ESSAYS IN ESTIMATION AND HYPOTHESIS TESTING IN NON-GAUSSIAN NONLINEAR TIME SERIES MODELS

Aryan Manafi Neyazi

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
GRADUATE PROGRAM IN ECONOMICS
YORK UNIVERSITY
TORONTO, ONTARIO

May 2026

Abstract

This dissertation develops new methods for estimation and hypothesis testing in strictly stationary non-Gaussian nonlinear time series models. Many economic and financial time series exhibit heavy tails, nonlinear dynamics, conditional heteroskedasticity, and locally explosive behavior that cannot be adequately captured by classical Gaussian linear frameworks. In such settings, linear autocovariances are insufficient to characterize dependence, and conventional diagnostic tools may fail to detect important dynamic features. This thesis proposes a unified approach based on nonlinear autocovariances within the Generalized Covariance (GCov) estimator and test framework to address these challenges.

The first essay introduces a test of the absence of linear and nonlinear serial dependence (NLSD) for univariate and multivariate non-Gaussian processes. The proposed portmanteau-type statistic is constructed from nonlinear transformations of the data and follows an asymptotic chi-square distribution under the null hypothesis of serial independence. Building on this idea, the dissertation develops the GCov specification test for semi-parametric dynamic models with independent and identically distributed non-Gaussian errors. The asymptotic distribution of the test statistic is derived under the null and under local alternatives, providing insight into its power properties and practical implementation.

The second essay addresses high-dimensional settings where the number of variables or nonlinear transformations is large relative to the sample size. A Regularized Generalized Covariance (RGCov) estimator is proposed, incorporating shrinkage techniques to stabilize covariance matrix estimation. The asymptotic properties of the estimator are established,

and Monte Carlo simulations demonstrate improved finite-sample performance in terms of bias, variance, and mean squared error.

The third essay extends shrinkage ideas to hypothesis testing by developing a regularized test for the absence of nonlinear serial dependence. The proposed procedure improves size control in high-dimensional environments while preserving strong power properties.

The theoretical contributions are complemented by simulation studies and empirical applications to financial and commodity price data, including models with mixed causal-noncausal dynamics. Overall, the dissertation advances the econometric analysis of nonlinear and non-Gaussian time series by providing robust, theoretically grounded, and practically applicable tools for estimation and diagnostic testing.

Acknowledgments

The completion of this dissertation marks the end of a long and rewarding journey, one that would not have been possible without the support, guidance, and encouragement of many people to whom I am deeply grateful.

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Joann Jasiak. Her guidance, mentorship, and unwavering support have been instrumental throughout my doctoral studies. I am grateful for the trust she placed in me, the opportunities she provided, and the countless discussions that shaped my development as a researcher. Her dedication to academic excellence and passion for research have been a constant source of inspiration. It has been a privilege to work under her supervision, and I will always be thankful for her mentorship, encouragement, and belief in my potential.

I am deeply thankful to the members of my dissertation committee for their time, expertise, and thoughtful feedback. In particular, I would like to thank Professor Antoine Djogbenou, whose advice, encouragement, and willingness to discuss research and academic matters made my Ph.D. journey considerably easier. His guidance helped me navigate many challenges and contributed greatly to my development as a researcher.

I would also like to thank the faculty members and students at Maastricht University for making my research visit a memorable and enriching experience. In particular, I am deeply grateful to Professor Alain Hecq, who welcomed me as a visiting researcher and served as my host supervisor. His generosity, support, and valuable feedback greatly contributed to my research and broadened my perspective on econometric theory and time series analysis.

I am also grateful to Professor Christian Gouriéroux for the insightful discussions and valuable suggestions that helped shape several aspects of my research. His pioneering work has been a constant source of inspiration throughout my doctoral studies.

My sincere gratitude also goes to the faculty members and staff of the Department of Economics at York University for creating a supportive and stimulating academic environment throughout my doctoral studies. I am particularly grateful to the chair of the department, Professor Wai-Ming Ho, and the Graduate Program Directors, Professor Ahmed Akyol, Professor Nippe Lagerlöf, and Professor Selçuk Özyurt, for their guidance and support at various stages of my Ph.D. journey. A special thank you goes to Professor Matias Cortes. Although he was not a member of my dissertation committee, he followed my progress closely throughout my doctoral studies.

I would also like to thank Regina Pinto and Sara Bang for their dedication and assistance in helping students navigate the many administrative aspects of graduate studies. Their support was invaluable throughout my time at York.

My sincere thanks go to Francesco Giancaterini, my co-author and friend. Our collaboration has been one of the most rewarding aspects of my academic journey. I am grateful for the many discussions, ideas, and shared experiences that contributed not only to our research but also to my personal and professional growth.

A special thank you goes to my fellow Ph.D. students at York University. The doctoral journey would have been far more difficult without their friendship, support, and companionship. In particular, I would like to thank Thinh Dao, Zahra Shahidi, Shadi Rahimfallah, Sukran Dinc Kilaguz, Yunus Emre Bozkurt, Romina Amirimehr, Emre Inan, and Cheng Zhong for the many conversations, encouragement, and memorable moments we shared over the years. The friendships I formed during this journey are among the most valuable things I take away from my time at York.

I am also grateful to my friends, who have been a constant source of support throughout this journey. In particular, I would like to thank Amin Saedi, Reza Shojaei, Raziye Falasiri, and Milad Ahmadi for their friendship, encouragement, and unwavering support during both the challenging and rewarding moments of my Ph.D.

Finally, and most importantly, I would like to thank my family.

To my mother and father, words cannot adequately express my gratitude. Everything I have achieved is built upon your sacrifices, unconditional love, and unwavering support. You believed in me even when I doubted myself and encouraged me to pursue opportunities that often took me far from home. Every milestone in this journey has been possible because of your faith in me, your patience, and the countless sacrifices you made along the way. This accomplishment belongs as much to you as it does to me.

To my grandmother, thank you for your unconditional love and care. No matter where I was in the world, I always carried with me the comfort of your love, prayers, and support. Your kindness and warmth have been a constant source of strength throughout my life.

To my brother, Arash, and my sister-in-law, Hasti, thank you for your encouragement, support, and confidence in me. Knowing that you were always there for me provided comfort and motivation during both the difficult and successful moments of this journey.

This dissertation is dedicated to all of you. Your love, support, sacrifices, and belief in me made this achievement possible, and for that, I will be forever grateful.

Contents

List of Tables

List of Figures

Introduction

Modern economic and financial time series are characterized by nonlinear dynamics, heavy-tailed distributions, structural asymmetries, and episodes of local explosiveness such as spikes and bubbles. These empirical features challenge the classical Gaussian linear time series paradigm that has long dominated econometric practice. While traditional ARMA and VAR frameworks remain useful benchmarks, they are often inadequate when the data exhibit non-Gaussianity, conditional heteroskedasticity, nonlinear serial dependence, or mixed causal-noncausal behavior. In such environments, estimation and diagnostic procedures built upon second-order linear dependence may fail to detect economically meaningful dynamics, and model misspecification can lead to misleading inference.

This dissertation develops new tools for estimation and hypothesis testing in strictly stationary but non-Gaussian and nonlinear time series models. The central theme is that nonlinear dependence structures cannot be fully characterized by standard autocovariances. Instead, higher-order and nonlinear transformations of the data contain essential information about the underlying dynamics. By systematically exploiting nonlinear autocovariances within a generalized covariance (GCov) framework, this thesis proposes unified methods for diagnostic checking, specification testing, and regularized estimation in both univariate and multivariate settings.

The first contribution introduces a new test of the absence of linear and nonlinear serial dependence (NLSD) for strictly stationary non-Gaussian processes. Classical portmanteau tests, such as Box-Pierce or Ljung-Box statistics, focus on linear autocorrelations and are primarily justified under Gaussian assumptions. However, for non-Gaussian processes, zero linear autocovariances do not imply independence. Nonlinear transformations of the series may reveal hidden dependence structures that remain undetected by conventional methods. Building on the Generalized Covariance methodology of Gourieroux and Jasiak (2023), the dissertation shows that the joint absence of linear and nonlinear serial dependence can be tested using a multivariate portmanteau statistic constructed from nonlinear autocovariance matrices. The resulting test has an asymptotic chi-square distribution under the null hypothesis and provides a convenient one-step alternative to multiple residual diagnostics.

Beyond proposing the NLSD test, the dissertation develops the GCov specification test for semi-parametric dynamic models with independent and identically distributed non-Gaussian errors. The GCov estimator is defined as the minimizer of a quadratic form in nonlinear autocovariances. When the model is correctly specified, this estimator is consistent, asymptotically normal, and semi-parametrically efficient for a suitable choice of transformations. Importantly, the associated specification test possesses a known asymptotic distribution, unlike several alternative dependence measures whose limiting laws are intractable and require bootstrap approximations. The thesis extends existing theoretical results by deriving the local asymptotic distribution of the GCov specification test under drifting alternatives. This analysis characterizes the non-centrality parameter and provides insight into the power properties of the procedure against economically relevant alternatives, including additional autoregressive lags or nonlinear variance components.

A second major contribution addresses the challenges posed by high dimensionality. In many contemporary applications—such as macroeconomic panels, financial portfolios, and sectoral indices—the number of variables or transformations can be large relative to the sample size. In such settings, the sample covariance matrix may be ill-conditioned or singular, undermining the stability of GCov estimation and hypothesis testing. To address this issue, the dissertation develops a Regularized Generalized Covariance (RGCov) estimator that incorporates shrinkage techniques inspired by Ledoit and Wolf. By shrinking the estimated covariance matrix toward a structured target, the RGCov estimator achieves improved numerical stability and reduced mean squared error in finite samples. The asymptotic properties of the regularized estimator are derived under both fixed and vanishing shrinkage sequences, and Monte Carlo experiments demonstrate substantial gains in bias-variance trade-offs when the dimension of the system or the number of nonlinear transformations increases.

The third contribution extends shrinkage ideas to hypothesis testing. When the number of nonlinear transformations grows large, the portmanteau statistic may suffer from size distortions due to estimation noise in high-dimensional covariance matrices. The dissertation proposes a shrinkage-regularized test for the absence of nonlinear serial dependence that stabilizes the test statistic while preserving its asymptotic properties. Simulation results indicate that the shrinkage-based procedure improves empirical size control without sacrificing power, especially in multivariate systems or when many transformations are employed.

The theoretical developments are complemented by extensive simulation studies and empirical applications. The empirical analysis illustrates how nonlinear autocovariances identify both causal and noncausal dynamics in mixed autoregressive models. Applications to commodity prices and renewable energy indices demonstrate that locally explosive patterns and nonlinear dependence structures can be effectively captured within the GCov framework. The results highlight the importance of robust diagnostic checking in models estimated under non-Gaussian assumptions and show that regularized GCov methods can enhance inference in high-dimensional environments.

Taken together, the essays in this dissertation contribute to the growing literature on nonlinear and non-Gaussian time series econometrics. They provide unified tools that bridge diagnostic testing and estimation, accommodate semi-parametric settings, and remain computationally feasible in high dimensions. By extending classical portmanteau concepts to nonlinear transformations and embedding them within a generalized covariance framework, the dissertation advances both the theoretical foundations and the practical implementation of modern time series analysis.

Chapter 1

GCov-Based Portmanteau Test

Coauthorship Statement

This chapter is based on joint work titled "GCov-Based Portmanteau Test" with Joann Jasiak. All authors contributed equally to the conceptualization, data preparation, empirical modeling, analysis, and manuscript writing.

