Optimization of the Generalized Covariance Estimator in Noncausal Processes

Gianluca Cubadda[a], Francesco Giancaterini[†b], Alain Hecq[b], Joann Jasiak[c]

[a] *Università di Roma "Tor Vergata", Italy*
[b] *Maastricht University, The Netherlands*
[c] *York University, Canada*

Abstract

This paper investigates the performance of the Generalized Covariance estimator ($GCov$) in estimating mixed causal and noncausal Vector Autoregressive (VAR) models. The $GCov$ estimator is a semi-parametric method that minimizes an objective function without making any assumptions about the error distribution and is based on nonlinear autocovariances to identify the causal and noncausal orders of the mixed VAR. When the number and type of nonlinear autocovariances included in the objective function of a $GCov$ estimator is 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, it is possible for the algorithm to converge to a local minimum, resulting in inaccurate estimates. In this regard, the paper proposes the use of the Simulated Annealing (SA) optimization algorithm as an alternative to conventional numerical optimization methods. The results demonstrate that SA performs effectively when applied to multivariate mixed VAR models, successfully eliminating the effects of local minima. Finally, the effectiveness of the proposed approach is demonstrated through an empirical application involving two bivariate commodity price series.

Keywords: Multivariate causal and noncausal models, Generalized covariance estimator, Simulated Annealing, Optimization, Commodity prices.
JEL: C32

1. Introduction

The causal vector autoregressive model (VAR) was introduced by Sims (Sims (1980)) as an alternative to the simultaneous equation models for macroeconomic variables. The VAR model explains the current value of a vector of

[†]Corresponding author. Email address:f.giancaterini@maastrichtuniversity.nl

macroeconomic variables as a function of their past values and is commonly used by the NBER, central banks, and European institutions.

Let us consider an extension of Sim's model to a n-dimensional strictly stationary VAR of order p, denoted VAR(p):

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

where $Y_t = (y_{1,t}, \ldots, y_{n,t})'$ is a $(n \times 1)$ vector, Θ_i are $(n \times n)$ coefficient matrices, and u_t is a vector $(n \times 1)$ of *i.i.d.* error terms with a non-Gaussian distribution, mean of zero and $Var(u_t) = \Sigma_u$. Whenever the process in (1) is characterized by roots outside the unit circle, that is $\det \Theta(z) \neq 0$ for $|z| \leq 1$, where $\Theta(z) = I_n - \sum_{j=1}^{p} \Theta_j z^j$, then, there exists a strictly stationary purely causal solution to equation (1) with the following one-sided moving average representation:

$$Y_t = \Psi(L) u_t = \sum_{j=0}^{\infty} \Psi_j u_{t-j}, \tag{2}$$

where $\Psi_0 = I_n$, and Ψ_j converge to zero at a geometric rate as $j \to \infty$. In this case, the error term u_t is defined as Y_t-fundamental since it only depends on the present and past values of Y_t (Alessi et al. (2008)). When some of the roots of $\det \Theta(z) = 0$ are inside the unit circle, the VAR stated in (1) is defined as noncausal (Lanne and Saikkonen (2011)). According to Alessi et al. (2008), this may occur because the model developed from the econometric theory is, by definition, nonfundamental or because the econometricians do not observe some relevant variables and thus not included in the analysis (Blanchard and Quah (1988), Lippi and Reichlin (1993), Fernández-Villaverde et al. (2007)). In the presence of a noncausal component, the moving average polynomial $\Psi(L)$ of order infinity in the lag operator L involves both positive and negative powers of L, implying the dependence of Y_t also on future errors (see Alessi et al. (2008), and Lanne and Saikkonen (2013)). Then, the assumption of a non-Gaussian distribution of serially i.i.d. errors is crucial for the identification of the model dynamics because it allows for distinguishing between the roots which are inside or outside the unit circle: $\det \Theta(z) \neq 0$ for $|z| = 1$ of the mixed causal and noncausal VAR process. In practice, the VAR process (1) with noncausal components exhibits nonlinear dynamics, such as local trends (bubbles) and conditional heteroscedasticity observed in the time series of commodity (oil) prices and cryptocurrency rates, for example. Moreover, when the roots inside the unit circle are considered, the matrix polynomial on the right-hand side of equation (1) no longer represents the conditional expectation of the linear model. Indeed, in the context of noncausality, the error term u_t is not orthogonal to the lags of Y_t (see Lanne and Saikkonen (2013), Hecq et al. (2016), Cubadda et al. (2019), and Gourieroux and Jasiak (2017) Gourieroux and Jasiak (2022b)).

Two strategies have been developed to estimate multivariate mixed causal and noncausal models expressed as in (1) with roots inside and/or outside the unit circle: the Approximate Maximum Likelihood Estimator (MLE) (see Davis and Song (2020)) and the Generalized Covariance estimator (see Gourieroux and

Jasiak (2017) and Gourieroux and Jasiak (2022a)). This paper focuses on the Generalized Covariance estimator ($GCov$), which is a semi-parametric estimator designed to minimize an objective function without requiring distributional assumptions on the error term (see Gourieroux and Jasiak (2017) and Gourieroux and Jasiak (2022a)). The objective function involves the nonlinear autocovariances, i.e., the autocovariances of nonlinear functions of model errors.

Our study shows that local minima can appear in the objective function of the GCov estimator when identification issues arise. In particular, we show that these issues may arise due to a difficulty in distinguishing the causal and noncausal dynamics because either the number of nonlinear autocovariances is insufficient or the nonlinear transformations are inadequate, or the error density is too close to the Gaussian distribution. These identification problems lead to local minima, to which the maximization algorithm can converge, depending on the starting point and optimization algorithm. This can occur in the application of the GCov estimator to either univariate or multivariate noncausal processes (however, the main focus of this paper is on the multivariate framework). Therefore, selecting a suitable optimization algorithm and appropriate initial values is crucial to ensure successful convergence. It is worth noting that this problem also arises when alternative parametric estimators are employed to estimate noncausal processes (see respectively Bec et al. (2020) and Hecq and Velasquez-Gaviria (2022) for the MLE and the spectral identification of univariate models).

This paper, to effectively address the issues related to local minima, proposes combining the GCov estimator with the Simulated Annealing (SA) optimization algorithm. SA is a powerful metaheuristic method designed specifically to capture the global minimum when the objective function contains numerous local minima. Initially proposed by Kirkpatrick et al. (1983), SA draws inspiration from the annealing process of solids to tackle optimization problems. Over the years, SA has demonstrated remarkable success in solving complex optimization problems in various fields, including computer (VLSI) design, image processing, molecular physics, and chemistry (see, for instance, Wong et al. (2012), Carnevali et al. (1987), Jones (1991), and Pannetier et al. (1990)). By integrating SA with the GCov estimator, this paper presents a significant improvement in the performance of the GCov estimator within the framework of mixed causal and noncausal models.

The rest of the paper is organized as follows. Section 2 discusses the mixed (causal-noncausal) VAR model. Section 3 introduces the $GCov$ estimator. Section 4 shows that its objective function can admit local minima under some conditions. Section 5 suggests the use SA to overcome the issue of local minima and optimize the choice of initial values. Section 6 investigates some commodity price series. Section 7 concludes.

2. Model representation

This section reviews the univariate and multivariate mixed causal-noncausal models that have been considered in the literature. In particular, it aims at

providing an overview of their key characteristics and alternative specifications.

Univariate mixed causal and noncausal models for a strictly stationary series $y_t, t = 1, ...T$ were introduced by Breidt et al. (1991) as autoregressive models with the feature of having their roots both inside and outside the unit circle:

$$\phi^+(L)\phi^-(L)y_t = \eta_t, \tag{3}$$

where η_t is serially *i.i.d.* and non-Gaussian. In particular, $\phi^+(L) = 1-\phi_1^+L\cdots-\phi_r^+L^r$, also called the causal polynomial, is characterized by r roots outside the unit circle, and $\phi^-(L) = 1 - \phi_1^-L\cdots-\phi_s^-L^s$, denoted as the noncausal polynomial, by s roots inside the unit circle (with $r + s = p$). We assume that y_t has zero mean to simplify the notation. Although representation (3) leads to locally explosive behaviors, the series does not globally diverge because of the correlation between the error term η_t and the noncausal polynomial's explanatory variable, y_{t+s} (see Breid et al. (1991), and Lanne and Saikkonen (2011)).

