upon the objective function to be minimized. In empirical studies, a common approach to investigate whether the global minimum has been found is to repeat the algorithm with a different initial state Θ^S. If the same global minimum is reached, it can be concluded with high confidence that convergence has been achieved. In the cases where a different result is obtained, it may be necessary to modify one or more of the parameters involved in the SA algorithm.

4.2 Performance of BFGS with SA starting values: univariate framework

In this section, we evaluate the performance of the SA algorithm for minimizing the objective function of $GCov22$ in mixed univariate causal-noncausal models. To this end, we conducted a Monte Carlo experiment to calculate the empirical density functions of $\hat{\theta}$, while maintaining the same DGP as specified in (10), that is, $\theta = (0.66, 0.9)$, $\nu = (4, 11)$, and $T = 500$. The coefficient θ obtained from SA optimization and known as θ_{SA}, serves as a starting point for the BFGS optimization of the estimator $GCov22$. This strategy offers two distinct benefits. First, it provides an opportunity to explore the impact of different initial value strategies on reaching the global optimum, thus facilitating comparison with the results presented in Figures 1-2 and Table 1. Second, it allows for the refinement of the solution obtained through the SA method. This refinement proves particularly valuable when θ_{SA} is close to the global

minimum, but there is room for improvement in its solution.

As previously mentioned, we begin with an initial temperature of T^o_{MAX}, and at each of the iterations Q, we allow it to decrease at a rate of r. After Q iterations, it reaches the minimum temperature, denoted T^o_{MIN}. It is worth noting that, in our approach, the final temperature T^o_{MIN} is a deterministic function of T^o_{MAX}, r, and Q. This is true because T^o_{MIN} is obtained by Q reductions at a rate of r from the initial value T^o_{MAX}, that is, $T^o_{MIN} = T^o_{MIN}(T^o_{MAX}, r, Q)$.

In the literature, it is common to employ a cooling rate of $r = 0.85$, as indicated in Goffe, Ferrier, and Rogers (1994) and Corana, Marchesi, Martini, and Ridella (1987). However, determining the appropriate values for T^o_{MAX} and Q, which later determine T^o_{MIN}, often requires an empirical approach. Therefore, before conducting our MC experiment, we perform a preliminary analysis of the behavior of DGP under investigation by initially setting T^o_{MAX} and Q at high values. This allows us to monitor the performance of the objective function throughout the optimization process. More specifically, in this preliminary analysis, we set $T^o_{MAX} = 5000$ and $Q = 150$. Figure 6-(a) illustrates the behavior of the value of a minimized objective function of $GCov22$ as a function of the number of iterations Q when $\theta = 0.66$ and $\nu = 4$. The value of the minimized objective function of $GCov22$ calculated from approximately 50 iterations fluctuates around an average value of 2.7 without making any improvements to our optimization problem. This suggests that $T^o_{MAX} = 5000$ is excessively high and results in inefficient time use. Approximately for $Q = 50$, corresponding to $T^o = 1.5$, the value of the minimized value objective function decreases toward the global minimum. These results are further confirmed in Figure 6-(b), which shows the behavior of the $GCov22$ estimator of the parameter θ as a function of Q. Based on the insights gained from this analysis, we set $T_{MAX} = 1.5$ and $Q = 100$ in our MC experiment. Furthermore, to effectively explore the search space, we set $M = 100$. The choice of a high value for M is crucial for a comprehensive exploration of the search space. The results are summarized in Table 1. The SA algorithm yields a significant improvement in the results compared to Section 3.1: $\hat{\theta}$ closely approximates the true value and the true dynamic is captured 90% of the time. It should be noted that cases where $GCov22$ incorrectly identified our process as purely causal can arise from certain replications of our MC experiment when the objective function requires higher values of T^o_{MAX}, Q, M, or a combination of them. As mentioned previously, these parameters are typically problem-specific, and their selection involves experimentation. However, for practical reasons, we maintain the same values for Q and M in all replications.