This is an Original Manuscript of an article published in Econometric Reviews on 14 Apr 2026, available online: https://doi.org/10.1080/07474938.2026.2643749

1.1 Introduction

Diagnostic checking is an important component of the traditional Box-Jenkins procedure for the identification and estimation of stationary linear ARMA models with serially uncorrelated Gaussian errors. It has been routinely conducted by applying the Box-Pierce or Ljung-Box tests to the residuals of an ARMA model to detect serial correlation. Analogously, the diagnostic checking of nonlinear ARCH-type models is often conducted by applying the Box-Pierce or Ljung-Box test twice to the residuals and their squares. Recently, there has been a growing interest in dynamic models with independent, identically distributed (i.i.d.) non-Gaussian errors, such as the structural vector autoregressive (SVAR) models for macroeconomic data, and the mixed causal-noncausal autoregressive (MAR), vector autoregressive (VAR) and double autoregressive (DAR) models for processes with local explosive features, such as spikes and bubbles, displayed by the commodity and cryptocurrency price processes, for example. Diagnostic checking in these models is commonly based on the Ljung Box tests applied to the residuals, their squares, and potentially higher powers, which is a cumbersome multi-step procedure.

This paper reviews convenient one-step diagnostic procedures for jointly testing various forms of nonlinear serial dependence in non-Gaussian time series before estimation, as well as in the residuals ex-post, as a specification test for dynamic models with i.i.d. non-Gaussian errors.

We introduce a new (non)linear serial dependence (NLSD) test for strictly stationary time series with non-Gaussian distributions. We show that in strictly stationary processes, the hypothesis of the absence of linear and nonlinear serial dependence can be tested by a multivariate portmanteau test involving nonlinear autocovariances, i.e., the autocovariance (matrices) of nonlinear transformations of a time series. This approach is available for univariate or multivariate processes, and the test statistic follows asymptotically a chi-square distribution under the null hypothesis of serial independence.

The NLSD test is inspired by the Generalized Covariance (GCov) specification test introduced by Gourieroux and Jasiak (2023). The GCov specification test allows for detecting nonlinear serial correlation in the residuals. It is a convenient one-step procedure that can replace multiple Box-Pierce and Ljung-Box tests for diagnostic checking of strictly stationary semi-parametric models with i.i.d. non-Gaussian errors. The advantage of the semi-parametric approach is that it does not require any distributional assumptions on the errors other than being i.i.d. and non-Gaussian. The GCov specification test statistic has an asymptotic chi-square distribution under the null when the dynamic model is estimated by the semi-parametric Generalized Covariance (GCov) estimator, which is consistent, asymptotically normally distributed, and semi-parametrically efficient[1] [Gourieroux and Jasiak (2023)]. We study analytically and through simulations the properties of the GCov specification test under local alternative hypotheses to provide convincing arguments and empirical evidence of its potential as a widely applicable diagnostic tool. We also extend the GCov specification test by considering a set of nonlinear autocovariance conditions increasing to infinity. For size adjustments and power computation in finite samples, we introduce a new GCov bootstrap test. It enlarges the domain of applications to residual diagnostics in parametric models estimated by a Maximum Likelihood (ML) method under a distributional assumption on the errors.

This paper contributes to the literature on univariate and multivariate portmanteau tests. The GCov specification test can be compared to the test of the martingale difference hypothesis of De Gooijer (2023), which relies on a portmanteau test statistic computed from the residuals and squared residuals. The GCov specification test is more general in the sense that it can include the autocovariances of various nonlinear functions of residuals rather than the residuals and their squares only. An alternative approach for specification testing is based on the auto-distance covariance [Fokianos and Pitsillou (2018) and references therein], or its auto-distance correlation (ADCF) counterpart [Davis et al. (2018)]. Wan and Davis (2022) consider the auto-distance covariance (ADCV) function and propose a specification test of the null hypothesis of serial independence of errors [see alsoChu (2023)]. The asymptotic distribution of the ADVC test statistic is generally intractable, and the asymptotic critical values have to be approximated by bootstrap. In comparison, the GCov specification test based on the GCov estimator has the advantage of a known asymptotic distribution. Another difference between the ADCV and GCov methods is in the weights, which in the GCov procedure are optimally selected and estimated rather than being predetermined. For data analysis in the frequency domain, the inference methods of Velasco and Lobato (2018) are available, but are not discussed here as the focus of our paper is on the time domain only.

[1] It achieves the parametric efficiency for a well-chosen set of nonlinear transformations.

This paper also makes a contribution to the practice of time series analysis. There exists a growing literature on the SVAR models, motivated by the fact that shock identification in these models is feasible by assuming that the structural shocks are independent and non-Gaussian. In applied research, residual diagnostics are often disregarded, likely because applying multiple Ljung-Box tests to the data and then to the model residuals and their powers, for example, is too cumbersome. Then, the model selection is based only on the goodness-of-fit criteria. For example, Gouriéroux et al. (2017), Lanne and Luoto (2021), Guay (2021), and Keweloh (2021) assume in their application that the errors are independent and non-Gaussian without reporting the results of tests for the absence of (non)linear serial dependence. Gouriéroux et al. (2020) write only that the Box and Pierce (1970) and Ljung-Box tests of the null hypothesis of absence of autocorrelation were applied to the residuals. Lanne et al. (2017) provide the most detailed description of applying the Ljung-Box test to residuals and report p-values. Thus, applied research would benefit from a convenient one-step test designed to detect nonlinear dependence in residuals.

The GCov-based specification test, in addition to being applicable to a variety of models with i.i.d. non-Gaussian errors, is particularly useful for testing the fit of causal-noncausal dynamic models with non-Gaussian errors. The reason is that nonlinear autocovariances identify both causal and noncausal dynamics [Chan et al. (2006)]. There is a growing interest in univariate and multivariate mixed causal-noncausal models for processes with locally explosive patterns, such as short-lasting trends, bubbles and spikes, and time-varying volatility [see e.g. Hecq et al. (2016), Gourieroux and Jasiak (2017, 2023), Gouriéroux and Zakoïan (2017), Fries and Zakoian (2019), Davis and Song (2020), Hecq and Voisin (2021), Swensen (2022)]. In practice, locally explosive patterns characterize the time series of commodity prices, including oil, soybean, nickel, and aluminum prices, as well as cryptocurrency prices of native cryptocurrencies, tokens, and some stablecoins. The estimators available for this class of models are the aforementioned semi-parametric GCov estimator [Gourieroux and Jasiak (2017, 2023)] and the parametric (Approximate) Maximum Likelihood (AML), or

ML estimators [Lanne and Saikkonen (2011), Davis and Song (2020)]. When the error distribution of a mixed model is misspecified, the AML estimator can be unreliable [Hecq et al. (2016)], adversely affecting the outcomes of any ML-based fit criteria, such as the AIC, Schwarz, and Hannan-Quinn criteria. Hence, in these models, the proposed tests arise as convenient tools for both detecting nonlinear serial dependence in the data and testing the goodness of fit.

The following notation is used. For any $m \times K$ matrix A with the kth column $a_k, k = 1, ..., K$ $vec(A)$ will denote the column vector of dimension mK defined as:

$$vec(A) = (a_1'..., a_k', ..., a_K')',$$

where the prime denotes a transpose. For any two matrices $A \equiv (a_{ik})$ and B, the Kronecker product $(A \otimes B)$ is the block matrix with the (i, k)th block denoted by $a_{ik}B$.

This chapter is organized as follows. Section 2.2 introduces the NLSD test for time series, inspired by the classical multivariate portmanteau test, which is briefly reviewed. Section 2.3 reviews the GCov estimator and describes the GCov specification test. It provides new results on the asymptotic properties of this test under a sequence of local alternatives. Section 2.4 introduces the GCov bootstrap test. Section 2.5 examines how the asymptotic performance of the GCov specification test can be improved by increasing the set of nonlinear autocovariances. Section 2.6 presents the simulation results. Section 2.7 presents an empirical application of the GCov specification test to a mixed causal-noncausal model fitted to monthly aluminum prices recorded between 2005 and 2024. Section 2.8 concludes. Appendices A and B provide technical results for this chapter, and Appendix C includes empirical investigations.

1.2 NLSD Test for Non-Gaussian Processes

In this section, we first recall the existing results on the weak white noise hypothesis test. Next, we introduce a new test of the absence of linear and nonlinear serial dependence (NLSD) for non-Gaussian processes based on a portmanteau statistic involving nonlinear transformations of strictly stationary univariate or multivariate time series with non-Gaussian marginal distributions and nonlinear dynamics.

1.2.1 Linear Serial Dependence in Time Series

This section summarizes the results on the test of the weak white noise hypothesis, i.e., the absence of linear serial dependence in time series.

1.2.1.1 Univariate Time Series

Let us consider a univariate stationary time series (y_t) with finite fourth-order moments[2]. The test of the weak white noise hypothesis $H_0 = \{\gamma(h) = 0, h = 1, ..., H\}$, with $\gamma(h) = Cov(y_t, y_{t-h})$ is commonly based on the test statistic:

$$\hat{\xi}_T(H) = T \sum_{h=1}^{H} \hat{\rho}(h)^2 = T \sum_{h=1}^{H} \frac{\hat{\gamma}(h)^2}{\hat{\gamma}(0)^2}, \tag{2.1}$$

where $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ are the sample autocovariance and autocorrelation of order h, respectively, computed from a sample of T observations[3].

Under the null hypothesis of serial independence and standard regularity conditions, this statistic follows asymptotically a chi-square distribution $\chi^2(H)$ with H degrees of freedom [see Box and Pierce (1970)].

[2]The existence of moments up to order 4 is needed to derive the asymptotic variance of $\hat{\gamma}(h)$ under the asymptotic normality.

[3]In this Section the index T of the estimators, e.g., $\hat{\gamma}_T(h)$, is omitted to simplify the notation.

1.2.1.2 Multivariate Time Series

Let us now consider a strictly stationary time series (Y_t) of dimension m with finite fourth-order moments. The null hypothesis is $H_0 = \{\Gamma(h) = 0, h = 1, ..., H\}$, where $\Gamma(h) = Cov(Y_t, Y_{t-h})$ is the autocovariance (matrix) of order h. The multivariate equivalent of the test statistic (2.1) is:

$$\hat{\xi}_T(H) = T \sum_{h=1}^{H} Tr[\hat{R}^2(h)], \tag{2.2}$$

where $\hat{R}^2(h)$ is the sample analogue of the multivariate R-square defined by:

$$R^2(h) = \Gamma(h)\Gamma(0)^{-1}\Gamma(h)'\Gamma(0)^{-1}. \tag{2.3}$$

Since

$$\hat{R}^2(h) = \hat{\Gamma}(0)^{1/2}[\hat{\Gamma}(0)^{-1/2}\hat{\Gamma}(h)\hat{\Gamma}(0)^{-1}\hat{\Gamma}(h)'\hat{\Gamma}(0)^{-1/2}]\hat{\Gamma}(0)^{-1/2}, \tag{2.4}$$

this matrix is equivalent up to a change of basis to the matrix within the brackets, which is symmetric and positive-definite. Therefore, it is diagonalizable, with a trace equal to the sum of its eigenvalues, which are the squares of the empirical canonical correlations between Y_t and Y_{t-h}, denoted by $\hat{\rho}_i^2(h), i = 1, ..., m$ [Hotelling (1992)]:

$$\hat{\xi}_T(H) = T \sum_{h=1}^{H} Tr[\hat{\Gamma}(h)\hat{\Gamma}(0)^{-1}\hat{\Gamma}(h)'\hat{\Gamma}(0)^{-1}] = T \sum_{h=1}^{H} [\sum_{i=1}^{m} \hat{\rho}_i(h)^2].$$

Under the null hypothesis of strong white noise, this statistic follows asymptotically a chi-square distribution $\chi^2(mH)$ [see, e.g., Robinson (1973), Anderson (1999), Section 7, Anderson (2002), Section 5].

1.2.2 Nonlinear Serial Dependence in Time Series

The NLSD test based on nonlinear functions of a strictly stationary process with a non-Gaussian distribution extends the approach presented in the previous section. For processes with Gaussian distributions, zero-valued autocovariances are equivalent to serial independence, which becomes the null hypothesis of interest. Then, the asymptotic distribution of the test statistics is determined under that null hypothesis. In the case of non-Gaussian processes, zero-valued linear autocovariances do not imply serial independence. Chan et al. (2006) and Gourieroux and Jasiak (2023) show that the nonlinear autocovariances, i.e., the autocovariances of nonlinear functions of non-Gaussian processes reveal and identify the nonlinear and noncausal serial dependencies. This suggests that the null hypothesis of the absence of (non)linear serial dependence in univariate or multivariate time series Y_t can be tested by applying the test statistic $\hat{\xi}_T(H)$ to a set of nonlinear transformations of Y_t. Thus, the NLSD test concerns the null hypothesis of zero autocovariances of the transformed series.