Lanne and Saikkonen (2011) write the noncausal polynomial as a lead polynomial, resulting in an alternative representation of univariate mixed causal and noncausal models, known as MAR(r,s), such that:

$$\phi^+(L)\varphi(L^{-1})y_t = \varepsilon_t, \tag{4}$$

where ε is *i.i.d.* and non-Gaussian, and $\varphi(L) = 1-\varphi_1L^{-1}-\cdots-\varphi_sL^{-s}$ represents the noncausal component. This alternative representation has the advantage that both backward and forward polynomials are now characterized by roots outside the unit circle (see Hecq et al. (2016), Gourieroux and Jasiak (2016), Gourieroux and Jasiak (2018), Hecq and Voisin (2021), Giancaterini and Hecq (2022)):

$$\phi^+(z) \neq 0 \;\; and \;\; \varphi(z) \neq 0, \quad for \;\; |z| \leq 1.$$

Note that when $\phi_1 = \cdots = \phi_r = 0$ the process in (4) is defined as purely noncausal (MAR(0,s)); on the other hand, when $\varphi_1 = \cdots = \varphi_s = 0$ it is purely causal (MAR(r,0)).

Lanne and Saikkonen (2013) propose the multivariate VMAR(r,s)

$$\Phi(L)\Pi(L^{-1})Y_t = \epsilon_t, \tag{5}$$

where $\Phi(L)$ and $\Pi(L^{-1})$ are matrix polynomials of order r and s respectively. Both polynomials are characterized by roots outside the unit circle:

$$det\Phi(z) \neq 0 \;\; \text{and } det\Pi(z) \neq 0, \quad \text{for } \; |z| \leq 1, \tag{6}$$

and ϵ_t is a sequence of *i.i.d.* random non-Gaussian $(n \times 1)$ vectors. However, unlike the univariate case, the model representation displayed in (5) presents two main differences. The first one is that the matrix multiplication is not commutative, and hence $\bar{\Pi}(L^{-1})\bar{\Phi}(L)Y_t = \bar{\epsilon}_t$ which is observationally equivalent, provides different coefficient matrices than (5). The second point about the multivariate noncausal representation (5) is that, unlike the univariate case, they cannot always be represented by the product of lead and lag polynomials

(see Lanne and Saikkonen (2013), Davis and Song (2020), and Gourieroux and Jasiak (2017).

These two issues have motivated Davis and Song (2020), Gourieroux and Jasiak (2017), Gourieroux and Jasiak (2022a), and Gourieroux and Jasiak (2022b) to consider a VAR(p) expressed as in (1). In this representation, a process is classified as purely causal if the eigenvalues of the companion-form matrix of the process (1), with dimensions of $n \times p$, lie inside the unit circle. Similarly, if the roots of the characteristic equation $|I_n - \Theta_1 z - \cdots - \Theta_p z^p| = 0$ in (1) are located outside the unit circle, it indicates a purely causal process. On the other hand, if the eigenvalues lie outside the unit circle or the roots of the characteristic equation are inside the unit circle, the process is considered purely noncausal. Finally, it is identified as mixed causal and noncausal when the roots, or eigenvalues, lie both inside and outside the unit circle.

Starting from Section 3, we use the notation VAR(n_1, n_2, p) to represent a mixed VAR(p) model with n_1 roots outside and n_2 roots inside the unit circle. Additionally, we define purely causal and non-causal processes as VAR($n_1, 0, p$) and VAR($0, n_2, p$), respectively.

3. Generalized covariance estimator

This section describes the two versions of the *GCov* estimator introduced respectively in Gourieroux and Jasiak (2017) and Gourieroux and Jasiak (2022a). Let us consider a mixed strictly stationary $n - dimensional$ VAR($n_1, n_2, 1$) process:

$$Y_t = \Theta Y_{t-1} + u_t, \tag{7}$$

where the error term u_t is a sequence of *i.i.d.* non-Gaussian random vectors of dimension $n \times 1$. We also assume that u_t are square-integrable, with $E(u_t) = 0$, and a variance-covariance matrix $V(u_t) = \Sigma$. The existence of a strictly stationary solution of (7), as well as the two-sided moving average representation of Y, is provided in Gourieroux and Jasiak (2017).

Representation theorem (*Gourieroux and Jasiak (2017)*): When a n-*dimensional* VAR($n_1, n_2, 1$) is considered, there exists an invertible $(n \times n)$ real matrix A, and two square real matrices: J_1 of dimension $(n_1 \times n_1)$ and J_2 of dimension $(n_2 \times n_2)$ such that all eigenvalues of J_1 (resp. J_2) are those of Θ with a modulus strictly less (resp. larger) than 1, and such that:

$$Y_t = A_1 Y^*_{1,t} + A_2 Y^*_{2,t} \tag{8}$$

$$Y^*_{1,t} = J_1 Y^*_{1,t-1} + u^*_{1,t}, \quad Y^*_{2,t} = J_2^{-1} Y^*_{2,t+1} - J_2^{-1} u^*_{2,t+1} \tag{9}$$

$$Y^*_{1,t} = A^1 Y_t, \quad Y^*_{2,t} = A^2 Y_t \tag{10}$$

$$u^*_{1,t} = A^1 u_t, \quad Y^*_{2,t} = A^2 u_t \tag{11}$$

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

From equation (9), we can see that $Y^*_{1,t}$ and $Y^*_{2,t}$ are, respectively, purely causal and purely noncausal processes. Hence, these two components can be interpreted as the causal and noncausal components of the process Y_t.

In addition, Gourieroux and Jasiak (2017) show that:

$$\Gamma^*_{12} = Cov(Y^*_{1,t}, Y^*_{2,t-h}) = 0 \ for \ h \leq 0. \tag{12}$$

Exploiting the moment condition (12) may yield a suitable VAR(n_1, n_2, p) estimator. However, distinguishing between mixed models $(n_1, n - n_1)$ and $(n - n_1, n_1)$ based solely on linear second-order moments of the component series of Y_t is not possible (Gourieroux and Jasiak (2017)). In this regard, nonlinear covariance-based conditions can be used to identify the model, as long as the error terms u_t are serially independent (Chan et al. (2006)). This led to the development of the $Gcov17$ and $Gcov22$ estimators by Gourieroux and Jasiak (2017) and Gourieroux and Jasiak (2022a), respectively. In particular, the estimator $Gcov22$ aims to minimize

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{h=1}^{H} Tr\big[\hat{\Gamma}(h)\hat{\Gamma}(0)^{-1}\hat{\Gamma}(h)'\hat{\Gamma}(0)^{-1}\big], \tag{13}$$

where H is the highest selected lag, $\hat{\Gamma}(h)$ is the sample autocovariance between $a_j(Y_t - \Theta Y_{t-1})$ and $a_k(Y_{t-h} - \Theta Y_{t-h-1})$, and a_k, $j, k = 1 \ldots, K$ are a set of the element by element functions which depends on the series of interest. In particular, a_k are linear and nonlinear transformations of the error term, and they have to be twice continuously differentiable with respect to Θ. For the asymptotic normality of $GCov22$, finite fourth moments of $a(u_t)$ are needed (see Gourieroux and Jasiak (2022a)).

The matrix in (13) is diagonalizable, with a trace equal to the sum of its eigenvalues, which are the squares of the canonical correlations between $a_j(Y_t - \Theta Y_{t-1})$ and $a_k(Y_t - \Theta Y_{t-1})$.

On the other hand, the estimate provided by $Gcov17$ minimizes:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{h=1}^{H} Tr\Big[\hat{\Gamma}(h)diag(\hat{\gamma}(0))^{-1}\hat{\Gamma}(h)'diag(\hat{\gamma}(0))^{-1}\Big], \tag{14}$$

where $diag(\hat{\gamma}(0))$ is the matrix with only the diagonal elements of $\hat{\Gamma}(0)$. The objective functions (13) and (14) are multivariate portmanteau tests statistics, examined in Cubadda and Hecq (2011).