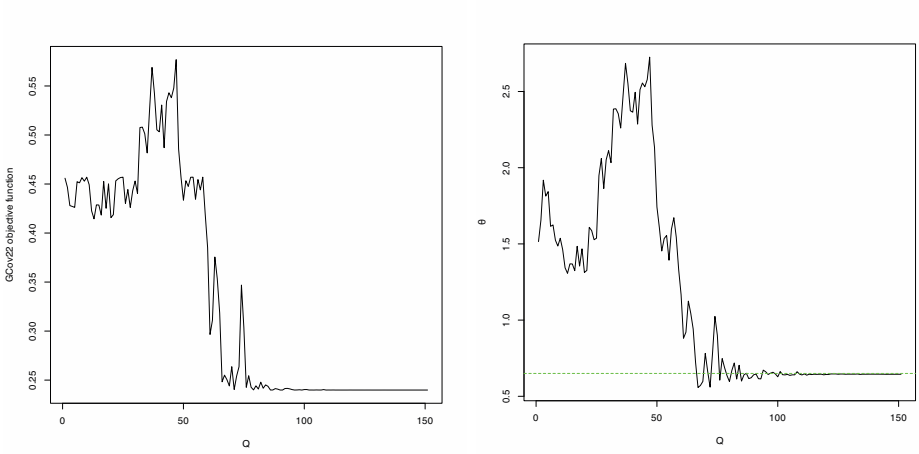
We find that improved results are obtained when SA is implemented for the estimation of the simulated $DGPs$ characterized by different autoregressive coefficients and degrees of freedom (see Table 1) of t-Student error distribution.

4.3 Performance of BFGS with SA starting values: multivariate framework

In this section, we evaluate the performance of the SA algorithm for optimizing the estimator $GCov22$ in multivariate mixed causal-noncausal models. As in Section 3.2, in each Monte Carlo replication, we simulate the time series and then estimate them by the BFGS algorithm, using the SA-provided starting value. For comparison, we maintain the same Monte Carlo input and autoregressive model specified in Section 3.2. In this way, we can explore the impact of different starting value choices on the convergence of BFGS to the global minimum and compare them with the results presented in Figures 3-4-5 and Table 2.

Using the approach employed in Section 4.2 on univariate models, we set $T_{MAX} = 800$, $Q = 200$, and $M = 2000$. The results are shown in Figure 7 and summarized in Table 2. In the multivariate framework, the SA algorithm yields a significant improvement in results compared to Section 3.1. The density of the estimator is now centered on the population value (12). Lastly, as in the previous section, it is worth noting that, for practical reasons, we maintain a constant value for T^o_{MAX}, Q, and M in each replication of the Monte Carlo experiment.

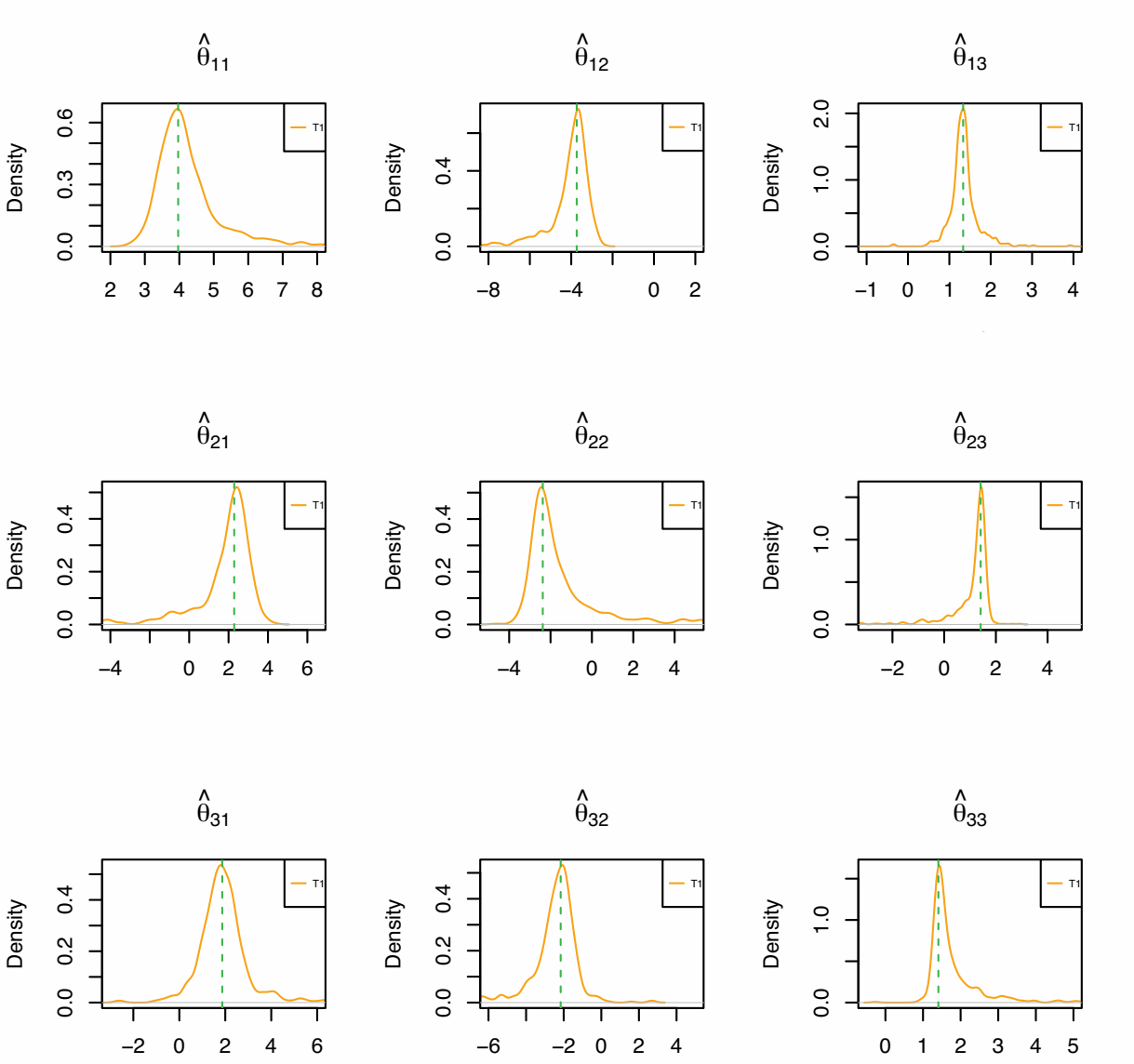
Figure 6: **Performance of SA in univariate noncausal process with $\theta = 0.66$, $\nu = 4$, and $T = 500$.**