We consider nonlinear transformations Y_t^a of a univariate or multivariate non-Gaussian process Y_t, where $a = \{a_1, a_2, ...\}$ is a set of nonlinear scalar functions satisfying the regularity conditions given in Gourieroux and Jasiak (2023). The transformations include Y_t to accommodate linear serial dependence and the powers of Y_t, for example, to account for nonlinear serial dependence. If the process has fat tails, the existence of the fourth-moment condition may not be satisfied by Y_t, but instead by a well-chosen transformation $a_k(Y_t)$. Hence, transformations such as $\sqrt{Y_t}$, $\log Y_t$, or $1/(1 + \exp(-Y_t))$ of univariate positive processes can be used depending on the tails of their distributions.

Suppose, we consider the functions $a_1, ..., a_K$ of the process (Y_t), transforming it into a multivariate process of dimension K with components $a_k(Y_t)$:

$$Y_t^a = \begin{pmatrix} a_1(Y_t) \\ \vdots \\ a_K(Y_t) \end{pmatrix},$$

where for a univariate time series y_t, the transformation $a_1(y_t) = y_t$ is the time series itself, allowing the test to capture the linear dependence. We compute the sample autocovariances of the transformations $a_k(Y_t), k = 1, ..., K$:

$$\hat{\Gamma}^a(h) = \frac{1}{T}\sum_{t=h+1}^{T} Y_t^a Y_{t-h}^{a'} - \frac{1}{T}\sum_{t=1}^{T-h} Y_t^a \frac{1}{T}\sum_{t=h+1}^{T} Y_{t-h}^{a'}.$$

Once a set of transformations is determined, the null hypothesis:

$$H_{0,a} = (\Gamma^a(h) = 0,\ h = 1, ..., H),$$

is used to test for the absence of (non)linear serial dependence. Because one cannot consider all the lags and nonlinear transformations, the null hypothesis $H_{0,a}$ is not equivalent to the serial independence condition, but is arbitrarily close to it, depending on the lag H and the set of nonlinear transformations considered [see Section 2.5]. Then, the NLSD test statistic:

$$\hat{\xi}_T(H)_a = T\sum_{h=1}^{H} Tr[\hat{R}_a^2(h)], \tag{2.5}$$

where:

$$\hat{R}_a^2(h) = \hat{\Gamma}_a(h)\hat{\Gamma}_a(0)^{-1}\hat{\Gamma}_a(h)'\hat{\Gamma}_a(0)^{-1}],$$

is computed from the nonlinear sample autocovariances, i.e., sample autocovariance matrices of a transformed univariate or multivariate process. In each case, the dimension of the process can be increased, i.e., become higher than that of the initial time series of interest. By analogy to the traditional literature, we use the independence hypothesis to obtain the asymptotic distribution of the test statistic. If the dimension of the transformed process is K, then under serial independence, the NLSD test statistic (2.5) follows asymptotically a $\chi^2(K^2H)$ distribution [see, e.g., Robinson (1973),Chitturi (1976), Anderson (1999), Section 7, Anderson (2002), Section 5]. The test of the null hypothesis $H_{0,a}$ at level α is performed as follows: the null hypothesis $H_{0,a}$ is rejected when $\hat{\xi}_T(H)_a > \chi^2_{1-\alpha}(K^2H)$ and $H_{0,a}$ is not rejected, otherwise.

1.3 GCov Specification Test for Semi-Parametric Models

The NLSD test introduced in Section 2.2 is a special case of the Generalized Covariance (GCov) specification test introduced by Gourieroux and Jasiak (2023) for semi-parametric nonlinear models of strictly stationary time series with i.i.d. errors and parameter vector θ describing their dynamics. In this context, $\hat{\xi}_T(H)$ is computed from a multivariate time series of residuals and their nonlinear transforms instead of an observed time series, which reduces the degrees of freedom of its limiting χ^2 distribution [Gourieroux and Jasiak (2023)]. The model and GCov test are reviewed below.

1.3.1 The Semi-Parametric Model

Let us consider a strictly stationary process (Y_t) satisfying a semi-parametric model studied in Gourieroux and Jasiak (2023):

$$g(\tilde{Y}_t; \theta) = u_t, \tag{3.1}$$

where g is a known function with $dim(g) = dim(u_t)$, $\tilde{Y}_t = (Y_t, Y_{t-1}, \dots, Y_{t-p})$, p is a non-negative integer, (u_t) is an i.i.d. sequence of errors (not necessarily with mean zero) with a common marginal density function f and θ is an unknown parameter vector of dimension $dim(\theta)$. We assume that the model is well-specified, the true value of parameter θ is θ_0, and the true error density is f_0. Moreover, u_t is not assumed to be independent of $\tilde{Y}_{t-1}$. Hence, errors u_t are not necessarily interpretable as innovations, assuming that $g(Y_t, Y_{t-1}, ..; \theta)$ is an invertible function of Y_t.

Similarly, as was done in Section 2.2, a process following model (3.1) can be transformed into a system of dimension K by considering (linear and) nonlinear scalar transformations of u_t. Let us introduce K nonlinear transformations $a_1, ..., a_K$. Then we have:

$$
\begin{aligned}
a_k[g(\tilde{y}_t;\theta)] &= a_k(u_t),\ k=1,...,K, \\
\text{or, equivalently}\ \ a[g(\tilde{y}_t;\theta)] &= a(u_t) \equiv v_t,
\end{aligned} \tag{3.2}
$$

where the transformed process (v_t) is also an i.i.d. process. Henceforth, the subscripts of transformations a are omitted for clarity.

1.3.2 The GCov Specification Test

We consider the null hypothesis of the absence of (non)linear serial dependence in i.i.d. errors whose distribution is left unspecified. The portmanteau test statistic for testing the model specification is:

$$
\hat{\xi}_T(H) = TL_T(\hat{\theta}_T), \ \ \text{where}\ \ L_T(\hat{\theta}_T) = \sum_{h=1}^{H} Tr[\hat{R}_T^2(h,\hat{\theta}_T)], \tag{3.3}
$$

and

$$
\hat{R}_T^2(h,\theta) = \hat{\Gamma}_T(h;\hat{\theta}_T)\hat{\Gamma}_T(0;\hat{\theta}_T)^{-1}\hat{\Gamma}_T(h;\hat{\theta}_T)'\hat{\Gamma}_T(0;\hat{\theta}_T)^{-1}, \tag{3.4}
$$

and the estimated autocovariances $\hat{\Gamma}_T(h;\hat{\theta}_T)$ are the sample autocovariances[4] of the residuals $\hat{u}_{t,T} = u_t(\hat{\theta}_T) = g(\tilde{Y}_t,\hat{\theta}_T)$ evaluated at the GCov estimator $\hat{\theta}_T$ of θ defined by:

$$
\hat{\theta}_T(H) = Argmin_{\theta} \sum_{h=1}^{H} Tr[\hat{R}_T^2(h,\theta)]. \tag{3.5}
$$

When model (3.1) is well-specified, and the true value θ_0 is identifiable from the zero nonlinear autocovariance conditions, the one-step GCov estimator is consistent, asymptotically normally distributed, and attains a semi-parametric efficiency bound, under standard regularity conditions [Gourieroux and Jasiak (2023), assumptions A.1, A.2 and Proposition

[4]They have to be divided by T instead of $(T-H-p)$ to ensure that the sequence of multivariate sample autocovariances remains positive semi-definite.

3]. Then, the GCov portmanteau test statistic $\hat{\xi}_T(H)$ follows asymptotically the chi-square distribution with degrees of freedom equal to $K^2H - dim(\theta)$ [see Gourieroux and Jasiak (2023), Proposition 4]. This result holds only for the GCov estimator $\hat{\theta}_T$. The GCov specification test can be applied as a diagnostic tool to a variety of dynamic models, with i.i.d. non-Gaussian errors, including the following ones:

Example 1: Double Autoregressive DAR(1) Model

$$y_t = \phi y_{t-1} + u_t\sqrt{w + \alpha y_{t-1}^2},$$

where $w > 0, \phi \geq 0, \alpha \geq 0$, $\theta = (w, \phi, \alpha)'$ and the u_t's are i.i.d. with a Gaussian or non-Gaussian distribution [Ling (2007)]. We assume that $E(log|\phi + \sqrt{\alpha} u|) < 0$ and the regularity conditions on functions θ, f are satisfied, ensuring the existence of a strictly stationary solution. Then, the semi-parametric representation (3.1) of this model is

$$g(\tilde{y}_t; \theta) = [(y_t - \phi y_{t-1})/\sqrt{w + \alpha y_{t-1}^2} = u_t,$$

In particular, this process is strictly stationary for $\phi = 1$ (i.e., in the unit root case) due to the volatility induced mean reversion.

Example 2: Mixed (Causal-Noncausal) Autoregressive MAR(r,s) Model

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_r L^r)(1 - \psi_1 L^{-1} - \psi_2 L^{-2} - ... - \psi_s L^{-s})y_t = u_t, \tag{3.6}$$

where the errors are i.i.d., non-Gaussian and such that $E(|u_t|^\delta) < \infty$ for $\delta > 0$. The polynomials $\Phi(L)$ and $\Psi(L)$ in L are of degrees r and s, respectively, with roots strictly outside the unit circle and such that $\Phi(0) = \Psi(0) = 1$. For $r = s = 1$, the MAR(1,1) model is obtained:

$$(1 - \phi L)(1 - \psi L^{-1})y_t = u_t, \tag{3.7}$$

where the autoregressive coefficients satisfy the condition $|\phi| < 1$ and $|\psi| < 1$ for strict stationarity. Coefficient ϕ represents the standard causal persistence, while coefficient ψ determines noncausal persistence, along with locally explosive patterns and conditional heteroscedasticity. For the MAR(1,1) we have $\theta = (\phi, \psi)'$, and the semi-parametric representation (3.1) of this model is:

$$g(\tilde{y}_t; \theta) = \Phi(L)\Psi(L^{-1})y_t = u_t.$$

Example 3: Causal SVAR(p) Model

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + Du_t,$$

where $\theta = [vec\Phi_1', ..., vec\Phi_p']'$, u_t is a multivariate, non-Gaussian, serially and cross-sectionally i.i.d. error and matrix D is square and invertible. The roots of the characteristic equation $det(Id - \Phi_1\lambda - \cdots \Phi_p\lambda^p) = 0$ are of modulus strictly greater than one. We have

$$g(\tilde{Y}_t, \theta) = D^{-1}[Y_t - \Phi_1 Y_{t-1} - \cdots - \Phi_p Y_{t-p}] = u_t.$$

This model is commonly used in macroeconomic applications.

Example 4: Causal-Noncausal VAR(p) Model

The model is specified as above except that the roots of the characteristic equation $det(Id - \Phi_1\lambda - \cdots \Phi_p\lambda^p) = 0$ are of modulus either strictly greater, or smaller than one and the errors are not necessarily assumed to be cross-sectionally independent. The roots located inside the unit circle create locally explosive patterns and conditional heteroscedasticity, like in the MAR(r,s) model. There exists a unique (strictly) stationary solution (Y_t) with a two-sided $MA(\infty)$ representation, which satisfies model (3.1) with:

$$g(\tilde{Y}_t, \theta) = Y_t - \Phi_1 Y_{t-1} - \cdots - \Phi_p Y_{t-p} = u_t.$$

The causal-noncausal VAR(p) model has been studied in Gourieroux and Jasiak (2017, 2023), Davis and Song (2020), and Swensen (2022).

1.3.3 Asymptotic Properties of the GCov Specification Test

This section presents new results on the asymptotic properties of the GCov specification test under local alternatives. By focusing on the local alternatives, we can study hypotheses on the parameters of a semi-parametric model, given the marginal error distribution and extend the results on the asymptotic distribution of the GCov test statistic derived in Gourieroux and Jasiak (2023).