Note that for $VAR(n_1, n_2, p)$ processes with $p \geq 2$, we can use the companion matrix to rewrite the process as $VAR(n_1, n_2, 1)$ so that the estimators (13) and (14) can easily be applied to the new representation of the model.

The $GCov$ estimator (13) is consistent, and asymptotically normally distributed when the 4th-order moments of $a(u_t)$ are finite. It is also semiparemetrically efficient because of the optimal choice of weights $\widehat{\Gamma}(0)^{-1}$ in the objective function (13). Its semi-parametric efficiency is achieved by an appropriate choice of weights $\widehat{\Gamma}(0)^{-1}$ in the objective functions (13). Moreover, it is a one-step estimator.

The GCov estimator (14) has similar properties except for semi-parametric efficiency. Its advantage is in application to high dimensional matrices when K and n are large, because the objective functions (14) does not require the inversion of $\hat{\Gamma}(0)$ matrices.

4. Finite sample performances of Gcov22

In this section, we investigate the behavior of the objective function associated with the $GCov$ estimator when applied to a 2-dimensional mixed VAR$(n_1, n_2, 1)$ model as described in (7). The underlying assumption is that the population matrix Θ_0 has one eigenvalue (j_1) inside the unit circle and another eigenvalue (j_2) outside it. The error term $u_t = (u_{1,t}, u_{2,t})'$ are serially $i.i.d.$ and follows a multivariate Student's-t distribution with $\nu = 4$ degrees of freedom, and a variance-covariance matrix $\Sigma = \nu/(\nu-2)I_2$. We focus on reporting results for the $Gcov22$ estimator, as both $Gcov22$ and $Gcov17$ perform similarly when Σ is diagonal. However, it is worth noting that $Gcov22$ exhibits greater effectiveness than $Gcov17$ when the assumption of a diagonal covariance matrix is relaxed, as it considers $\Gamma(0)$ as a whole.

Our analysis consists of two cases. In Case 1, in addition to considering a diagonal variance-covariance matrix Σ (i.e., errors contemporaneously cross-sectionally independent), we also take into account a diagonal autoregressive matrix Θ_0 in (7). This setup allows us to evaluate the estimator's performance in both univariate and multivariate scenarios. By examining the objective function of $Gcov22$ with respect to θ_{11} (while keeping the other elements of the matrix constant at their true values), we can analyze its behavior when applied to the purely causal univariate process $y_{1,t} = \theta_{11} y_{1,t-1} + u_t$. Similarly, studying the objective function of $Gcov22$ with respect to θ_{22} (while holding the other elements constant at their true values) provides valuable insights into its performance with the purely noncausal process y_2. On the other hand, to assess the estimator's performance in the multivariate setting, we conduct Monte Carlo experiments and calculate the empirical density function of matrix Θ, considering various initial values for the optimization problem.

In Case 2, while we maintain the assumption of contemporaneously cross-sectionally independent errors, we relax the assumption of a diagonal population matrix Θ_0. We proceed with a similar analysis to explore the implications of this relaxation.

Finally, for each case, we consider two DGPs. In one DGP, the eigenvalues of the population matrix for the causal and noncausal components are significantly smaller and larger than the modulus of 1, respectively. In the other DGP, the eigenvalues are slightly smaller and larger than the modulus of 1.

4.1. ***Case 1: Analysis of a 2-dimensional process with independent variables***

4.1.1. ***Presence of Bimodality Issue***

We examine a $2-dimensional$ mixed VAR(1,1,1) process, denoted as (7), where the cleaned Y_t components $y_{1,t}$ and $y_{2,t}$ are purely causal and noncausal

processes, respectively. Specifically, the autoregressive matrix exhibits a Jordan decomposition with the following coefficient matrices:

$$\begin{aligned}\Theta_0 &= AJA^{-1} \\ &= \begin{bmatrix} 2.1 & 0 \\ 0 & 1.5 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2.1 & 0 \\ 0 & 1.5 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix}.\end{aligned} \tag{15}$$

It is important to emphasize, based on Proposition 4 in Gourieroux and Jasiak (2017), that distinguishing the true bivariate mixed process from either the purely causal process (with eigenvalues j_1 and j_2^{-1}) or the purely noncausal process (with eigenvalues j_1^{-1} and j_2) solely based on the linear autocovariance function becomes impossible when both Σ and Θ_0 are diagonal. We denote the matrix with eigenvalues j_1 and j_2^{-1} as the causal counterpart of Θ and the matrix with eigenvalues j_1^{-1} and j_2 as the noncausal counterpart of the population matrix. Furthermore, as previously stated, the diagonality of the two matrices allows us to investigate how the estimator performs in the univariate framework, i.e. when expressed as a function of θ_{11} (or θ_{22}). Figure 1 shows the $one-dimensional$ $Gcov22$'s objective function with $T = 1000$ observations, $H = 10$ in (13) and using the following nonlinear transformations (a_k):

- T1: $a_1(u)$=u_1, $a_2(u)$=u_2, $a_3(u)$=u_1^2, $a_4(u)$=u_2^2, $a_5(u)$=u_1^3, $a_6(u)$=u_2^3, $a_7(u)$=u_1^4, $a_8(u)$=u_2^4;
- T2: $a_1(u)$=u_1, $a_2(u)$=u_2, $a_3(u)$=$log(u_1^2)$, $a_4(u)$=$log(u_2^2)$
- T3: $a_1(u)$=$sign(u_1)$, $a_2(u)$=$sign(u_2)$, $a_3(u)$=u_1^2, $a_4(u)$=u_2^2,
- T4: $a_1(u)$=$sign(u_1)$, $a_2(u)$=$sign(u_2)$, $a_3(u)$=$log(u_1^2)$, $a_4(u)$=$log(u_2^2)$.

Our findings indicate that the objective function of $Gcov22$ displays local minima when applied to the univariate framework. Specifically, we observe that for the causal process y_1, the objective function exhibits a local minimum that corresponds to its noncausal counterpart j_1^{-1} (or θ_{11}^{-1}), while for the noncausal process y_2, the objective function has a local minimum that corresponds to its causal counterpart j_2^{-1} (or θ_{22}^{-1}). The reason behind this is that, as mentioned in Section 3, also in the univariate framework, it is not possible to distinguish between a pure causal representation and a pure noncausal representation based solely on knowledge of the linear second-order moment of Y. To accurately determine the correct specification, it is necessary to consider nonlinear second-order moments. The $GCov$ estimator effectively leverages this characteristic. However, it is crucial to note that if the number of nonlinear autocovariances considered is insufficient or if the chosen nonlinearities are inadequate, it may result in the presence of local minima during the estimation process. Finally, Figure 1 shows that T1 is the nonlinear transformation that better performs in amplifying the difference between the global and local minima, improving the

estimator accuracy in identifying the process.

To gain insights into the performance of the estimator in the multivariate setting, we conduct three Monte Carlo experiments. In each experiment, we compute the empirical density function of Θ ($\hat{\Theta}$) using a different initial point for the optimization problem. We consider $T = 1000$ observations and $N = 1000$ replications. Furthermore, we discard the initial and final 10% of observations from each simulated time series and set $H = 10$ in equation (13). Finally, we employ the Broyden Fletcher Goldfarb Shanno (BFGS) numerical optimization algorithm. The BFGS algorithm is a well-known deterministic optimization method that approximates the inverse gradient of the objective function to locate the minimum. The algorithm starts with an initial minimum estimate and gradually improves this estimation using gradient information and an approximation of the inverse Hessian matrix. We also implemented other conventional numerical optimization algorithms in our simulation studies, such as the Nelder-Mead, conjugate gradient method, and limited-memory BFGS. However, since they provide similar results, they are available only upon request.