(a) *Gcov22 objective function: minimized value as a function of Q, during the SA optimization*

(b) *Gcov22 estimator $\hat{\theta}$ as a function of Q, during the SA optimization*

Figure 7: Empirical density function of $\hat{\Theta}$ with SA starting values



The empirical density function of BFGS optimized Gcov22 with SA starting values and Θ in (12), as the true parameter matrix. The vertical green lines in the plot indicate the corresponding parameter values.

Table 1: **Estimated dynamics: Univariate framework**

a_j	ν	θ	Starting value	T	**Purely noncausal AR(1)**	Purely causal AR(1)
T1	4	0.66	0.66^{-1}	500	61.4%	38.6%
T2	4	0.66	0.66^{-1}	500	16.8%	83.2%
T3	4	0.66	0.66^{-1}	500	74.5%	25.5%
T4	4	0.66	0.66^{-1}	500	31.5%	68.5%
T1	4	0.66	0.66	500	99.4%	0.6%
T1	4	0.66	θ_{SA}	500	94.5%	5.5%
T1	10	0.66	0.66^{-1}	500	13.3%	86.7%
T2	10	0.66	0.66^{-1}	500	6.0%	94.0%
T3	10	0.66	0.66^{-1}	500	24.2%	75.8%
T4	10	0.66	0.66^{-1}	500	14.8%	85.2%
T1	10	0.66	0.66	500	94.6%	5.4%
T1	10	0.66	θ_{SA}	500	74.9%	25.1%
T1	4	0.9	0.9^{-1}	500	70.8%	29.2%
T2	4	0.9	0.9^{-1}	500	32.0%	68.0%
T3	4	0.9	0.9^{-1}	500	41.7%	58.3%
T4	4	0.9	0.9^{-1}	500	34.2%	65.8%
T1	4	0.9	0.9	500	86.3%	13.7%
T1	4	0.9	θ_{SA}	500	82.6%	17.4%
T1	10	0.9	0.9^{-1}	500	36.5%	63.5%
T2	10	0.9	0.9^{-1}	500	21.3%	78.7%
T3	10	0.9	0.9^{-1}	500	26.7%	73.3%
T4	10	0.9	0.9^{-1}	500	25.4%	74.6%
T1	10	0.9	0.9	500	83.9%	16.1%
T1	10	0.9	θ_{SA}	500	73.7%	26.3%

The table illustrates the performance of the BFGS algorithm in optimizing the $GCov22$ estimator of mixed causal-noncausal univariate processes. a_k indicates the linear and nonlinear transformations used, while the starting value of the column indicates the strategy adopted to select the starting value for the optimization algorithm

Table 2: Estimated dynamics: Multivariate framework

a_j	Starting values	VAR($n_1 = 3, n_2 = 0, p = 1$)	**VAR($n_1 = 2, n_2 = 1, p = 1$)**	VAR($n_1 = 1, n_2 = 2, p = 1$)	VAR($n_1 = 0, n_2 = 3, p = 1$)
T1	Θ_{OLS}	34.7%	55%	7.6%	0.7%
T2	Θ_{OLS}	100%	0.0%	0.0%	0.0%
T3	Θ_{OLS}	95.5%	4.3%	0.2%	0.0%
T4	Θ_{OLS}	96.7%	3.2%	0.0%	0.1%
T1	$\tilde{\Theta}$	0.1%	2.3%	48.6%	49%
T2	$\tilde{\Theta}$	0.0%	0.0%	0.0%	100%
T3	$\tilde{\Theta}$	0.0%	0.0%	7.0%	93%
T4	$\tilde{\Theta}$	0.0%	0.0%	0.5%	99.5%
T1	Θ	0.0%	94.5%	3.1%	2.4%
T2	Θ	0.0%	97.5%	2.4%	0.1%
T3	Θ	0.0%	97.6%	2.3%	0.1%
T4	Θ	0.0%	97.3%	2.3%	0.4%
T1	Θ_{SA}	1.4%	64.6%	28.4%	5.6%