1.3.3.1 Null Hypothesis and Asymptotic Size

Let us discuss the null hypothesis in the semi-parametric framework. As mentioned earlier, there are two types of parameters: vector θ defining the dynamics, and functional parameter f defining the error distribution. Hence, the theoretical autocovariances $\Gamma(h;\theta,f)$ are functions of both θ and f. Suppose that θ_0, f_0 are identifiable from the condition: $\Gamma_0(h;\theta,f)=0,\ h=1,...,H$ [5]. Then, the null hypothesis becomes:

$$H_0 : \{\Gamma_0(h)=0,\ h=1,...,H\},$$

where $\Gamma_0(h)$ is the true autocovariance. In terms of parameters θ, f, it corresponds to the parameter set $\Theta_0 = \{\theta, f : (\Gamma(h;\theta,f)=0,\ h=1,...,H\}$. So far, the transformation a has not been introduced to simplify the notation. When these transformations are accounted for, the null hypothesis and the associated parameter set become:

$$H_{0,a} \approx \Theta_{0,a} = \{\theta, f : \Gamma_{0,a}(h;\theta,f)=0,\ h=1,...,H\}.$$

[5]For example, in the SVAR model, identification can be ensured by assuming that the components of errors u_t are cross-sectionally independent and non-Gaussian. A detailed discussion of identification is out of the scope of this paper.

where $\Gamma_{0,a}(h;\theta,f)$ is the true autocovariance of the transformed errors at lag h. When the error terms u_t are serially independent, the nonlinear autocovariances $\Gamma_{0,a}(h;\theta_0,f_0)=0,\ \forall h$. Hence, if the model is correctly specified, the GCov test does not reject the null hypothesis $H_{0,a}$ of the absence of linear and nonlinear dependence in the errors u_t.

The test of hypothesis $H_{0,a}$ at level α is performed as follows: the null hypothesis $H_{0,a}$ is rejected when $\hat{\xi}_T(H)_a > \chi^2_{1-\alpha}(K^2H - dim(\theta))$ and $H_{0,a}$ is not rejected otherwise. The asymptotic size tends to the nominal size when $T\to\infty$:

$$\lim_{T\to\infty} P_{\theta,f}[\hat{\xi}_T(H)_a > \chi^2_{1-\alpha}(K^2H - dim(\theta))] = \alpha,\ \forall\theta, f\in\Theta_0,\ \forall\alpha.$$

1.3.3.2 Local Alternatives and Local Power

Rather than evaluating the performance of our test against a fixed alternative, we can examine a sequence of local alternatives that drift towards the null at rate $\frac{1}{\sqrt{T}}$. The drifting sequence of local alternatives makes it harder and harder to reject the null as the sample grows. The evaluation of the test under several types of local alternatives reveals the direction of alternative to consider when the null hypothesis is rejected [Davidson and MacKinnon (2006), Section 3, Escanciano (2007), Section 2.3, Dovonon et al. (2020), Section 4].

Because the semi-parametric model depends not only on the vector of dynamic parameters θ but also on the functional parameter f, the general alternative hypothesis is difficult to formulate. In particular, the alternative could contain models in which u_t's are serially dependent, and we would need to introduce additional parameters to accommodate these cases.

For clarity, we omit in this Section the subscripts a of transformation, and assume a semi-parametric alternative model defined by:

$$g^*(\tilde{Y}_t;\theta,\gamma) = u_t,\ \text{where } u_t \text{ are i.i.d.,} \tag{3.8}$$

with an additional (scalar) parameter γ, a known function g^*, such that $g^*(\tilde{Y}_t;\theta,0) = g(\tilde{Y}_t;\theta)$ and dimension $dim(g^*) = dim(u_t)$. Under this alternative model, the autocovariances depend on θ, γ and the marginal distribution f of errors u_t, and are denoted by $\Gamma(h;\theta,\gamma,f)$.

The alternative hypothesis based on autocovariances is parametrized by θ, γ, f. It is written, in terms of the parameters as:

$$H_1 \approx \Theta_1 = \{\theta, \gamma, f : \Gamma(h;\theta,\gamma,f) = 0,\ h = 1, ..., H\}$$

and the null hypothesis is:

$$H_0 = \{\theta, \gamma, f : \Gamma(h;\theta,0,f) = 0,\ h = 1, ..., H\} = \{\gamma = 0\} \cap H_1.$$

Let us examine the fixed and local alternatives in the context of the DAR(1) model. The alternatives concern the presence of serial dependence at higher lags in 1. the conditional mean, and 2. the conditional variance.

Example 4: DAR(1) model cont. The (fixed) alternative models are:
1. $y_t = \phi_1 y_{t-1} + \gamma y_{t-2} + u_t\sqrt{w + \alpha y_{t-1}^2}$, or $u_t = (y_t - \phi_1 y_{t-1} - \gamma y_{t-2})/\sqrt{w + \alpha y_{t-1}^2} = g^*(\tilde{Y}_t;\theta,\gamma)$
2. $y_t = \phi_1 y_{t-1} + u_t\sqrt{w + \alpha y_{t-1}^2 + \gamma y_{t-2}^2}$, or $u_t = (y_t - \phi_1 y_{t-1})/\sqrt{w + \alpha y_{t-1}^2 + \gamma y_{t-2}^2} = g^*(\tilde{Y}_t;\theta,\gamma)$

The above alternative models 1-2 depend on an additional parameter γ. They are equivalent when $\gamma = 0$, or extend the model in various directions when $\gamma \neq 0$. Thus, we have a set of directional alternatives.

The local alternatives are defined in a neighborhood of the true value θ_0, f_0 satisfying the null hypothesis. We consider parametric directional alternatives where:

$$\theta_T \approx \theta_0 + \mu/\sqrt{T}, \quad \gamma_T \approx \nu/\sqrt{T}, \quad f_T \approx f_0.$$

Example 5: DAR(1) model cont. The local alternatives of the DAR model are obtained by the first-order Taylor expansion about $\gamma = 0$ of alternative models 1 to 2 given above. We get:

1. $y_t = \phi_1 y_{t-1} + \gamma y_{t-2} + u_t\sqrt{w + \alpha y_{t-1}^2}$, or $u_t = (y_t - \phi_1 y_{t-1} - \gamma y_{t-2})/\sqrt{w + \alpha y_{t-1}^2} = g^*(\tilde{Y}_t; \theta, \gamma)$
2. $y_t \approx \phi_1 y_{t-1} + u_t\left[\sqrt{w + \alpha y_{t-1}^2} + \frac{\gamma}{2}\frac{y_{t-2}^2}{\sqrt{w+\alpha y_{t-1}^2}}\right]$, or $u_t \approx (y_t - \phi_1 y_{t-1})/\left[\sqrt{w + \alpha y_{t-1}^2} + \frac{\gamma}{2}\frac{y_{t-2}^2}{\sqrt{w+\alpha y_{t-1}^2}}\right] \approx g^*(\tilde{Y}_t; \theta, \gamma)$

Under the sequence of local alternatives, we consider doubly indexed sequences $(y_{T,t})$, i.e., a sequence of processes indexed by T (triangular array). In this framework, what matters is the local impact on autocovariances, i.e.,:

$$\Gamma(h; \theta_T, \gamma_T, f_0) \approx \Gamma(h; \theta_0, 0, f_0) + \frac{\partial \Gamma(h; \theta_0, 0, f_0)}{\partial \theta'}(\theta_T - \theta_0) + \frac{\partial \Gamma(h; \theta_0, 0, f_0)}{\partial \gamma'}(\gamma_T - \gamma_0),$$

with $\Gamma(h; \theta_0, 0, f_0) = 0$. This leads to local alternatives written on the autocovariances:

$$\Gamma(h; \theta_T, \gamma_T, f_0) = \Delta(h; \theta_0, f_0, \mu, \nu)/\sqrt{T}, \tag{3.9}$$

with:

$$\Delta(h; \theta_0, f_0, \mu, \nu) = \frac{\partial \Gamma(h; \theta_0, 0, f_0)}{\partial \theta'}\mu + \frac{\partial \Gamma(h; \theta_0, 0, f_0)}{\partial \gamma'}\nu. \tag{3.10}$$

whose vec representation is denoted by $\delta(h; \theta_0, f_0, \mu, \nu) = vec\Delta(h; \theta_0, f_0, \mu, \nu)$. Then, the asymptotic local power of the test, given f_0 fixed, is

$$\lim_{T\to\infty} P_{\theta_T, \gamma_T, f_0}[\hat{\xi}_T(H) > \chi^2_{1-\alpha}(K^2 H - dim(\theta))] = \beta(\theta_0, f_0, \mu, \nu; \alpha),$$

for any μ, ν, α and $(\theta_0, f_0) \in \Theta_0$.

Proposition 1: Under the sequence of local alternatives, and the regularity conditions given in Appendices B.1-B.2:

i) The autocovariance estimator:

$$\hat{\Gamma}_T(h;\theta) = \frac{1}{T}\sum_{t=1}^{T} g(y_{T,t};\theta)g'(y_{T,t-h};\theta) - \frac{1}{T}\sum_{t=1}^{T} g(y_{T,t};\theta)'\frac{1}{T}\sum_{t=1}^{T} g(y_{T,t};\theta)$$

converges in probability:

$$\hat{\Gamma}_T(h;\theta)\xrightarrow{p}\Gamma(h;\theta_0,0,f_0),$$

for all θ, h where the limit is computed under the null hypothesis. This convergence is uniform in θ.

ii) The GCov estimator converges in probability to θ_0:

$$\hat{\theta}_T\xrightarrow{p}\theta_0.$$

Proof: i) The proof is based on the Law of Large Numbers (LLN) for doubly indexed sequences [see e.g., Andrews (1988), Newey (1991) and Appendices B.1-B.2 for regularity conditions]. The LLN implies the convergence of the estimated autocovariances.

ii) Under the sequence of local alternatives, the objective function:

$$L_T(\theta) = \sum_{h=1}^{H} Tr[\hat{\Gamma}_T(h;\theta)\hat{\Gamma}_T(0;\theta)^{-1}\hat{\Gamma}_T(h;\theta)'\hat{\Gamma}_T(0;\theta)^{-1}], \tag{3.11}$$

tends to

$$L_\infty(\theta) = \sum_{h=1}^{H} Tr[\Gamma(h;\theta_0,0,f_0,\theta)\Gamma(0;\theta_0,0,f_0,\theta)^{-1}\Gamma(h;\theta_0,0,f_0,\theta)'\Gamma(0;\theta_0,0,f_0,\theta)^{-1}]. \tag{3.12}$$

This limit is the same as the limit of the objective function under the null hypothesis. Then, the consistency is proven as in Gourieroux and Jasiak (2023). QED

Let us now show that the GCov specification test statistic converges in distribution under local alternatives to a non-central chi-square distributed variable with a non-centrality parameter involving the direction of the alternative. The distribution of the test statistic computed from the residuals of the model $g(y_t, \theta) = u_t$ estimated by the GCov estimator under the sequence of local alternatives is given in Proposition 2 below.