Figure 2 shows the density function of $\hat{\Theta}$ when the causal counterpart of Θ_0 (Θ_C) is used as the initial point for the optimization problem, with:

$$\Theta_C = \begin{bmatrix} j_1 & 0.1 \\ 0.1 & j_2^{-1} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}. \tag{16}$$

Figure 3 shows the estimator's performance when Θ_0 is selected as the starting point. Figure 4 shows the density function of $\hat{\Theta}$ when the noncausal counterpart of Θ_0 (Θ_{NC}) is used as the initial point:

$$\Theta_{NC} = \begin{bmatrix} j_1^{-1} & 0 \\ 0 & j_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}. \tag{17}$$

Our findings reveal that when Θ_C is selected as the initial point, the empirical density function of $\hat{\Theta}$ shows two peaks (Figure 2). The shorter peak corresponds to the population matrix Θ_0, while the higher peak corresponds to its causal counterpart Θ_C. Therefore, the process is often identified as purely causal in this framework (see Table 1). On the other hand, selecting Θ_{NC} as the initial point also yields two peaks in the empirical density function of $\hat{\Theta}$. The smaller one remains associated with Θ_0, but the higher peak is now associated with Θ_{NC}. As a result, Table 1 indicates that the process is often identified as purely noncausal under these conditions. Finally, selecting Θ_0 as the starting point for the optimization problem leads to a significant improvement in the estimator's performance. The numerical algorithm BFGS can now capture the global minimum, resulting in the disappearance of any local minimum.

The findings reveal that the objective function domain of the *Gcov*22 estimator consists of three sets, characterized by the matrices yielding specific roots:

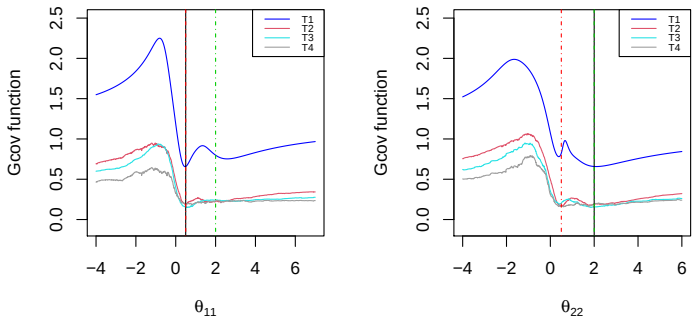
- Set 1: Matrices satisfying $det\Theta(z) \neq 0$ for $|z_1| \leq 1$ and $|z_2| \leq 1$;
- Set 2: Matrices satisfying $det\Theta(z) \neq 0$ for $|z_1| \geq 1$ and $|z_2| \geq 1$;

- Set 3: Matrices satisfying $det\Theta(z) \neq 0$ for $|z_1| \leq 1$ and $|z_2| \geq 1$, or $|z_1| \geq 1$ and $|z_2| \leq 1$.

Each set contains a matrix that minimizes the objective function when considering linear transformations of the autocovariance, that is, Θ_C, Θ_0, and Θ_{NC}. However, when nonlinear transformations of the autocovariance of the error term are considered in (13), only Θ_0 represents the global minimum, while the causal and noncausal counterparts of Θ_0 act as local minima. Indeed, Figures 2-4 show that in the *Gcov*22's domain set where $\det\Theta(z) \neq 0$ for $|z_1| \leq 1$ and $|z_2| \leq 1$, the local minimum is associated with Θ_C. On the other hand, in the set where $\det\Theta(z) \neq 0$ for $|z_1| \geq 1$ and $|z_2| \geq 1$, the local minimum is associated with Θ_{NC}. These findings show the *GCov* estimator's ability to capture matrices (Θ_C and Θ_{NC}) that produce the same autocovariance function as the underlying process. This remains true despite the estimator taking into account both linear and nonlinear transformations of the autocovariance of the error term. Consequently, local minima corresponding to these matrices arise.

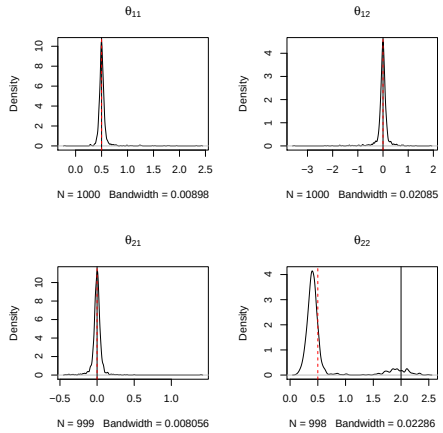
In conclusion, if the optimization problem starts within a set where a local minimum occurs, it is likely that the numerical optimization algorithm will get trapped in that set and converge to the local minimum instead of the global one.

Figure 1: Univarite *Gcov*22's objective function



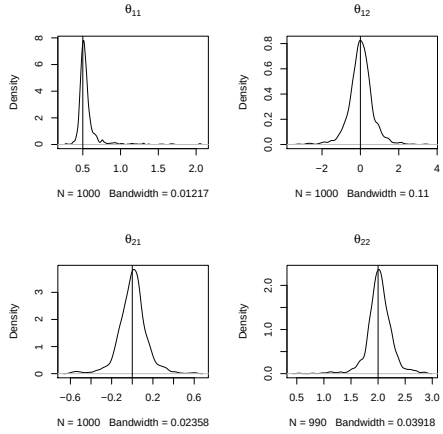
*The Gcov*22*'s objective function is shown when expressed as a function of* θ_{11} *and* θ_{22} *while holding the other elements of the same matrix constant and equal to the population matrix* Θ_0*, expressed as in (15). The black, red, and green dashed vertical lines represent the coefficients of matrices* Θ_0*,* Θ_C*, and* Θ_{NC}*, respectively. Case 1 is considered.*

Figure 2: Empirical density function of Θ with Θ_C as starting point



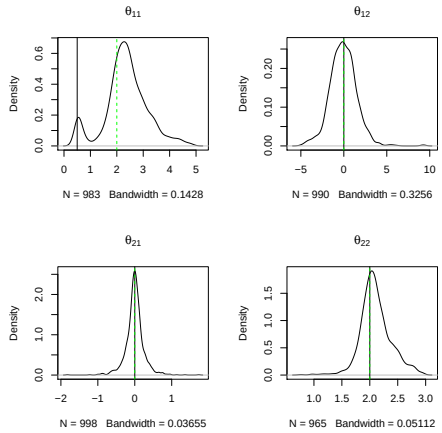
Empirical density function of Θ obtained when Θ_C is selected as the starting point, and the population matrix Θ_0 is expressed as in (15). The black vertical lines in the figure represent the true values of the matrix Θ_0, while the red dashed lines correspond to the elements of the matrix Θ_C. Case 1 is considered.

Figure 3: Empirical density function of Θ with Θ_0 as starting point



Empirical density function of Θ when the population matrix Θ_0, expressed as in (15), is considered. The initial value in the optimization problem is also set to Θ_0. The black vertical lines in the figure represent the true values of the matrix Θ_0.

Figure 4: Empirical density function of Θ with Θ_{NC} as starting point



Empirical density function of Θ obtained when Θ_{NC} is selected as the starting point, and the population matrix Θ_0 is expressed as in (15). The black vertical lines in the figure represent the true values of the matrix Θ_0, while the green dashed lines correspond to the elements of the matrix Θ_{NC}. Case 1 is considered.