The table illustrates the performance of the BFGS algorithm in optimizing the GCov22 estimator of mixed causal-noncausal multivariate VAR processes. a_k indicates the linear and nonlinear transformations used, while the column starting value indicates the strategy adopted to select the starting value for the optimization algorithm. In this table, VAR($n_1, n_2, 1$) indicates a VAR(1), with n_1 roots outside the unit circle and roots n_2 inside the unit circle. Θ is defined as in (12) and $\nu = 4$.

5 Empirical analysis

We conduct an empirical analysis of a bivariate time series consisting of 363 daily observations on the CBOT closing prices of wheat and soybean futures in US Dollars, over the medium term. For this analysis, we use the same data range as Gourieroux and Jasiak (2022), covering the period from October 18, 2016, to March 29, 2018. The dataset was obtained from https://ca.finance.yahoo.com, with the wheat futures represented by the ticker ZW = F and the soybean futures by the ticker ZS = F. Figure 8 shows the demeaned data, while Figure 9 presents the kernel-smoothed density estimators of the series.

Our primary objective is to evaluate the performance of the BFGS optimized estimator *GCov*22. Additionally, we seek to identify the presence of speculative bubbles in agricultural commodity markets. Detecting such bubbles has significant implications for various stakeholders, including market participants, policymakers, and investors, as it directly impacts decision-making in both the agricultural and financial sectors. It should be noted that the examined series does not exhibit global trends or other widespread and persistent explosive patterns. Instead, they display local trends and spikes, often sharing similar patterns with concurrent spikes. To gain insight into the interactions among these variables and determine

whether noncausal components drive these processes, we proceed to estimate a 2-dimensional VAR(p).

Following Gourieroux and Jasiak (2017), we select the value of autoregressive order $p = 2$ that eliminates serial autocorrelation from residuals and apply the BFGS optimized $GCov22$ to the demeaned data. Furthermore, we reject the null hypothesis of the Gaussianity of the residuals of the estimated model.

We use both the OLS estimate of the bivariate process (Θ_{OLS}) and the starting value obtained from the SA as starting points for the BFGS algorithm. The results are presented in Table 3, which indicates that the choice of starting value significantly affects the results. When we use the OLS starting values, the optimized BFGS $GCov22$ identifies the process as a purely causal VAR(2). However, when the SA starting values are used, we obtain a lower value of the objective function $GCov22$, and the bivariate process is identified as a mixed causal and noncausal VAR(2) with three roots outside the unit circle and one root inside the unit circle: $j_1 = 0.972$, $j_2 = 0.88$, $j_3 = 0.604$, $j_4 = -4.355$ and:

$$\hat{\Theta}_1 = \begin{bmatrix} 0.44 & 1.23 \\ 3.31 & -2.34 \end{bmatrix}, \quad \hat{\Theta}_2 = \begin{bmatrix} 0.52 & -1.21 \\ -3.19 & 3.10 \end{bmatrix}.$$

The associated matrix $\hat{B}$ (Section 2.1) is:

$$\hat{B} = \begin{bmatrix} -0.68 & -0.45 & 0.38 & -0.29 \\ -0.15 & -0.47 & 0.34 & 0.92 \\ -0.70 & -0.52 & 0.63 & 0.07 \\ -0.15 & -0.54 & 0.57 & -0.21 \end{bmatrix}.$$

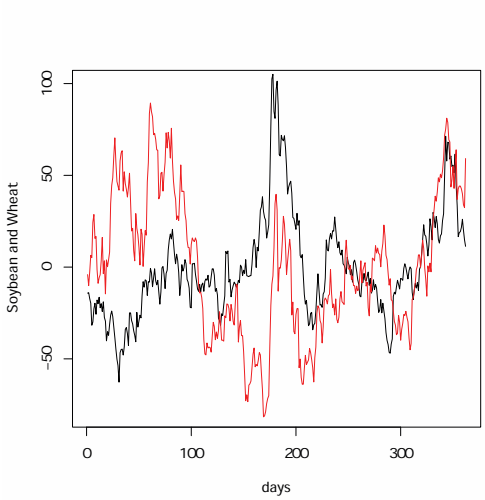
Since j_4 lies outside the unit circle, it implies the simultaneous occurrence of "speculative" bubbles in the two series considered. Furthermore, the negative value of this eigenvalue underscores the fact that these bubbles display fluctuations (see Gourieroux and Zakoian (2017) and Hecq and Voisin (2021)).