Proposition 2: Let us consider the specification test of the null hypothesis:

$$H_0 = \Theta_0 = \{\theta, f : \Gamma(h; \theta, f) = 0 \ \ \forall h = 1, ..., H\},$$

against the sequence of local alternatives:

$$H_{1,T} = \Theta_{1,T} = \{\theta = \theta_0 + \mu/\sqrt{T}, \gamma = \nu/\sqrt{T}, f = f_0, \text{ with } (\theta_0, f_0) \in \Theta_0\}.$$

The expansion of the test statistic under the sequence of local alternatives is:

$$\hat{\xi}_T(H) = T \sum_{h=1}^{H} \{vec[\sqrt{T}\hat{\Gamma}_T(h; \theta_T, \gamma_T, f_0)]\Pi(h; \theta_0, f_0)vec[\sqrt{T}\hat{\Gamma}_T(h; \theta_T, \gamma_T, f_0)]\} + o_p(1), \quad (3.13)$$

where

$$\begin{aligned}
\Pi(h; \theta_0, f_0) \quad = \quad & [\Gamma_0(0, \theta_0, f_0)^{-1} \otimes \Gamma_0(0, \theta_0, f_0)^{-1}] - [\Gamma_0(0, \theta_0, f_0)^{-1} \otimes \Gamma_0(0, \theta_0, f_0)^{-1}]\frac{\partial vec\Gamma(h, \theta_0, f_0)}{\partial \theta'} \\
& \left\{\frac{\partial vec\Gamma(h, \theta_0, f_0)'}{\partial \theta}[\Gamma_0(0, \theta_0, f_0)^{-1} \otimes \Gamma_0(0, \theta_0, f_0)^{-1}]\frac{\partial vec\Gamma(h, \theta_0, f_0)}{\partial \theta}\right\}^{-1} \\
& \times \frac{\partial vec\Gamma(h, \theta_0, f_0)'}{\partial \theta'}[\Gamma_0(0, \theta_0, f_0)^{-1} \otimes \Gamma_0(0, \theta_0, f_0)^{-1}].
\end{aligned}$$

Then, under the sequence of local alternatives, $\hat{\xi}_T(H) \overset{d}{\sim} \chi^2(K^2H - dim(\theta), \lambda(\theta_0, f_0, \mu, \nu))$, where the non-centrality parameter is:

$$\lambda(\theta_0, f_0, \mu, \nu) = \sum_{h=1}^{H} \delta(h; \theta_0, f_0, \mu, \nu)'\Pi(h; \theta_0, f_0)\delta(h; \theta_0, f_0, \mu, \nu)),$$

with $\delta(h; \theta_0, f_0, \mu, \nu) = \frac{\partial \Gamma(h; \theta_0, 0, f_0)}{\partial \theta'}\mu + \frac{\partial \Gamma(h, \theta_0, 0, f_0)'}{\partial \gamma'}\nu$.

Proof: The proof of Proposition 2 is based on the Central Limit Theorem (CLT) for doubly indexed sequences [see, e.g., Wooldridge and White (1988)].

Under the regularity conditions 1-6, it is easy to see that expansions similar to (B.3)-(B.4) are still valid under the sequence of local alternatives by using the LLN for triangular arrays and the convergence of order $1/\sqrt{T}$ of the estimated autocovariances that follows from the CLT. For example, we still have the Taylor series expansion:

$$\hat{\xi}_T(H) = T \sum_{h=1}^{H} \{vec[\sqrt{T}\hat{\Gamma}_T(h; \theta_T, \gamma_T, f_0)]\Pi(h; \theta_0, f_0)vec[\sqrt{T}\hat{\Gamma}_T(h; \theta_T, \gamma_T, f_0)]\} + o_p(1),$$

similar to expansion (B.4) where the vectors $vec[\sqrt{T}\hat{\Gamma}_T(h; \theta_T, \gamma_T, f_0)]$, $h = 1, ..., H$, are now asymptotically independent with the distribution:

$N[\delta(h; \theta_0, f_0, \mu, \nu), \Gamma(0; \theta_0, f_0) \otimes \Gamma(0; \theta_0, f_0)]$,

by the CLT.

Then, under the sequence of local alternatives, the asymptotic distribution of $\hat{\xi}_T(H)$ is a chi-square distribution with the non-centrality parameter λ:

$$\lambda(\theta_0, f_0, \mu, \nu) = \sum_{h=1}^{H} \delta(h, \theta_0, f_0, \mu, \nu)'\Pi(h; \theta_0, f_0)\delta(h, \theta_0, f_0, \mu, \nu),$$

and a degree of freedom equal to the rank of matrix $\Pi(H; \theta_0, f_0) = diag[\Pi(h; \theta_0, f_0)]$, where diag denotes a diagonal matrix.

These results can be applied to test the null hypothesis $H_0 = (y_t = u_t)$ without the parameter θ (NLSD test) and other forms of local alternatives.

Let us consider the test of the absence of linear dependence in the time series $y_t = u_t, t = 1, ..., T$ against the local alternatives of an autoregressive form. More specifically, we test

$$H_0 : \{\Gamma_0(h) = 0, \forall h = 1, ..., H\} = \{B_1 = \cdots = B_H = 0\},$$

against the local alternatives. The local alternatives can be defined in terms of the autoregressive parameters $B_1, ..., B_H$, or equivalently in terms of the autocovariances $\Gamma(h), h = 1, ..., H$.

Thus, the additional parameter γ is not necessarily a scalar. We follow the latter approach with the sequence of local alternatives:

$$H_{1,T} = \{\Gamma_T(h) = \Delta(h)/\sqrt{T},\ h = 1, ..., H\} = \{vec\Gamma_T(h) = \delta(h)/\sqrt{T},\ h = 1, ..., H\},$$

with $\delta(h) = vec\Delta(h)$.

Under the sequence of local alternatives, the estimated autocovariances are asymptotically independent with the asymptotic normal distributions.

$$vec[\sqrt{T}\hat{\Gamma}_T(h)'] \overset{a}{\sim} N[\delta(h), \Gamma(0) \otimes \Gamma(0)].$$

Hence:

$$[\Gamma(0)^{-1/2} \otimes \Gamma(0)^{-1/2}]vec[\sqrt{T}\hat{\Gamma}_T(h)'] \overset{a}{\sim} N[(\Gamma(0)^{-1/2} \otimes \Gamma(0)^{-1/2})\delta(h), Id]$$

It follows that the portmanteau statistic $\hat{\xi}_T(H)$ has asymptotically, under the sequence of local alternatives, a chi-square $\chi^2(K^2H, \lambda)$ distribution with the non-centrality parameter λ, where

$$\lambda = \sum_{h=1}^{H} \delta(h)'[\Gamma(0)^{-1/2} \otimes \Gamma(0)^{-1/2}][\Gamma(0)^{-1/2} \otimes \Gamma(0)^{-1/2}]\delta(h) = \sum_{h=1}^{H} \delta(h)'[\Gamma(0)^{-1} \otimes \Gamma(0)^{-1}]\delta(h), \tag{3.14}$$

is the non-centrality parameter.

Let the cumulative distribution function (c.d.f.) of the non-central chi-square distribution be denoted by $F(x; \kappa, \lambda)$, where κ denotes the degrees of freedom and λ is the non-centrality parameter. Then, $F(x; \kappa, \lambda) = 1 - Q_{\kappa/2}(\sqrt{\lambda}, \sqrt{x})$, where $Q_\delta(a, b)$ is a Marcum Q-function.

For positive integer values of δ, the Marcum Q-function is defined by:

$$Q_\delta(a,b) = \begin{cases} H_\delta(a,b) & a < b, \\ 0.5 + H_\delta(a,a), & a = b, \\ 1 + H_\delta(a,b), & a > b, \end{cases}$$

where $H_\delta(a,b) = \frac{\zeta^{1-\delta}}{2\pi} exp(-\frac{a^2+b^2}{2}) \int_0^{2\pi} \frac{cos(\delta-1)w - \zeta cos\,\delta\, w}{1-2\zeta cos w + \zeta^2} exp(ab\, cos\, w) dw$ and $\zeta = a/b$. Then, we get:

Corollary 1: The local asymptotic power of the GCov specification test is defined as:

$$\beta(\theta_0, f_0, \mu, \nu; \alpha) = Q_{(K^2 H - dim(\theta))/2}[\sqrt{\lambda(\theta_0, f_0, \mu, \nu)}, \sqrt{\chi^2_{1-\alpha}(K^2 H - dim(\theta))}],$$

where $\lambda(\theta_0, f_0, \mu, \nu)$ is given in Proposition 2.

From the monotonicity property of the Q-function, it follows that the local asymptotic power function is strictly decreasing in the non-centrality parameter $\lambda(\theta_0, f_0, \mu, \nu)$.

1.4 Bootstrap-Adjusted GCov Test

Even though the asymptotic distribution of the GCov test statistic is known to be a χ^2 distribution [resp. non-central χ^2 distribution] under the null hypothesis [resp. local alternatives], and asymptotically valid critical values are available, one may be interested in finding critical values for hypothesis testing in finite samples. This section describes the bootstrap for GCov test statistic under the regularity conditions given in the end of this section.

1.4.1 Bootstrap Under the Null Hypothesis

We approximate the distribution of $\tilde{\xi}_T$ by computing the test statistic $\xi(H, \hat{\theta}^s_T, \hat{f}^s_T)$ with $\hat{\theta}^s_T$ and $\hat{f}^s_T$ obtained from the bootstrapped residuals, and the bootstrapped values $y^s_1, ..., y^s_T$. We assume $\tilde{Y}_t = (Y_t, Y_{t-1}, ..., Y_{t-p})$ and denote by c the inverse of function g with respect to y_t.

Then, the critical value of the test statistic $\tilde{\xi}_T(H)$ can be found by parametric bootstrap along the following steps:

1. Draw randomly with replacements T residuals $\hat{u}^s_{T,t}, t = 1, ..., T$ from residuals $\hat{u}_{T,t} = g(\tilde{Y}_t, \hat{\theta}_T), t = 1, ..., T$.

2. Build the bootstrapped time series of length T: $y^s_{T,t} = c(y^s_{T,t-1}, ..., y^s_{T,t-p}, \hat{u}^s_{T,t}, \hat{\theta}_T), t = 1, ..., T,\ s = 1, ..., S$ [6].

3. Re-estimate the parameter vector θ by GCov from $y^s_{T,t},\ t = 1, ..., T$, providing $\hat{\theta}^s_T,\ s = 1, ..., S$. Under the regularity conditions discussed in Appendix B.7, the asymptotic distribution of $\sqrt{T}(\hat{\theta}^s_T - \hat{\theta}_T)$ conditional on the sample $y_t, t = 1, ..., T$ is the same as the asymptotic distribution of $\sqrt{T}(\hat{\theta}_T - \theta_0)$.

4. Compute the test statistic $\hat{\xi}^s_T(H)$ from $\hat{u}^s_{T,t}, t = 1, ..., T$, their nonlinear transforms and $\hat{\theta}^s_T$.

5. Rank the test statistics $\hat{\xi}^s_T(H),\ s = 1, ..., S$, and use the 95th percentile $\hat{q}_{T,95\%}$, say, as the critical value for testing the null hypothesis of absence of nonlinear or linear serial dependence. Then, the null hypothesis is rejected if:

$$\hat{\xi}_T(H) > \hat{q}_{T,95\%},$$

and it is not rejected, otherwise. Henceforth, the bootstrap size-adjusted test is called the GCov bootstrap test.

1.4.2 Bootstrap Analysis of the Local Power of Test

The bootstrap can also be used to obtain an approximation of the local power of the size-adjusted GCov bootstrap test (see Escanciano (2007), Sections 2.3 and 3). The approach

[6] with the starting values $y^s_{T,-1} = y_{-1}, ..., y_{T,-p} = y_{-p}$, if the stationary process is causal and such that u_t is independent of $y_{t-1}, ..., y_{t-p} \forall t$. If the process is a mixed VAR, the bootstrapped values $y^s_1, ..., y^s_T$ can be computed from $\hat{u}^s_{T,t}$ using the formulas in Gourieroux and Jasiak (2017). For a VAR model in a multiplicative representation, see Lanne and Saikkonen (2013). For univariate MAR(r,s) processes, see Gourieroux and Jasiak (2016) and Section 6.1.

involves two types of bootstrap performed under the null and under the alternative, respectively. The local power approximation is obtained as follows:

Step 1: Estimate the model $g(\tilde{y}_t, \theta, \gamma) = u_t,\ t = 1, .., T$ under the alternative to get $\hat{\theta}_{1,T}, \hat{\gamma}_{1,T}$.

Step 2: Compute the residuals under the alternative: $\hat{u}_{1,T,t} = g(y_t, \hat{\theta}_{1,T}, \hat{\gamma}_{1,T})$.

Step 3: Get the bootstrapped residuals $\hat{u}^s_{1,b,T,t}$ by drawing in the sample distribution of $\hat{u}_{1,T,t}, t = 1, ..., T$ for $s = 1, ..., S$.