4.1.2. ***The Absence of Bimodality Issue***

We proceed to investigate whether any local minima vanish when the eigenvalues of the population matrix for the causal and noncausal components are slightly smaller and larger than 1, respectively, while assuming cross-sectionally independent errors. As Figure 1 suggests, this scenario may reduce the global and local minima gap, prompting us to analyze if any local minima disappear. To achieve this goal, we will focus on the following population matrix:

$$\Theta_0^* = \begin{bmatrix} \theta_{11}^* & \theta_{12}^* \\ \theta_{21}^* & \theta_{22}^*, \end{bmatrix} = \begin{bmatrix} 0.85 & 0 \\ 0 & 1.2 \end{bmatrix}. \tag{18}$$

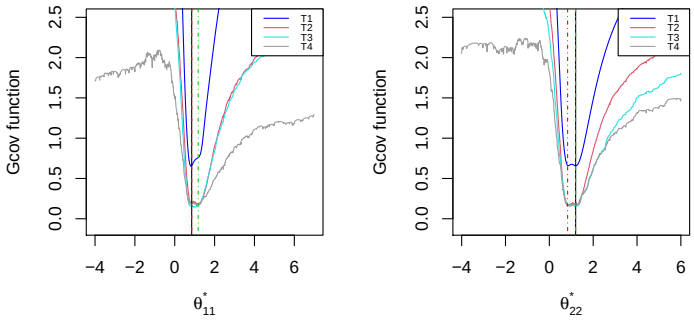
Even in this case, we derive the one-dimensional *Gcov*22's objective function expressed as a function of θ_{11}^* and θ_{22}^*, respectively, while holding the other elements of the same matrix constant and equal to Θ_0^* (Figure 5). Finally, we conduct a new Monte Carlo experiment with the same inputs as the previous simulation studies and compute the empirical density function of Θ^*, when its causal counterpart (Θ_C^*) is considered as the starting point for the optimization problem:

$$\Theta_C^* = \begin{bmatrix} 0.85 & 0 \\ 0 & (1.2)^{-1} \end{bmatrix}$$

Figures 5 and 6 show the results. Furthermore, Table 1 displays the frequency with which the correct model is identified in the Monte Carlo experiment.[2]

Our findings suggest that under this new DGP, the distance between the global and local minimums is not significant enough to give rise to the convergence of the algorithm to local minima. In other words, the reduced distance between the global and local minima facilitates the numerical algorithm in identifying the global minimum, even if initiated within a set where a local minimum occurs. This is true regardless of the nonlinear transformation a_k employed.

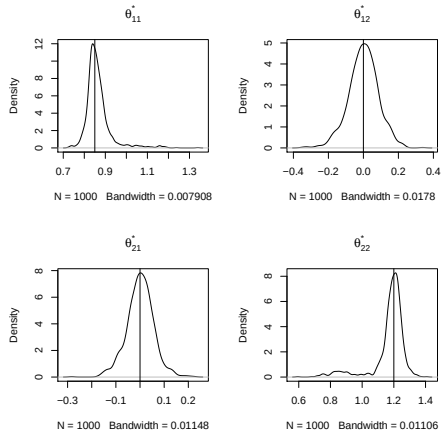
Figure 5: Univarite *Gcov*22's objective function



*The Gcov*22*'s objective function is shown when expressed as a function of* θ^*_{11} *and* θ^*_{22} *while holding the other elements of the same matrix constant and equal to the population matrix* Θ^*_0*, expressed as in (18). The black, red, and green dashed vertical lines represent the coefficients of matrices* Θ^*_0*,* Θ^*_C*, and* Θ^*_{NC}*, respectively. Case 1 is considered.*

[2]In the alternative case where the noncausal counterpart of Θ^*_0 is selected as the initial point for the optimization problem, similar results are obtained and are available upon request.

Figure 6: Empirical density function of Θ^* with Θ_C^* as starting point



Empirical density function of Θ^* obtained when Θ_C^* is selected as the starting point, and the population matrix Θ_0^* is expressed as in (18). The black vertical lines in the figure represent the true values of the matrix Θ_0^*, while the red dashed lines correspond to the elements of the matrix Θ_C^*. Case 1 is considered.

Table 1: Dynamic captured by *Gcov*22; Case 1

PM	Initial value	$VAR(2,0,1)$	$VAR(1,1,1)$	$VAR(0,2,1)$
Θ_0	Θ_C	87.6%	12%	0.4%
Θ_0	Θ_0	0.3%	99.3%	0.4%
Θ_0	Θ_{NC}	0%	9,8%	90.2%
Θ_0^*	Θ_C^*	8.8%	78.2%	13.0%

*The table illustrates the performance of Gcov*22 *in estimating the dynamic of a 2-dimensional VAR(1,1,1). 'PM' denotes the population matrix. Case 1 is considered.*

4.2. ***Case 2***

4.2.1. ***Presence of Bimodality Issue***

We now relax the assumption of independence between $y_{1,t}$ and $y_{2,t}$, and maintain the assumption of cross-sectional independent error term. Specifically, we consider the following non-diagonal autoregressive population matrix:

$$\Theta_0 = AJA^{-1} = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & -1 \end{bmatrix} \begin{bmatrix} 0.4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 1.17 & -0.58 \\ -1.10 & 1.23. \end{bmatrix} \quad (19)$$

This relaxation allows us now to leverage the second-order moment of the process to differentiate the true mixed process from its causal and noncausal counterpart (see Gourieroux and Jasiak (2017)).

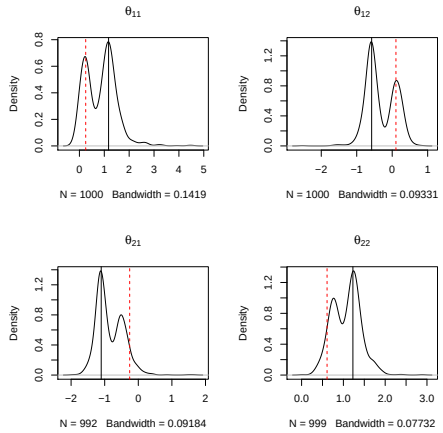
In contrast to Section 4.1, analyzing the objective function in terms of θ_{11} and θ_{22} does not provide information about the univariate case due to the lack of independence between y_1 and y_2. Therefore, we directly perform three new Monte Carlo experiments and compute the empirical density functions of Θ for different initial points in the optimization problem. We set $T = 1000$ observations and $N = 1000$ replications, discard the first and last 10% of observations in each simulated time series, and set $H = 10$ in (13). Finally, we consider the nonlinear transformation of the error term T1. The results are presented in Figures 7-9 and Table 2.

Figure 7 displays the results obtained using the OLS estimate of the process Y as the initial point. It has been shown in Gourieroux and Jasiak (2017) that the OLS estimator is not capable of capturing the noncausal component of the process, and it estimates a matrix with eigenvalues j_1 and j_2^{-1}. This is because the OLS estimator assumes Gaussian errors, in which case the causal and noncausal dynamics cannot be distinguished, and the model is not identifiable. Therefore, the OLS estimate of Θ (Θ_{OLS}) provides the causal counterpart of the matrix (19). Figure 8 shows the density function of $\hat{\Theta}$ when the noncausal counterpart of Θ is selected as the initial point. To obtain this matrix, it is necessary to estimate a VMAR(0,1) process by OLS and invert the estimated matrix. We denote this matrix as $\widetilde{\Theta}$, which is characterized by j_1^{-1} and j_2 as eigenvalues. Finally, Figure 9 presents the results when the initial value is Θ_{MIX}, a randomly selected matrix that lies in the same set as the global minimum (one root inside and one root outside the unit circle):

$$\begin{aligned}\Theta_{MIX} = AJA^{-1} &= \begin{bmatrix}-0.8 & 0.75\\ 0.46 & -1\end{bmatrix}\begin{bmatrix}0.8 & 0\\ 0 & 1.2\end{bmatrix}\begin{bmatrix}-0.8 & 0.75\\ 0.46 & -1\end{bmatrix}^{-1}\\ &= \begin{bmatrix}0.50 & -0.53\\ 0.40 & 1.50.\end{bmatrix}\end{aligned} \tag{20}$$

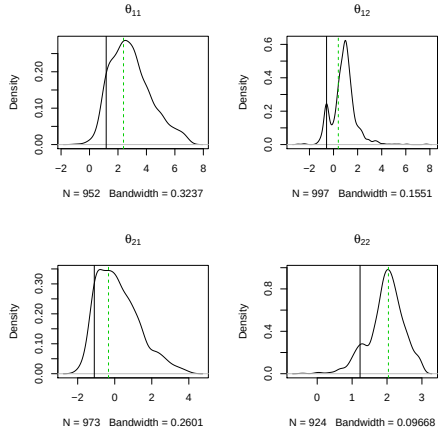
The results reveal that when the optimization process starts at the parameter values corresponding to the reciprocals of the true eigenvalues of the autoregressive matrix, the maximization algorithm becomes stuck at a local minimum. Consequently, in practical applications, identification issues may emerge, and local minima manifest at parameter values associated with incorrect orders. Therefore, depending on the initial starting point, the GCov optimization algorithm might converge towards a local minimum. Indeed, similar to Case 1, the results show that when the optimization is initiated within the domain set of $Gcov22$ where $det\Theta(z) \neq 0$ for $|z_1| \leq 1$ and $|z_2| \leq 1$, the algorithm converges towards a local minimum associated with the matrix having eigenvalues j_1 and j_2^{-1} (Θ_{OLS}). In contrast, when the numerical optimization algorithm is initialized in the set where $det\Theta(z) \neq 0$ for $|z_1| \geq 1$ and $|z_2| \geq 1$, the local minimum is associated with a matrix having eigenvalues j_1^{-1} and j_2 ($\widetilde{\Theta}$). Finally, the elimination of convergence to local minima is observed when Θ_{MIX} is chosen as the initial point. This finding emphasizes the importance of selecting a matrix that belongs to the same set as the population matrix. Table 2 further confirms these results.