After computing B^{-1}, we obtain a noncausal component of dimension 1 representing a common bubble in commodity prices:

$$X^*_{2,t} = 1.54Y_{1,t} - 1.21Y_{2,t} + 0.78Y_{1,t-1} - 0.66Y_{2,t-1}$$

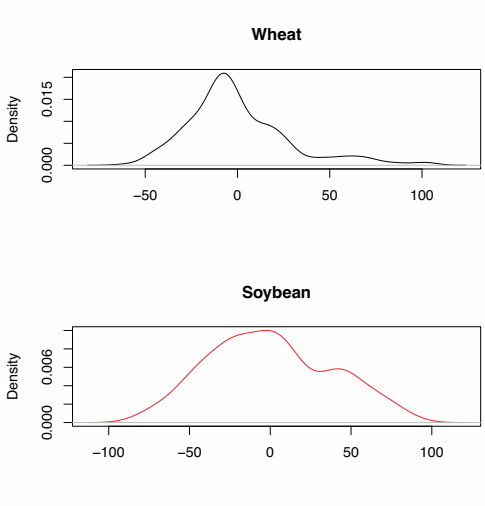
These findings underscore the importance of combining SA with the BFGS optimization of the $GCov22$ or another routinely used optimization algorithm. Employing SA to get the starting value is crucial in this case, enabling us to identify a noncausal component of the process, which, in turn, allows us to capture the nonlinear features that define these series.

Figure 8: **Empirical investigation: wheat and soybean**



The graph shows the demeaned prices of wheat (black line) and soybean (red line) futures from October 18, 2016, to March 29, 2018.

Figure 9: **Marginal sample densities of demeaned daily future price series.**



Marginal sample densities of demeaned daily future price series are non-Gaussian.

Table 3: **Estimated coefficients of mixed bivariate VAR(2)**

Soybean and Wheat							
SV		Θ_1		Θ_2		f.v.	Model
		$\phi_{j,1}$	$\phi_{j,2}$	$\phi_{j,1}$	$\phi_{j,2}$		
Θ_{OLS}	$\phi_{1,j}$	1.16	0.24	-0.29	-0.29	2.00	VAR($n_1 = 4, n_2 = 0, p = 2$)
	$\phi_{2,j}$	1.06	1.05	-0.04	-0.09		
SA	$\phi_{1,j}$	0.44	1.23	0.52	-1.21	1.50	VAR($n_1 = 3, n_2 = 1, p = 2$)
	$\phi_{2,j}$	3.31	-2.34	-3.19	3.10		

*The column "*SV*" and "*f.v.*" display the choice of starting values of the algorithm and the value of the objective function at the minimum. In this table, VAR($n_1, n_2, 1$) indicates a VAR(1), with n_1 roots outside the unit circle and roots n_2 inside the unit circle.*

6 Conclusions

In this paper, we have investigated the performance of the BFGS algorithm for the optimization of the $GCov22$ estimator in mixed causal-noncausal models. The $GCov22$ estimator is a semi-parametric method, which does not require any distributional assumptions on the model errors other than serial independence and non-Gaussianity. It minimizes a portmanteau-type criterion based on nonlinear autocovariances, providing consistent estimates and consequently allowing for the identification of the causal and noncausal orders of the mixed VAR.

Our findings highlight the importance of considering an adequate number and type of nonlinear autocovariances in the objective function of the estimator $GCov22$. When these autocovariances are insufficient or inadequate or when the error density closely resembles the Gaussian distribution, identification issues can arise. This manifests itself in the presence of local minima in the objective function, occurring at parameter values associated with the incorrect causal and noncausal orders. Consequently, the optimization algorithm may converge to a local minimum, leading to inaccurate estimates.

To avoid the optimization problem due to local minima and improve the accuracy of the estimation, we propose the use of the SA optimization algorithm as an alternative to conventional numerical optimization methods. The SA algorithm effectively manages the identification issues caused by local minima, successfully eliminating their effects. By exploring the parameter space more robustly and flexibly, SA provides a reliable solution for obtaining

more accurate estimates of the causal and noncausal orders. However, it is worth noting that, in high-dimensional frameworks, Simulated Annealing (SA) may be time-consuming, due to the computational complexity involved in finding the global minimum. For future research in this framework, it would be beneficial to investigate the optimization of $GCov22$ using more complex optimization algorithms. For example, it would be interesting to explore how Dynamic Mode Decomposition performs in this context (see Tu (2013), Schmid (2010), and Gu, Lin, Lee, and Qiu (2024)). Regardless, one should reestimate the model using several different sets of starting values and compare the minimized values of the objective functions to ensure a correct outcome.