Step 4: Calculate the bootstrapped pseudo-observations $y^s_{1,b,T,t}$ from

$$g(\tilde{y}^s_{1,b,T,t}; \hat{\theta}_{1,T}, \hat{\gamma}_{1,T}) = \hat{u}^s_{1,b,T,t},\ s = 1, ..., S.$$

$$\Longleftrightarrow\ y^s_{1,b,T,t} = c(y^s_{1,b,T,t-1}, ..., y^s_{1,b,T,t-p}, \hat{u}^s_{1,b,T,t}, \hat{\theta}_{1,T}, \hat{\gamma}_{1,T}),\ s = 1, ..., S.$$

Step 5: From $y^s_{1,b,T,t}, t = 1, ..., T$ bootstrapped under the alternative, compute the values of the estimates $\hat{\theta}^s_{0,b,T}$ of θ under the null and the associated test statistic $\hat{\xi}^s_{0,b,T}, s = 1, ..., S$. Let the test statistic computed under the null be denoted by $\hat{\xi}_{0,T}$. Then, the empirical distribution of $\hat{\xi}^s_{0,b,T} - \hat{\xi}_{0,T}, s = 1, ..., S$ approximates the distribution of $\hat{\xi}_{0,T}$ under the local alternative.

Step 6: Evaluate the bootstrapped local power of the (size-adjusted) GCov bootstrap test as: $1 - \hat{F}^s_{b,T}(\hat{q}_{T,95\%})$, where $\hat{F}^s_{b,T}(\alpha)$ denotes the bootstrap c.d.f. under the alternative and $\hat{q}_{T,95\%}$ is defined in Section 4.1.

1.4.3 An Extended Continuous Updating GMM

To discuss the relationship with the current literature on bootstrap applied to overidentification tests based on moment conditions, we consider a comparable extension of the Continuous Updating GMM (CUGMM) estimator. This extended CUGMM estimates an additional vector of parameters β that is needed to center each moment condition based on nonlinear transformations in order to obtain moment conditions equivalent to nonlinear autocovariance conditions of the GCov.

For ease of exposition, we assume that $H = 1$ and enlarge the parameter set to θ, β, where $dim(\beta) = K$. Then the set of centered moment conditions equivalent to zero nonlinear autocovariance conditions is:

$$\begin{aligned} E[a_k[g(\tilde{Y}_t, \theta)] - \beta_k] &= 0, \; k = 1, ..., K, \\ E\{[a_j[g(\tilde{Y}_t, \theta)] - \beta_j][a_k[g(\tilde{Y}_{t-1}, \theta)] - \beta_k]\} &= 0, \; j, k = 1, ..., K. \end{aligned} \tag{4.1}$$

This set of moment conditions jointly identifies the parameter θ and the additional parameter β.

Proposition 3:

i) Under the condition of just-identification, $dim(\theta) = K^2$ and $H = 1$, the extended CUGMM concentrated in β and applied to the set of moment conditions (4.1) yields the estimators $\hat{\theta}_T, \hat{\beta}_T$ and the objective function $\hat{\xi}_T^*(\hat{\theta}_T, \hat{\beta}_T)$ such that $\hat{\theta}_T$ is equal to the GCov estimator of θ, and the minimum of $\hat{\xi}_T^*(\hat{\theta}_T, \hat{\beta}_T)$ is equal to the minimum value of zero of the GCov objective function.

ii) Under overidentification, $dim(\theta) < K^2$, the GCov and the concentrated extended CUGMM estimators of θ are equivalent at order $1/\sqrt{T}$ and differ at order $1/T$. The associated test statistics have the same asymptotic chi-square distributions at the first order, but differ at higher orders.

Proof: i) If $dim(\theta) = K^2$, due to the just identification of β for a given θ, we can concentrate β out of the CUGMM objective function by using:

$$\hat{\beta}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} a[g(\tilde{Y}_t, \theta)].$$

Next, after substituting $\hat{\beta}_T(\theta)$ into the CUGMM objective function, we find that at the minimum, the concentrated objective function of the extended CUGMM is equal to the minimum of the GCov objective function. The result follows.

ii) Under the overidentification, the asymptotic distribution of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is normal, and both estimators are semi-parametrically efficient for the given set of autocovariance conditions and have the same asymptotic variance. This implies a common chi-square distribution of test statistics at the first order. The asymptotic Taylor expansions of The GCov and extended concentrated CUGMM estimators show a difference in asymptotic biases at order $1/T$ due to different weighting, which implies that these estimators differ at order $1/T$.

QED.

When the lag H is increased to $H > 1$ and higher, the dimensions of the set of moment conditions and of the weighting matrix in the objective function of the extended CUGMM increase by a factor KH. In practice, this may cause an invertibility issue, which will not arise in the GCov where the dimensions of the objective function remain unchanged when H increases.

Since the GCov and extended concentrated CUGMM estimators differ asymptotically by the bias at order $1/T$, and the bootstrap methods applied to the above estimators are expected to adjust for these biases, sufficient regularity conditions for the validity of the bootstrap for the GCov and extended concentrated CUGMM, respectively, will be the same [see Dovonon and Gonçalves (2017)]. An important difference arises when the set of moment conditions increases to infinity [see Section 2.5]. Then, the number of parameters β in the extended CUGMM tends to infinity too, which is at odds with the results established in the literature on CUGMM with an infinite set of moment conditions and a parameter of fixed dimension.

1.4.4 Regularity Conditions for Bootstrapping the Test Statistic

In the literature, an approach with a finite number of nonlinear moment conditions in u_t has not yet been developed. Inoue and Shintani (2006) consider the block bootstrap with a finite number of moment conditions that are linear in u. Escanciano (2007), Section 3 considers wild bootstrap with an infinite number of moment conditions also linear in u, while Wan and Davis (2022) study a parametric bootstrap with infinite number of nonlinear sine and cosine transformations[7]. The approach proposed in our paper is based on a statistic that is nonlinear in u with a finite number of nonlinear transformations and a finite lag H, and differs from the literature in this respect.

For ease of exposition, let us consider $H = 1$, $dim(u_t) = 1$, $u_t(\theta_0)$ with a symmetric density f_0 satisfying:

$$\sup_u f_0(u)/\exp(b|u|)<\infty, \text{ for } b > 0, \tag{4.2}$$

and a GCov estimator based on a finite number of transformations $a_k(u) = |u|^{p_k} \exp(-t_k|u|), k = 1, ..., K$, where p_k, t_k are fixed. The tail condition (4.2) ensures the uniform integrability of all moments $Ea_k(u_t(\theta_0))$ by considering transformations that reduce the effect of the tail. We also assume $g(\tilde{y}_t; \theta) = g(y_t, ..., y_{t-p}; \theta)$, and we distinguish the following regularity conditions for bootstrap validity:

i) Regularity conditions for deriving the asymptotic distribution of the test statistic under the null hypotheses of i.i.d. errors $u_t = u_t(\theta_0)$.

ii) Conditions concerning the third term in the Edgeworth expansion for the refinement of the bootstrap procedure

iii) Additional regularity conditions to ensure the consistency of the bootstrap approach. Let us now discuss these regularity conditions.

i) Regularity conditions for deriving the asymptotic distribution of the bootstrap-adjusted portmanteau test statistic under the null hypothesis of serial independence.

These conditions have already been discussed in Appendices B.1-B.4. As a consequence

[7] Cavaliere et al. (2020) develop a bootstrap approach for the OLS estimator of autoregressive coefficient ρ in a noncausal AR(1) model: $y_t = \rho y_{t+1} + u_t$, where u_t has a stable distribution. In their framework the coefficient ρ is not equal to $Cov(y_t, y_{t-1})/Var(y_{t-1})$, because the theoretical moments may not exist. Hence, the OLS estimator is not a moment (or covariance) estimator, and their theoretical results cannot be applied to our framework.

of these regularity conditions, we have the asymptotic equivalence of the GCov estimator:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T} m_T(\underline{u_t(\theta_0)}, \theta_0) + o_p(1), \tag{4.3}$$

where $\underline{u_t(\theta_0)} = (u_t, u_{t-1}, ...)$, and $m_T(\underline{u_t}, \theta_0)$ is a vector-valued function satisfying the martingale difference sequence condition $E(m_T(\underline{u_t}, \theta_0)|\underline{u_{t-1}}) = 0$, $E||m_T(\underline{u_t}), \theta_0||^2 < \infty$. An expression of $m_T(u_t(\theta_0), \theta_0)$ can be easily deduced from (B.3). Then, we observe that the empirical process $(\frac{1}{\sqrt{T}} m_T(\underline{u_t(\theta_0)}, \theta_0)$ converges in distribution to a Gaussian process (as a triangular array), and we deduce the asymptotic distribution of the test statistic from its asymptotic expansion.

ii) Conditions for refinement in the third-order Edgeworth expansion

The function m_T in (4.3) is not uniquely defined, since it can always be modified by a term $B(\theta_0)/T$ of order $1/T$. Additional conditions can be introduced, especially on the third-order derivatives of $g(\tilde{y}_t, \theta)$ with respect to θ [see, e.g., Leucht and Neumann (2009)]. Then, we get:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T} m_T(\underline{u_t(\theta_0)}, \theta_0) + o_p(1/T). \tag{4.4}$$

In this expansion, the function m_T depends on the function g and the transforms $a_k, k = 1, ..., K$. It has a closed-form expression, even though it is complicated. In fact, we only need the existence of such an expansion, and the convergence in distribution of the empirical process $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} m_T(\underline{u_t(\theta_0)}, \theta_0)$ to a Gaussian process. This will imply an asymptotically pivotal test statistic at order 1 under (4.3), and at order $1/T$ under (4.4).

iii) Regularity conditions to ensure the consistency of the bootstrap under the null hypothesis of serial independence.

These conditions have to be introduced for the validity of the bootstrap under the null hypothesis $H_{0,iid}$. Below, we describe sets of sufficient conditions introduced in Wan and

Davis (2022) that are specific to this bootstrap procedure[8] and concern functions m_T in the expansion (4.3) (resp.(4.4)). These high level regularity conditions correspond to the assumptions M3, M3', M1' in Wan and Davis (2022), provided below and adapted to our framework of a finite set of nonlinear autocovariance conditions. They are valid for causal models when u_t is a nonlinear innovation of process (y_t) [Gourieroux and Jasiak (2005)][9]

Assumption M3: $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|a_k(\hat{u}_{T,t})-a_k(\hat{u}_{\infty,t})|^m = o_p(1), k=1,...,K,\ m=1,2,$

where $\hat{u}_{\infty,t}$ denotes the fitted residual based on an infinite sequence of observations.

Assumption M3': For any $\epsilon > 0$,

$P^*[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}|a_k(\hat{u}^s_{T,t})-a_k(\hat{u}^s_{\infty,t})|^m > \epsilon] \to 0, k=1,...,K,\ m=1,2,$

when T tends to infinity, where $\hat{u}^s_{\infty,t}$ denotes the bootstrapped residuals based on an infinite sequence of observations.

In comparison with the analogous assumption in Wan and Davis (2022), the condition $u_t(\theta_0) = g(y_t, ..., y_{t-p}; \theta_0)$ is equivalent to $\exp(-t_k u_t(\theta_0)) = \exp(-t_k g(y_t, ..., y_{t-p}; \theta_0))$ for example.

Assumption M1': For any $\epsilon > 0$ and some $\tau > 0$:

$P^*(|\frac{1}{T}\sum_{t=1}^{T}E^*[m'_T(\underline{\hat{u}^s_{T,t}}, \hat{\theta}_T)m_T(\underline{\hat{u}^s_{T,t}}, \hat{\theta}_T)|\underline{\hat{u}^s_{T,t-1}}] - \tau^2| > \epsilon) \xrightarrow{p} 0,$

when $T \to \infty$ and

$P^*(\frac{1}{T}\sum_{t=1}^{T}E^*[m'_T(\underline{\hat{u}^s_{T,t}}, \hat{\theta}_T)m_T(\underline{\hat{u}^s_{T,t}}, \hat{\theta}_T)1_{||m_T(\underline{\hat{u}^s_{T,t}}, \hat{\theta}_T)|| > \sqrt{T}\epsilon}|\underline{\hat{u}^s_{T,t-1}}] > \epsilon) \xrightarrow{p} 0,$

when $T \to \infty$, where P^*, E^* are taken with respect to the bootstrap sampling conditional on the observations $Y_1, ..., Y_T$, and $\xrightarrow{p}$ denotes the convergence in probability.