Figure 7: Empirical density function of Θ with Θ_{OLS} as starting point



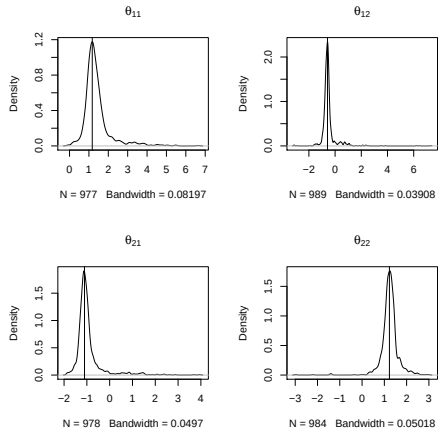
Empirical density function of Θ obtained when Θ_C is selected as the starting point, and the population matrix Θ_0 is expressed as in (19). The black vertical lines in the figure represent the true values of the matrix Θ_0, while the red dashed lines correspond to the elements of the matrix Θ_{OLS}. Case 2 is considered.

Figure 8: Empirical density function of Θ with $\widetilde{\Theta}$ as starting point



Empirical density function of Θ obtained when $\widetilde{\Theta}$ is selected as the starting point, and the population matrix Θ_0 is expressed as in (19). The black vertical lines in the figure represent the true values of the matrix Θ_0, while the green dashed lines correspond to the elements of the matrix $\widetilde{\Theta}$. Case 2 is considered.

Figure 9: Empirical density function of Θ with Θ_{MIX} as starting point



Empirical density function of Θ obtained when Θ_{MIX} is selected as the starting point, and the population matrix Θ_0 is expressed as in (19). The black vertical lines in the figure represent the true values of the matrix Θ_0.

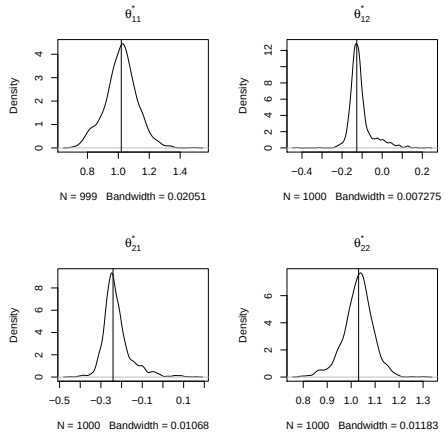
4.2.2. ***The Absence of Bimodality Issue***

We proceed to examine whether local minima vanish when the causal and noncausal eigenvalues are marginally smaller and larger than one, respectively, even when the processes $y_{1,t}$ and $y_{2,t}$ are not independent. Furthermore, we maintain the assumption of cross-sectional independent errors. To this end, we consider the following autoregressive matrix:

$$\Theta_0^* = AJA^{-1} = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & -1 \end{bmatrix} \begin{bmatrix} 0.85 & 0 \\ 0 & 1.2 \end{bmatrix} \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 1.02 & -0.13 \\ -0.24 & 1.03, \end{bmatrix} \quad (21)$$

and implement a Monte Carlo simulation study considering the OLS estimate of Θ^* (Θ^*_{OLS}) as the initial value (see Figure 10 and Table 2). The results highlight that the density function of $\hat{\Theta}^*$ is well-shaped around the true population matrix. In the alternative scenario where a matrix with both eigenvalues outside the unit circle is selected as the starting point, similar results are obtained and available upon request. Therefore, as also shown in Section 4.1, we conclude that in some cases, the distance between the global and local minima is insufficient to give rise to convergence to local minima. In other words, the small gap between the global and local minima makes it easier for the numerical algorithm to locate the global minimum, even if initiated within a set where a local minimum occurs.

Figure 10: Empirical density function of Θ^* with Θ^*_{OLS} as starting point



Empirical density function of Θ^* *obtained when* Θ^*_{OLS} *is selected as the starting point, and the population matrix* Θ^*_0 *is expressed as in (21). The black vertical lines in the figure represent the true values of the matrix* Θ^*_0*. Case 2 is considered.*

Table 2: Dynamic captured by $Gcov22$; Case 2

PM	Initial value	$VAR(2,0,1)$	$VAR(1,1,1)$	$VAR(0,2,1)$
Θ_0	Θ_{OLS}	34.2%	65.4%	0.4%
Θ_0	$\widetilde{\Theta}$	0.2%	15.3%	84.5%
Θ_0	Θ_{MIX}	0.1%	91.7%	8.2%
Θ^*_0	Θ^*_{OLS}	6.6%	90.2%	3.2%

The table illustrates the performance of $Gcov22$ *in estimating the dynamic of a 2-dimensional VAR(1,1,1). 'PM' denotes the population matrix. Case 2 is considered.*

5. Simulated Annealing algorithm

In the previous section, we stressed the importance of selecting an initial point that belongs to the same set as the global minimum. Specifically, selecting an initial matrix with the same n_1 and n_2 as the population matrix is crucial for achieving successful convergence of the BFGS optimization algorithm. However, in practical applications, determining the number of roots that lie within and outside the unit circle of the population matrix can be challenging. Therefore, in this section, we use SA (Kirkpatrick et al. (1983), Černỳ (1985), and Goffe et al. (1994)) to investigate whether it can find an initial point in the same set where the global minimum lies. If successful, this approach would allow the BFGS optimization algorithm to converge successfully.

SA is an optimization method that was inspired by the annealing process used in metallurgy. In metallurgy, materials are cooled gradually to eliminate any imperfections and achieve a more stable state. The algorithm starts at a high temperature (T^o) and gradually cools down over time to reduce the likelihood of getting stuck in a local minimum. Hence, 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, enabling 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 et al. (1987), Goffe et al. (1992), Goffe et al. (1994), and Goffe (1996)).

Let us now explain how the SA algorithm works when applied to the $GCov$ estimator. We consider a VAR($n_1, n_2, 1$) as (7), and we use f to denote the objective function of either $GCov17$ or $GCov22$. It is not necessary to specify which $GCov$ is being employed for the purposes of illustrating how the algorithm works. SA aims to select the matrix Θ close to the global optimum. We assume that the (i,j)-th element of matrix Θ (denoted as $\theta_{i,j}$, for $i,j = 1 \dots n$) lies within the interval $[\theta_{MIN}, \theta_{MAX}]$. This step helps us to restrict the domain, $\mathbb{R}^{n \times n}$, of our objective function and hence narrow the search area. Additionally, we denote the maximum and minimum temperature values of T^o as T^o_{MAX} and T^o_{MIN}, respectively.

To begin the optimization process, a random initial state of Θ, denoted as Θ^S, is selected, and its corresponding function value, $f(\Theta^S)$, is computed. Specifically, each element of the matrix Θ^S is randomly chosen from θ_{MIN} to θ_{MAX}. Subsequently, a new value of Θ, represented by Θ', is generated. We implement the following computation to derive the ij-th element of matrix Θ':

$$\theta'_{ij} = \theta^S_{ij} + m_{ij} \quad \forall i,j = 1, \dots, n, \tag{22}$$

where m_{ij} is the ij-th element of an $(n \times n)$ matrix M and is chosen randomly from the interval $[-\theta^S_{ij} - \theta_{MIN}, -\theta^S_{ij} + \theta_{MAX}]$. We choose this interval to ensure that θ'_{ij} remains within the range $[\theta_{MIN}, \theta_{MAX}]$.