The proposed method is applied to the $GCov22$ estimator of the causal-noncausal vector autoregressive model of a series of bivariate commodity prices. The results highlight the existence of local minima in this application and the advantage of the SA algorithm in providing reliable results in empirical research.

Acknowledgments

The authors would like to thank three anonymous referees and participants of CFE-CMStatistics 2022 in London, IAEE 2023 in Oslo, and ICEEE 2023 in Cagliari for their valuable comments and suggestions. The first two authors acknowledge support from MUR under the 2020WX9AC7 (PRIN 2020) and 20223725WE (PRIN 2022) grants. The first three authors acknowledge support from Maastricht University under a Graduate School of Business and Economics (GSBE) grant. The second and last authors acknowledge support from Mitacs Globalink research funding. The last author acknowledges support from the Natural Sciences and Engineering Council of Canada (NSERC). The usual disclaimers apply.

Appendix

Figure 10: **Graph of the simulated time series in (11)-(12)**

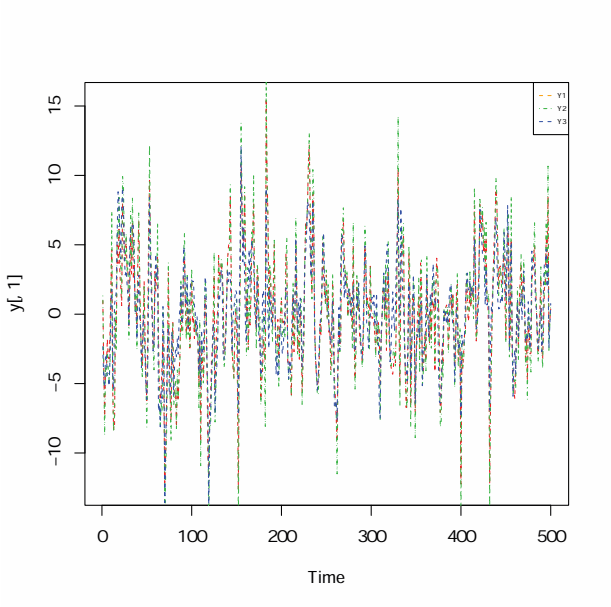
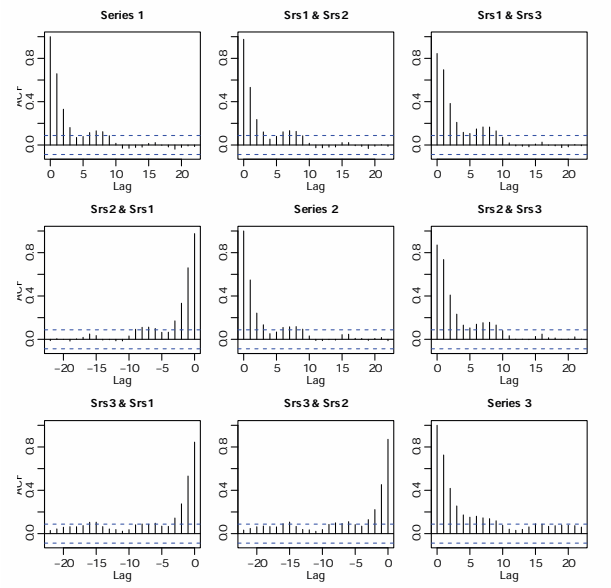


Figure 11: **Autocorrelation function of process generated by (11)-(12)**



References

Aarts, Emile, Jan Korst, and Wil Michiels (2005). "Simulated annealing". In: *Search methodologies: introductory tutorials in optimization and decision support techniques*, pp. 187–210.

Bec, Frederique, Heino Bohn Nielsen, and Sarra Saidi (2020). "Mixed causal–noncausal autoregressions: Bimodality issues in estimation and unit root testing 1". In: *Oxford Bulletin of Economics and Statistics* 82.6, pp. 1413–1428.

Breidt, F Jay, Richard A Davis, Keh-Shin Lh, and Murray Rosenblatt (1991). "Maximum likelihood estimation for noncausal autoregressive processes". In: *Journal of Multivariate Analysis* 36.2, pp. 175–198.

Broyden, Charles George (1970). "The convergence of a class of double-rank minimization algorithms 1. general considerations". In: *IMA Journal of Applied Mathematics* 6.1, pp. 76–90.