Assumption M1' ensures that the martingale difference sequence condition is asymptotically satisfied at order 1 by the bootstrapped residuals. Then, we can apply Theorem 4.2 in Wan and Davis (2022) to obtain the consistency of the bootstrap-adjusted GCov test statistic:

[8] See also Escanciano (2007), Assumption A.5 for a sufficient high-level condition for the validity of the bootstrap, when the moment condition is linear in u_t.

[9] They can also be used for pure noncausal processes, i.e., when u_t is a nonlinear innovation in reversed time. In this case, the innovations as well as their bootstrapped values have to be defined in reverse time.

$$\sup_z |P^*[\hat{\xi}_T^s < z] - P[\tilde{\xi}_T < z]| \xrightarrow{p} 0,$$

where P^* stands for conditional on data, and $\tilde{\xi}_T$ is the value of ξ_T adjusted for its first-order bias.

The local power analysis in Section 4.2 is based on Escanciano (2007), Section 3, Theorem 6, where a two-step parametric bootstrap is considered. The validity of the second step bootstrap is established for the wild bootstrap and extended to parametric bootstrap with conventional drawing on page 126.

1.5 Increasing the Set of Autocovariance Conditions

The GCov estimator has been applied so far using a finite set of zero nonlinear autocovariance conditions, that is, a given order H and a given set of K nonlinear error transformations. In this section, we discuss how the asymptotic performance of the GCov specification test can be improved by increasing the set of nonlinear transformations denoted by $\mathcal{A}$ and possibly by taking into account an infinite set of autocovariance conditions.

1.5.1 Implicit Null and Alternative Hypotheses

Let us compare the implicit null and alternative hypotheses considered with those of potential interest. We distinguish the following hypotheses concerning the parameter θ and the distribution f of the process (u_t)[10]:

i) $H_{0,\mathcal{A}} = \{\theta, f : Cov[a(u_t), \tilde{a}(u_{t-h})] = 0, \ \forall h, \ \forall a, \tilde{a} \in \mathcal{A}\}$ and the associated alternative $H_{1,\mathcal{A}}$. These hypotheses depend on the selected set of scalar transformations $\mathcal{A}$.

ii) $H_{0,pair} = \{\theta, f : u_t, \text{ and } u_{t-h} \text{ are independent}, \ \forall h\}$ and its alternative $H_{1,pair}$.

iii) $H_{0,ind} = \{\theta, f : u_t, u_{t-1}, ..., u_{t-h} \text{ are independent } \forall t, h\}$ and its alternative $H_{1,ind}$.

[10]As in previous sections, we assume a well-specified model (3.1): $u_t = g(\tilde{Y}_t, \theta)$ and the true value θ_0 of θ. To simplify the notation, the same symbol u_t is used for errors that depend on either θ or θ_0.

iv) $H_{0,iid} = \{\theta, f : u_t's \text{ are iid}\}$ and its alternative $H_{1,iid}$.

We have $H_{0,\mathcal{A}} \supset H_{0,pair} \supset H_{0,ind} \supset H_{0,iid}$ and $H_{1,\mathcal{A}} \subset H_{1,pair} \subset H_{1,ind} \subset H_{1,iid}$.

The test of $H_{0,\mathcal{A}}$ is consistent and of asymptotic power of 1 against $H_{1,\mathcal{A}}$. However, this test is not of asymptotic power 1 against $H_{1,pair}, H_{1,ind}$ and $H_{1,iid}$. By increasing the set of nonlinear transformations a in $\mathcal{A}$ appropriately, we expect to improve the asymptotic efficiency of the GCov estimator and to increase the set of alternatives against which the GCov specification test is asymptotically consistent.

1.5.2 The Semi-Parametric Efficiency Bound

The asymptotic performance of the GCov specification test depends on the asymptotic properties of the GCov estimator of parameter θ. We can distinguish the semi-parametric models and the associated semi-parametric efficiency bounds, which are:

i) The semi-parametric efficiency bound that accounts for the pairwise independence between u_t and u_{t-h} for any h.

ii) The semi-parametric efficiency bound that takes into account the joint independence of $u_t, u_{t-1}, ..., u_{t-h}$, since the pairwise independence does not imply the joint independence.

iii) The semi-parametric efficiency bound that takes into account the fact that the u_t's are identically distributed. It coincides with the parametric efficiency bound, as shown in Example 6.

The GCov estimator attains the first semi-parametric efficiency bound i), but not the parametric efficiency bound under $H_{0,iid}$, which requires a two-step estimator illustrated below.

Example 6: Adaptive Estimator

Let $\hat{\theta}_T$ denote a GCov estimator based on a finite set $\mathcal{A}$ of nonlinear transformations. Given this consistent estimator, we can compute the residuals $\hat{u}_{T,t}, t = 1, ..., T$ of the model and then approximate nonparametrically the true density f_0 of u_t by the kernel-based density $\hat{f}_T(u)$ based on the residuals $\hat{u}_{T,t}, t = 1, ..., T$. Next, in the second step, parameter θ can be estimated by a pseudo-maximum likelihood method, where the true density $f_0(u)$ is replaced by $\hat{f}_T(u)$ [see e.g., Bickel (1982) and Newey (1988)]. Under standard regularity conditions, this leads to a two-step estimator of θ that reaches the parametric efficiency bound. Due to this improvement of the asymptotic properties of the GCov estimator, the pseudo-likelihood ratio test based on this two-step estimator has better asymptotic power properties than the GCov test based on a finite set $\mathcal{A}$.

1.5.3 How to Choose an Infinite Set of Transformations

The asymptotic properties of the GCov estimator and of the associated test statistics can be improved by increasing the finite set of nonlinear autocovariance conditions to a larger finite or infinite set. The extension to an infinite set is easy when these conditions correspond to an orthonormal basis of the Hilbert space $L^2(u_t, u_{t-1}, ...)$. In our framework, we can increase the set of nonlinear autocovariance conditions by increasing the maximum lag H, or the set of transformations $\mathcal{A}$. We saw that, under the null hypothesis, the orthogonality of nonlinear autocovariance conditions with respect to the lag is satisfied. This explains the simplified form of the test statistic written as a sum of terms associated with lags $h = 1, ..., H$. In contrast, the set of nonlinear transformations does not necessarily correspond to an orthonormal basis, and inverting the variance matrix of a large dimension can become a problem from both the theoretical and computational perspectives.

We follow the approach of Bierens (1990) and build a sequence of orthonormal bases in $L^2(u_t, u_{t-1}, ..., u_{t-H})$ of square-integrable transformations of $(u_t, u_{t-1}, ..., u_{t-H})$ (or, equivalently in $L^2(u_t)$). We proceed as follows:

1. Consider a countable subset of transformations $\mathcal{A}$, called a system of generators, that allows identifying the unknown distribution f_0 of u_t;

2. Select from that set $\mathcal{A}$ an increasing sequence of finite subsets of transformations $\mathcal{A}_n$, such that $\bigcup_n \mathcal{A}_n$ is dense;

3. Orthonormalize within each subset;

(See Bierens (1990), Corollary 1 for the system of generators based on exponential transforms). In practice, it is difficult to find a system of generators that is informative about θ. An important feature of the causal-noncausal models considered in our paper is the presence of extreme risks and persistence, creating locally explosive patterns, including spikes and bubbles. This implies that the model errors do not necessarily have higher power moments. Moreover, some parameters driving the tails of their distributions, including those distinguishing between the negative and positive tails, may be non-identifiable from the transformations $u_t u_t'$ only.

In this respect, some standard linear systems of generators are not convenient. For example, the polynomial transformations of u_t cannot be used if u_t has no moments of order greater than 3. Moreover, the sine-cosine transformations used in the ADCV literature may not be informative of the tail driving parameters [see, e.g., Wan and Davis (2022), Fokianos and Pitsillou (2018) for the ADCV test of the independence hypothesis based on a joint characteristic function]. The same remark applies to the standard spline bases.

A preferred system of generators would assign weights to the power transformations, ensuring their square integrability. For ease of exposition, consider a univariate process with positive errors u_t and with trajectories admitting only positive and increasing bubbles caused by a large positive shock to u_t[11]. Then, a countable linear system of generators is:

$\mathcal{A} = \{a_{t,p}(u) = u^p \exp(-tu),\ p \in \mathbb{N},\ t \in \mathcal{Q} \subset [0,1]\},$

where $\mathcal{Q}$ is the set of rational numbers, with an increasing sequence of finite subsets given by: $\mathcal{A}_n = \{a_{t_j,n,p}(u) = u^p \exp(-t_{j,n}u),\ p \in \mathbb{N},\ t_{j,n} \in \mathcal{Q},\ j = 1, ...n\}$,

where $(t_{1,n}, ..., t_{j,n})$ are the elements of a grid that is getting finer as n increases.

We observe that the sequence $\{\mathcal{A}_n\}$ is such that $\bigcup_n \mathcal{A}_n$ is dense in L^2 and allows identifying the distribution of u, based on the inversion formula of the real Laplace transform called Post's inversion formula [Post (1930), and Feller (1991), Chapter 13, for the modern proof]. We make the following assumption:

[11]The extension to a multivariate u_t is done by finding the products of generators for univariate u_t's.

Assumption 1: U is positive and has a distribution with continuous density $f_0(u)$ such that: $\sup_{u>0} f_0(u)/\exp(bu) < \infty$, for some $b > 0$.

This condition on the density function implies that the distribution cannot have a large probability mass at zero, and its right tail can be of any size.

Under Assumption 1, the Laplace transform: $\Psi(t) = E[\exp(-tU)]$, $t \in [0,1]$ characterizes the distribution of U. However, as mentioned earlier $\mathcal{A}^* = \{a_t(u) = \exp(-tu),\ t \in [0,1]\}$ is not a convenient system of generators, because we may not be able to write f_0 as the limit of linear combinations of decreasing exponentials. A better system of generators follows from the Post's inversion formula.

Proposition 4: Post's inversion formula [Post (1930)]

Under Assumption 1, we have:

$$f_0(v) = \lim_{n\to\infty} \frac{1}{n!}\left(\frac{n}{v}\right)^{n+1} E[U^n \exp\left(-\frac{n}{v}U\right)],\ \forall v,\ a.e.$$

This suggests the linear system of generators:

$$\mathcal{A} = \{a_{t,p}(u) = u^p \exp(-tu),\ p \in \mathbb{N}, t \in [0,1]\},$$

where the power transforms are weighted by decreasing exponentials to ensure their square integrability, and the associated sequence given by:

$$\mathcal{A}_n = \{a_{t_{j,n},p}(u) = u^p \exp(-t_{j,n}u),\ p \in \{1,...,n\}, t_{j,n} \in \mathcal{Q} \subset [0,1],\ j = 1,...,n\}.$$

In practice, only weights with t close to 0 are useful. To see that, let us recall that the real Laplace transform Ψ of a positive variable is characterized by its Taylor series expansion:

$$E[\exp(-tU)] = \sum_{j=0}^{\infty} \frac{t^j}{j!}\mu_j.$$

If U admits power moments of any order, we have $\mu_j = E(U^j), \ \forall j$. In our framework of heavy-tailed U, such moments may not exist, but regardless, this series expansion exists and Taylor's coefficients μ_3, μ_4 define new notions of skewness and kurtosis measures.

Remark 5: In practice, the Post's inversion formula does not provide a tractable means of inverting the Laplace transform, and it is an ill-posed problem [see Bryan (2006), Conclusion]. In this respect, the regularized GCov estimator introduced later in this section can be seen as a solution to this ill-posed problem when the goal is to estimate θ. We use Post's inversion formula only to show that $\mathcal{A}$ is an adequate linear system of generators.