The function $f(\Theta')$ is then computed and compared with $f(\Theta^S)$. If $f(\Theta') < f(\Theta^S)$, Θ' is accepted and the algorithm progresses downhill. In the opposite case, $f(\Theta') > f(\Theta^S)$, the eventual acceptance of Θ' is based on the Metropolis criterion. According to this criterion, we compute the variable p^o such that

$$p^o = e^{-\frac{(f(\Theta') - f(\Theta^S))}{T^o}}, \tag{23}$$

to compare it with p^*, a number selected randomly from 0 to 1. If $p^o < p^*$, Θ' is rejected, and the algorithm remains on the same point of the function. On the other hand, if $p^o > p^{o,\prime}$, we accept Θ' and move uphill. The acceptance probability in simulated annealing is determined by Equation (23), which is controlled by the parameter T^o. 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}.

The SA method presents two main drawbacks. First, it requires exploring a large number of candidate solutions, especially for large values of n and p. To mitigate this challenge, setting boundaries, θ_{MIN} and θ_{MAX}, to restrict the search area can be helpful. By limiting the number of candidate solutions that need to be evaluated, we can reduce the computational time required to find the global minimum using the SA method. Secondly, 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 contingent upon the objective function to be minimized. Goffe et al. (1994) provides some tips on how to set these parameters, but, in general, their choice requires experimentation.

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 cases where a different result is obtained, it may be necessary to modify one or more of the parameters involved in the simulated annealing algorithm.

We now conduct a Monte Carlo experiment to evaluate the effectiveness of SA on the *GCov* estimator. The experiment uses the same DGP described in Section 4, which is known to produce local minima. Specifically, we employ the population matrix Θ_0 displayed in (19). As Σ is diagonal in this case, we obtained similar results when applying SA to either $Gcov22$ or $Gcov17$. However, to make a more accurate comparison with the findings of the previous section, we only present the results obtained by $Gcov22$, while the results from $Gcov17$ are available upon request.

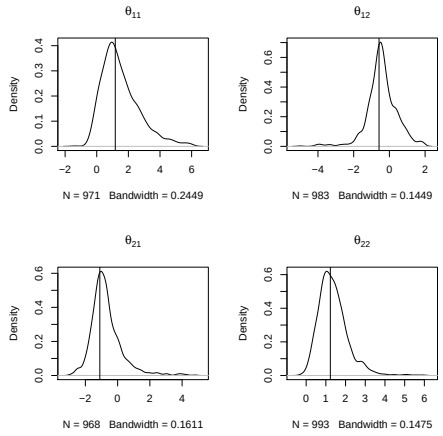
Given the time needed to run the $Gcov22$ estimator with SA, we reduced the number of observations to $T = 250$ observations. We keep $N = 1000$ replications. The initial temperature T^o is set to $T^o_{MAX} = 1600$, and we let it decreases at a rate of $r = 0.85$ for $Q = 200$ iterations until it reaches T^0_{MIN}.[3] In order to efficiently explore the search area, our experiments suggest setting $M = 1000$ for each T^o. Setting M too high may result in inefficiencies in terms of time. Conversely, setting M too low may restrict the ability of the algorithm to explore the search area thoroughly. It's worth noting that in some replications of our experiment, the objective function may require higher values of either Q, M, or both. However, for practical reasons, we set the values of Q and M to be the same for all replications. We set θ_{MIN} and θ_{MAX} to -3.5 and 3.5, respectively. As previously mentioned, these parameters are typically problem-dependent, and their selection requires experimentation. The matrix Θ obtained through the SA technique is then used as the initial point for our optimization problem, which is then solved using the $BFGS$ algorithm. The aim of the Monte Carlo experiment is to investigate which strategy for capturing the initial values, namely SA, Θ_{OLS}, and $\widetilde{\Theta}$ gets us closer to the global minimum. Figures

[3] The final temperature, T^o_{MIN}, is a function of T^o_{MAX}, r, and Q since it is obtained after $Q - th$ reductions, at rate r, from T^o_{MAX}. It is defined as $T^o_{MIN} = T^0_{MIN}(T^0_{MAX}, r, Q)$.

11 and Table 3 show the results.

The SA algorithm significantly improves the results compared to the previous section despite the fact that the sample size considered in this section is significantly smaller. Bimodality issues disappear, and the $Gcov22$ estimator accurately identifies the process dynamics with a success rate of approximately 78 percent. In the previous section, when the same DGP is considered, the correct dynamics were identified with success rates of 65.4% and 15.3% when using Θ_{OLS} and $\widetilde{\Theta}$ as initial values, respectively, and $T = 1000$ observations.

Figure 11: Empirical density function of Θ with SA



The empirical density function of $\hat{\Theta}$ is obtained using $Gcov22$ when the SA strategy is implemented to select the starting values. We assume that Θ_0, as defined in (19), represents the population matrix. The vertical lines in the plot indicate the corresponding population values.

Table 3: Dynamic captured by $Gcov22$

PM	Initial value	$VAR(2,0,1)$	$VAR(1,1,1)$	$VAR(0,2,1)$
Θ_0	SA	8.8%	78.2%	13.0%

Estimated dynamics captured by the estimator $Gcov22$ when the SA is implemented to capture the initial values. Θ_0, as expressed in (19), is selected as the population matrix (PM). Case 2 is considered.

6. Empirical investigations

This section describes two empirical investigations of bivariate processes. The first investigation examines the wheat and corn futures from October 18, 2016, to March 29, 2018. The second one examines soybean and wheat futures

using the same time period as the previous process.[4] The demeaned data is presented in Figures 12-13. The primary objective of this study is to evaluate the performance of the $Gcov22$ estimator when applied to real-world data and to investigate whether any eventual local minima arise.

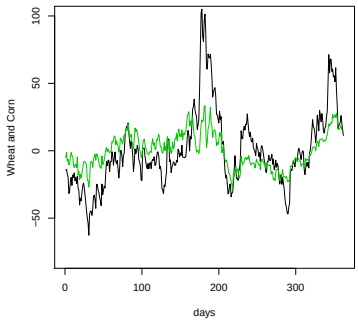
The series does not exhibit global trends or other widespread and persistent explosive patterns; instead, they display localized trends, bubbles, and spikes. Therefore, we assume that the series is strictly stationary.

Since the standard normality tests reject the null hypothesis of normality in all the considered time series, we can apply $Gcov22$ to our demeaned data. In this regard, we use both the OLS estimate (Θ_{OLS}) and the results obtained by the SA method as starting points for the optimization problems. Table 4 shows the results. $Gcov22$ identifies the process related to wheat and corn as mixed causal and noncausal (VAR$(1,1,1)$). In particular, the same estimated matrix is found regardless of the strategy employed to capture the initial guess, and eigenvalues $\hat{j}_1 = 0.94$ and $\hat{j}_2 = 1.12$ are found. This result confirms the findings of Section 4: there are no local minima when the causal and noncausal eigenvalues are respectively slightly smaller and larger than 1.

When analyzing the bivariate process of soybean and wheat, we find that the results depend on the starting point chosen for the optimization problem. In particular, when the OLS strategy is used to capture the initial value, $Gcov22$ identifies the process as purely causal (VAR(4,0,2)). On the other hand, when the SA method is used, a lower function value is achieved, and the bivariate process is identified as VAR(3,1,2), with eigenvalues $j_1 = 0.972$, $j_2 = 0.88$, $j_3 = 0.604$, and $j_4 = -4.355$. Even in this case, the empirical investigation confirms our previous findings. Specifically, local minima arise when either the causal or the noncausal or both eigenvalues are significantly smaller and larger than 1.