Byrd, Richard H, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu (1995). "A limited memory algorithm for bound constrained optimization". In: *SIAM Journal on scientific computing* 16.5, pp. 1190–1208.

Carnevali, Paolo, Lattanzio Coletti, and Stefano Patarnello (1987). "Image processing by simulated annealing". In: *Readings in Computer Vision*. Elsevier, pp. 551–561.

Cavaliere, Giuseppe, Heino Bohn Nielsen, and Anders Rahbek (2020). "Bootstrapping noncausal autoregressions: with applications to explosive bubble modeling". In: *Journal of Business & Economic Statistics* 38.1, pp. 55–67.

Černỳ, Vladimır (1985). "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm". In: *Journal of optimization theory and applications* 45, pp. 41–51.

Chan, Kung-Sik, Lop-Hing Ho, and Howell Tong (2006). "A note on time-reversibility of multivariate linear processes". In: *Biometrika* 93.1, pp. 221–227.

Corana, Angelo, Michele Marchesi, Claudio Martini, and Sandro Ridella (1987). "Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm—Corrigenda for this article is available here". In: *ACM Transactions on Mathematical Software (TOMS)* 13.3, pp. 262–280.

Cubadda, Gianluca, Alain Hecq, and Elisa Voisin (2023). "Detecting Common Bubbles in Multivariate Mixed Causal–Noncausal Models". In: *Econometrics* 11.1, p. 9.

Davis, Richard A and Li Song (2020). "Noncausal vector AR processes with application to economic time series". In: *Journal of Econometrics* 216.1, pp. 246–267.

Dennis Jr, John E and Robert B Schnabel (1996). *Numerical methods for unconstrained optimization and nonlinear equations*. SIAM.

Findley, David F (1986). "The uniqueness of moving average representations with independent and identically distributed random variables for non-Gaussian stationary time series". In: *Biometrika* 73.2, pp. 520–521.

Fletcher, Roger (1970). "A new approach to variable metric algorithms". In: *The computer journal* 13.3, pp. 317–322.

— (2000). *Practical methods of optimization*. John Wiley & Sons.

Fries, Sébastien and Jean-Michel Zakoian (2019). "Mixed causal-noncausal ar processes and the modelling of explosive bubbles". In: *Econometric Theory* 35.6, pp. 1234–1270.

Giancaterini, Francesco, Alain Hecq, and Claudio Morana (2022). "Is climate change time-reversible?" In: *Econometrics* 10.4, p. 36.

Goffe, William L (1996). "SIMANN: A global optimization algorithm using simulated annealing". In: *Studies in Nonlinear Dynamics & Econometrics* 1.3.

Goffe, William L, Gary D Ferrier, and John Rogers (1992). "Simulated annealing: An initial application in econometrics". In: *Computer Science in Economics and Management* 5, pp. 133–146.

— (1994). "Global optimization of statistical functions with simulated annealing". In: *Journal of econometrics* 60.1-2, pp. 65–99.

Goldfarb, D (1970). "A family of variable metric updates derived by variational means, v. 24". In: *Mathematics of Computation*, pp. 21–55.

Gourieroux, Christian and Joann Jasiak (2017). "Noncausal vector autoregressive process: Representation, identification and semi-parametric estimation". In: *Journal of Econometrics* 200.1, pp. 118–134.

— (2018). "Misspecification of noncausal order in autoregressive processes". In: *Journal of Econometrics* 205.1, pp. 226–248.

— (2022a). "Generalized Covariance Estimator". In: *Journal of Business & Economic Statistics* In press, pp. 1–31.

— (2022b). "Nonlinear forecasts and impulse responses for causal-noncausal (S) VAR models". In: *arXiv preprint arXiv:2205.09922*.

Gourieroux, Christian and Jean-Michel Zakoian (2017). "Local explosion modelling by noncausal process". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 79.3, pp. 737–756.

Gu, Mengyang, Yizi Lin, Victor Chang Lee, and Diana Y Qiu (2024). "Probabilistic forecast of nonlinear dynamical systems with uncertainty quantification". In: *Physica D: Nonlinear Phenomena* 457, p. 133938.

Hecq, Alain, Lenard Lieb, and Sean Telg (2016). "Identification of mixed causal-noncausal models in finite samples". In: *Annals of Economics and Statistics/Annales d'Économie et de Statistique* 123/124, pp. 307–331.

Hecq, Alain and Daniel Velasquez-Gaviria (2022). "Spectral estimation for mixed causal-noncausal autoregressive models". In: *arXiv preprint arXiv:2211.13830*.

Hecq, Alain and Elisa Voisin (2021). "Forecasting bubbles with mixed causal-noncausal autoregressive models". In: *Econometrics and Statistics* 20, pp. 29–45.

Hencic, Andrew and Christian Gouriéroux (2015). "Noncausal autoregressive model in application to bitcoin/USD exchange rates". In: *Econometrics of risk*, pp. 17–40.

Jasiak, Joann and Aryan M Neyazi (2023). "GCov-Based Portmanteau Test". In: *arXiv preprint arXiv:2312.05373*.

Jones, Robert O (1991). "Molecular structures from density functional calculations with simulated annealing". In: *Angewandte Chemie International Edition in English* 30.6, pp. 630–640.

Kirkpatrick, Scott, C Daniel Gelatt Jr, and Mario P Vecchi (1983). "Optimization by simulated annealing". In: *science* 220.4598, pp. 671–680.

Lanne, Markku and Pentti Saikkonen (2011). "Noncausal autoregressions for economic time series". In: *Journal of Time Series Econometrics* 3.3.

— (2013). "Noncausal vector autoregression". In: *Econometric Theory* 29.3, pp. 447–481.

Pannetier, Jean, J Bassas-Alsina, Juan Rodriguez-Carvajal, and Vincent Caignaert (1990). "Prediction of crystal structures from crystal chemistry rules by simulated annealing". In: *Nature* 346.6282, pp. 343–345.

Schmid, Peter J (2010). "Dynamic mode decomposition of numerical and experimental data". In: *Journal of fluid mechanics* 656, pp. 5–28.

Shanno, David F (1970). "Conditioning of quasi-Newton methods for function minimization". In: *Mathematics of computation* 24.111, pp. 647–656.

Swensen, Anders (2022). "On causal and non-causal cointegrated vector autoregressive time series". In: *Journal of Time Series Analysis* 43.2, pp. 178–196.

Tu, Jonathan H (2013). "Dynamic mode decomposition: Theory and applications". PhD thesis. Princeton University.

Wong, DF, Hon Wai Leong, and HW Liu (2012). *Simulated annealing for VLSI design*. Vol. 42. Springer Science & Business Media.

RECENT PUBLICATIONS BY *CEIS Tor Vergata*

Caring Connections: Immigrant Caregivers and Long-Term Elderly Care in Italy
Lisa Capretti, Joanna A. Kopinska, Rama Dasi Mariani and Furio Camillo Rosati
CEIS *Research Paper*, 573 April 2024

The Cost of Coming Out
Enzo Brox and Riccardo Di Francesco
CEIS *Research Paper*, 572 April 2024

The Time-Varying Multivariate Autoregressive Index Model
Gianluca Cubadda, Stefano Grassi and Barbara Guardabascio
CEIS *Research Paper*, 571 January 2024

A Further Look at the Gender Gap in Italian Academic Careers
Marianna Brunetti, Annalisa Fabretti and Mariangela Zoli
CEIS *Research Paper*, 570 December 2023

The Wind of Populism: Voter Turnout and Political Distance
Leonardo Becchetti and Gianluigi Conzo
CEIS *Research Paper*, 569 December 2023

Funding Liquidity and Stocks' Market Liquidity: Structural Estimation From High-Frequency Data
Gian Piero Aielli and Davide Pirino
CEIS *Research Paper*, 568 November 2023

On The Nonlinearity of the Finance and Growth Relation: the Role of Human Capital
Alberto Bucci, Boubacar Diallo and Simone Marsiglio
CEIS *Research Paper*, 567 November 2023

Distance Work and Life Satisfaction after the COVID-19 Pandemics
Leonardo Becchetti, Gianluigi Conzo and Fabio Pisani
CEIS *Research Paper*, 566 September 2023

Scitovsky Was Right...and There Is More: Comfort Goods, Stimulus Goods, Education and Subjective Wellbeing
Leonardo Becchetti and Chiara Lubicz
CEIS Research Paper, 565 August 2023

Information Campaigns and migration perceptions
Erminia Florio
CEIS Research Paper, 564 August 2023

DISTRIBUTION

Our publications are available online at www.ceistorvergata.it

DISCLAIMER

The opinions expressed in these publications are the authors' alone and therefore do not necessarily reflect the opinions of the supporters, staff, or boards of CEIS Tor Vergata.

COPYRIGHT

MEDIA INQUIRIES AND INFORMATION

For media inquiries, please contact Barbara Piazzi at +39 06 72595652/01 or by e-mail at piazzi@ceis.uniroma2.it. Our web site, www.ceistorvergata.it, contains more information about Center's events, publications, and staff.

DEVELOPMENT AND SUPPORT

For information about contributing to CEIS Tor Vergata, please contact at +39 06 72595601 or by e-mail at segr.ceis@economia.uniroma2.it