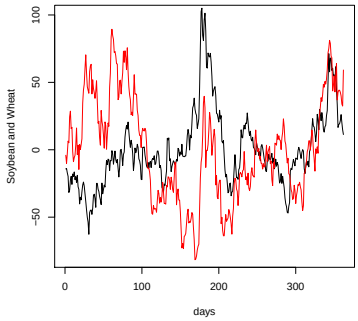
[4]Data obtained from https://ca.finance.yahoo.com (wheat futures: ticker ZW=F, corn futures: ticker ZC=F, and soybean futures: ticker ZS=F)

Figure 12: Bivariate process of wheat and corn futures



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

Figure 13: Bivariate process of wheat and soybean futures



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

Table 4: Estimated coefficients of our empirical analysis

Wheat and Corn							
SV		$\hat{\Phi}_1$		$\hat{\Phi}_2$			
		$\phi_{j,1}$	$\phi_{j,2}$	$\phi_{j,1}$	$\phi_{j,2}$	f.v.	Model
Θ_{OLS}	$\phi_{1,j}$	0.84	0.30	/	/	3.09	VAR(1,1,1)
	$\phi_{2,j}$	-0.09	1.22	/	/		
SA	$\phi_{1,j}$	0.84	0.30	/	/	3.09	VAR(1,1,1)
	$\phi_{2,j}$	-0.09	1.22	/	/		
Soybean and Wheat							
SV		$\hat{\Phi}_1$		$\hat{\Phi}_2$			
		$\phi_{j,1}$	$\phi_{j,2}$	$\phi_{j,1}$	$\phi_{j,2}$	f.v.	Model
Θ_{OLS}	$\phi_{1,j}$	1.16	0.24	-0.29	-0.29	2.00	VAR(4,0,2)
	$\phi_{2,j}$	0.06	1.05	-0.04	-0.09		
SA	$\phi_{1,j}$	0.44	1.23	0.52	-1.21	1.50	VAR(3,1,2)
	$\phi_{2,j}$	3.31	-2.34	-3.19	3.10		

The column "SV" *and* "f.v." *display the strategy adopted to capture the starting values and the value of the function at the estimated values.*

7. Conclusions

In this paper, we have investigated the performance of the *GCov* in estimating mixed causal and noncausal models. The *GCov* estimator, being a semi-parametric method, offers the advantage of not assuming any specific error distribution. By utilizing a portmanteau-type criterion based on nonlinear autocovariances, it ensures consistent estimates and consequently allows 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 *GCov* estimator. When these autocovariances are insufficient or inadequate, or when the error density closely resembles the Gaussian distribution, identification issues can arise. This manifests in the presence of local minima in the objective function, occurring at parameter values associated with incorrect causal and noncausal orders. Consequently, the optimization algorithm may converge to a local minimum, leading to inaccurate estimates.

To overcome the problem of local minima and improve the estimation accuracy of mixed VAR models, 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 in a more robust and flexible manner, SA provides a reliable solution for obtaining more accurate estimates of the causal and noncausal orders.

Finally, the paper applies the *Gcov*22 method to two bivariate commodity price series and assesses its performance by employing various initial points for the optimization problem. The results highlight the existence of local minima

in one of the bivariate processes, emphasizing the importance of carefully selecting suitable initial values to obtain accurate estimates when utilizing the $GCov$ estimator.

References

Alessi, L., Barigozzi, M., Capasso, M., 2008. A review of nonfundamentalness and identification in structural var models .

Bec, F., Nielsen, H.B., Saïdi, S., 2020. Mixed causal–noncausal autoregressions: Bimodality issues in estimation and unit root testing 1. Oxford Bulletin of Economics and Statistics 82, 1413–1428.

Blanchard, O.J., Quah, D., 1988. The dynamic effects of aggregate demand and supply disturbances.

Breid, F.J., Davis, R.A., Lh, K.S., Rosenblatt, M., 1991. Maximum likelihood estimation for noncausal autoregressive processes. Journal of Multivariate Analysis 36, 175–198.

Breidt, F.J., Davis, R.A., Lh, K.S., Rosenblatt, M., 1991. Maximum likelihood estimation for noncausal autoregressive processes. Journal of Multivariate Analysis 36, 175–198.

Carnevali, P., Coletti, L., Patarnello, S., 1987. Image processing by simulated annealing, in: Readings in Computer Vision. Elsevier, pp. 551–561.

Černỳ, V., 1985. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. Journal of optimization theory and applications 45, 41–51.

Chan, K.S., Ho, L.H., Tong, H., 2006. A note on time-reversibility of multivariate linear processes. Biometrika 93, 221–227.

Corana, A., Marchesi, M., Martini, C., Ridella, S., 1987. Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm—corrigenda for this article is available here. ACM Transactions on Mathematical Software (TOMS) 13, 262–280.

Cubadda, G., Hecq, A., 2011. Testing for common autocorrelation in data-rich environments. Journal of Forecasting 30, 325–335.

Cubadda, G., Hecq, A., Telg, S., 2019. Detecting co-movements in non-causal time series. Oxford Bulletin of Economics and Statistics 81, 697–715.

Davis, R.A., Song, L., 2020. Noncausal vector ar processes with application to economic time series. Journal of Econometrics 216, 246–267.

Fernández-Villaverde, J., Rubio-Ramírez, J.F., Sargent, T.J., Watson, M.W., 2007. Abcs (and ds) of understanding vars. American economic review 97, 1021–1026.

Giancaterini, F., Hecq, A., 2022. Inference in mixed causal and noncausal models with generalized student's t-distributions. Econometrics and Statistics .

Goffe, W.L., 1996. Simann: A global optimization algorithm using simulated annealing. Studies in Nonlinear Dynamics & Econometrics 1.

Goffe, W.L., Ferrier, G.D., Rogers, J., 1992. Simulated annealing: An initial application in econometrics. Computer Science in Economics and Management 5, 133–146.

Goffe, W.L., Ferrier, G.D., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. Journal of econometrics 60, 65–99.

Gourieroux, C., Jasiak, J., 2016. Filtering, prediction and simulation methods for noncausal processes. Journal of Time Series Analysis 37, 405–430.

Gourieroux, C., Jasiak, J., 2017. Noncausal vector autoregressive process: Representation, identification and semi-parametric estimation. Journal of Econometrics 200, 118–134.

Gourieroux, C., Jasiak, J., 2018. Misspecification of noncausal order in autoregressive processes. Journal of Econometrics 205, 226–248.

Gourieroux, C., Jasiak, J., 2022a. Generalized covariance estimator. Journal of Business & Economic Statistics , 1–31.

Gourieroux, C., Jasiak, J., 2022b. Nonlinear forecasts and impulse responses for causal-noncausal (s) var models. arXiv preprint arXiv:2205.09922 .

Hecq, A., Lieb, L., Telg, S., 2016. Identification of mixed causal-noncausal models in finite samples. Annals of Economics and Statistics/Annales d'Économie et de Statistique , 307–331.

Hecq, A., Velasquez-Gaviria, D., 2022. Spectral estimation for mixed causal-noncausal autoregressive models. arXiv preprint arXiv:2211.13830 .

Hecq, A., Voisin, E., 2021. Forecasting bubbles with mixed causal-noncausal autoregressive models. Econometrics and Statistics 20, 29–45.

Jones, R.O., 1991. Molecular structures from density functional calculations with simulated annealing. Angewandte Chemie International Edition in English 30, 630–640.

Kirkpatrick, S., Gelatt Jr, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. science 220, 671–680.

Lanne, M., Saikkonen, P., 2011. Noncausal autoregressions for economic time series. Journal of Time Series Econometrics 3.

Lanne, M., Saikkonen, P., 2013. Noncausal vector autoregression. Econometric Theory 29, 447–481.

Lippi, M., Reichlin, L., 1993. The dynamic effects of aggregate demand and supply disturbances: Comment. The American Economic Review 83, 644–652.

Pannetier, J., Bassas-Alsina, J., Rodriguez-Carvajal, J., Caignaert, V., 1990. Prediction of crystal structures from crystal chemistry rules by simulated annealing. Nature 346, 343–345.

Sims, C.A., 1980. Macroeconomics and reality. Econometrica: journal of the Econometric Society , 1–48.

Wong, D., Leong, H.W., Liu, H., 2012. Simulated annealing for VLSI design. volume 42. Springer Science & Business Media.