Remark 6: Similar results can be obtained if additional information is available on the size of the tail of the distribution of U, like, for example, the tails are Pareto. Then, we can apply the Hardy-Littlewood Tauberian theorem with a Pareto weighting function to define an alternative system of generators [Feller (1991)].

Remark 7: The causal-noncausal MAR(r,s) processes are often applied to positive time series, such as commodity prices. Then, their solutions have a two-sided moving average representation $y_t = \sum_{j=-\infty}^{\infty} c_j u_{t-j}$ in errors taking values in $(0, \infty)$. The positivity restriction on process (y_t) implies that $c_j \geq 0, \forall j$, and $u_t \geq 0, \forall t$ (up to sign identification), validating the assumption $U > 0$. If the error u_t takes both positive and negative values, we can observe "positive" and "negative" bubbles. Then, the weights have to be replaced by $\exp(-t|U|)$, and the even and odd powers of U have to take positive or negative signs, respectively. If u_t has a symmetric distribution, the generators become $a_{t,p} = |u|^p \exp(-t|u|)$.

1.5.4 Regularized Orthonormalization of the System of Generators

It is impossible to construct an exact orthonormal basis from a system of generators because the true distribution of U under the i.i.d. hypothesis is unknown. Hence, a two-step approach is required. We derive below the orthonormalization in the multidimensional case and discuss the form of the portmanteau statistic and its asymptotic distribution in the following subsection. The empirical orthonormalization in the spirit of the Gram-Schmidt forward regression [Chen et al. (2025)] proceeds in the following steps:

step 1. Consider a finite set $\mathcal{A}_0$ of transformations from which θ is identifiable. Estimate θ by a consistent estimator $\tilde{\theta}_{n,T}$ and compute $\hat{u}_{T,t},\ t = 1, ..., T$.

step 2. Consider an infinite set $\mathcal{A} = \{a(u) = u_1^{p_1}, ..., u_J^{p_J} \exp(-\tau' u),\ p_1, ..., p_J \in \mathcal{N},\ \tau \in [0,1]^J\}$ and the increasing sequence of finite sets $\mathcal{A}_n$:

$\mathcal{A}_n = \{a(u) = u_1^{p_1}, ..., u_J^{p_J} \exp(-\tau'_{j,n} u),\ p_1, ..., p_J \in (0, 1, ..., P_n),\ \tau_{1,n}, ..., \tau_{n,n} \in [0,1]^J\}$,

where $\bigcup_n (\tau_{1,n}, ..., \tau_{n,n})$ is dense in $[0,1]^J$ when $n \to \infty$.

The number of transformations in $\mathcal{A}_n$ is $P_n^J J^n$, where $K_n \equiv P_n^J J^n$ with the elements $a_{k,n}, k = 1, ..., K_n$ arranged in an increasing order: $\mathcal{A}_n = \{a_{k,n},\ k = 1, ..., K_n = P_n^J J^n\}$.

step 3. (Regularized) Orthonormalization

The mapping of $a_{k,n},\ k = 1, ..., P_n^J J^n$ into an orthonormal basis $a^*_{k,n},\ k = 1, ..., K_n$ is obtained as follows. We run regressions to get at step n the orthonormal functions $a^*_{k,n},\ k = 1, ..., K_n$, with zero mean.

a) We start from regressing $a_{1,n}(\hat{u}_{T,t})$ on the constant:

$$a_{1,n}(\hat{u}_{T,t}) = \alpha_{1,n,T} + \hat{w}_{1,n,T,t},$$

$t = 1, ..., T$ with the residuals $\hat{w}_{1,n,T,t}$. Next, compute:

$$a^*_{1,n,T}(u) = \hat{w}_{1,n,T}(u)/||\hat{w}_{1,n,T}||_T,$$

where $\hat{w}_{1,n,T}(u) = a_{1,n}(u) - \alpha_{1,n,T}$. Let $R^2_{1,n,T}$ denote the R-square in the above regression. Then $a^*_{1,n,T}$ is used if $1 - R^2_{1,n,T}$ is sufficiently different from 0, i.e., $1 - R^2_{1,n,T} > \epsilon_{n,T}$, where $\epsilon_{n,T}$ is an appropriately chosen regularization tuning parameter [The regularization through tuning parameters is an empirical analogue of the theoretical condition in the proof of Corollary 1 in Bierens (1990)]. Otherwise, disregard $a_{1,n}$ and start from $a_{2,n}$.

b) In the next step, $a_{2,n}$ is projected on $a^*_{1,n,T}$ as follows. We run the regression:

$$a_{2,n}(\hat{u}_{T,t}) = \alpha_{2,n,T} + \beta_{2,n,T} a^*_{1,n,T}(\hat{u}_{T,t}) + \hat{w}_{2,n,T,t},$$

$t = 1, ..., T$ with the residuals $\hat{w}_{2,n,T,t}$. Next, compute:

$$a^*_{2,n,T}(u) = \hat{w}_{2,n,T}(u)/||\hat{w}_{2,n,T}||_T,$$

where $\hat{w}_{2,n,T}(u) = a_{2,n}(u) - \alpha_{2,n,T} - \beta_{2,n,T} a^*_{1,n,T}(u)$. This can be done if $||\hat{w}_{2,n,T}||_T$ is sufficiently different from 0, which is the case when $1 - R^2_{2,n,T} > \epsilon_{n,T}$. Otherwise, proceed with $a_{3,n}$ instead.

c) The third step is

$$a_{3,n}(\hat{u}_{T,t}) = \alpha_{3,n,T} + \beta_{3,1,n,T} a^*_{1,n,T}(\hat{u}_{T,t}) + \beta_{3,2,n,T} a^*_{2,n,T}(\hat{u}_{T,t}) + \hat{w}_{3,n,T,t},$$

and so on, up to $k = K_n$.

We end up with a selected set of transformations $\mathcal{A}^*_n = \{a^*_{k,n}, k = 1, ..., K^*_n\}$, which are zero mean and orthonormal with respect to the sample distribution of $\hat{u}_{T,t},\ t = 1, ..., T$, and a random K^*_n, such that $K^*_n \leq K_n$. Since the autocovariance conditions concern pairs u_t, u_{t-h} and transformations of type $a(u_t), \tilde{a}(u_{t-h})$, we need to consider transformations in $\mathcal{A}^2_n$, where $dim(\mathcal{A}^2_n) = P_n^{2J} J^{2n} = K^2_n$.

The above regularized orthonormalization depends on the estimator $\tilde{\theta}_{n,T}$ of θ selected in the first step. In practice, we can consider either: a) $\tilde{\theta}_{n,T} = \hat{\theta}_{n,T}$ equal to the GCov estimator, or b) $\tilde{\theta}_{n,T}$, which is a Diagonal GCov estimator with $\hat{\Gamma}(0)$ replaced by a matrix containing only the diagonal elements of $\hat{\Gamma}(0)$ in the objective function [Gourieroux and Jasiak (2017)]. For a large number of transformations, the Diagonal GCov estimator can be easily computed, while the GCov estimator becomes computationally challenging.

1.5.5 Two-step Portmanteau Statistic and its Asymptotic Distribution

The orthonormalization simplifies the expression of the objective function and/or facilitates the derivation of the asymptotic distribution of the GCov specification test statistics. The orthonormalization approach outlined in subsection 2.5.4 provides an identity weighting matrix, so that the resulting objective function has the following expression:

$$L_{n,T}(\theta) = \sum_{h=1}^{H} \sum_{j=1}^{K_n} \sum_{k=1}^{K_n} \left(\frac{1}{T} \sum_{t=1+h}^{T} a^*_{j,n,T}[g(\tilde{y}_t;\theta)] a^*_{k,n,T}[g(\tilde{y}_{t-h};\theta)] \right)^2, \tag{5.1}$$

and $\hat{\xi}_{n,T} = T L_{n,T}(\theta)$, since the inverse of the variance-covariance matrix is equal to the identity matrix. Note that the transformations $a^*_{j,n,T}$ depend on the starting values $\tilde{\theta}_{n,T}$.

Below, we examine two types of GCov test statistics, depending on the initial value $\hat{\theta}_{n,T}$, or $\tilde{\theta}_{n,T}$ used in the orthonormalization algorithm. Then, the GCov test statistics based on the minimizers of (5.1) are:

$$\hat{\xi}_{n,T}(Argmin_\theta \, L_{n,T}(\theta|\hat{\theta}_{n,T})) \equiv \xi_{n,T}(\hat{\hat{\theta}}_{n,T}|\hat{\theta}_{n,T}) \text{ and } \tilde{\xi}_{n,T}(Argmin_\theta \, L_{n,T}(\theta|\tilde{\theta}_{n,T})) \equiv \xi_{n,T}(\tilde{\tilde{\theta}}_{n,T}|\tilde{\theta}_{n,T}),$$

where $\hat{\hat{\theta}}_{n,T}$ and $\tilde{\tilde{\theta}}_{n,T}$ denote the minimizers of the objective function (5.1). Note that $\hat{\hat{\theta}}_{n,T}$ can be computationally challenging for large K and different from $\tilde{\tilde{\theta}}_{n,T}$, although they are both asymptotically equivalent.

Proposition 5

Under $H_{0,iid}$ and the regularity conditions given in Appendix B.8, we have:

$$(\hat{\xi}_{n,T} - HK_n^{*2})/\sqrt{2HK_n^{*2}} \xrightarrow{d} N(0,1) \text{ and } (\tilde{\xi}_{n,T} - HK_n^{*2})/\sqrt{2HK_n^{*2}} \xrightarrow{d} N(0,1),$$

when T and K_n tend to infinity, K_n^6/T tends to 0, and the regularization tuning parameter tends to 0 at a rate such that there exist δ_1, δ_2, $0 < \delta_1 \leq \delta_2 < 1$ with $\delta_1 < EK_n^*/K_n < \delta_2$.

Proof: We consider below the GCov statistic $\tilde{\xi}_{n,T}$ based on $\tilde{\theta}_{n,T}$. The asymptotic behavior of this statistic is derived along the following steps:

step 1: We consider the infeasible portmanteau test statistic:

$$\xi^0_{n,T} = T \sum_{h=1}^{H} \sum_{j=1}^{K_n} \sum_{k=1}^{K_n} \left(\frac{1}{T} \sum_{t=1+h}^{T} a^*_{j,n}[g(\tilde{y}_t;\theta_0)] a^*_{k,n}[g(\tilde{y}_{t-h};\theta_0)] \right)^2,$$

where $a^*_{k,n}$ are the transformations obtained from orthonormalization performed given the true distribution f_0. The analysis of the asymptotic properties of this theoretical infeasible statistic needs to take into account the increasing sequence $\mathcal{A}_n$ and the theoretical regularization used in the orthonormalization.

step 2: We take into account the uncertainties due to the estimation of θ_0 and f_0, and the replacement of θ_0 by the starting value $\tilde{\theta}_{n,T}$.

Step 1 Regularity Conditions

These regularity conditions concern the theoretical orthonormal bases of $\mathcal{A}_n$. i.e., $a^*_{j,n}, j = 1, ..., n$. Let us assume that u_t has non-negative components, and a continuous density $f_0(u)$ that satisfies a tail condition:

$$\sup_{u>0} f_0(u)/\exp(b'u) < \infty, \text{ for some } b,$$

with strictly positive components, ensuring that the transformations in $\mathcal{A}_n$ are uniformly integrable. Then, under the hypothesis $H_{0,iid}$ of i.i.d. $u_t = g(\tilde{y}_t, \theta_0)$, for any fixed K_n, the multidimensional Lindeberg-Feller condition for the convergence in distribution of:

$$\frac{1}{T} \sum_{t=1+h}^{T} a^*_{j,n}[u_t] a^*_{k,n}[u_{t-h}], \ j,k = 1, ..., K_n, \ h = 1, ..., H.$$

to a standard normal $N(0, Id_{K_n^2 H})$ is satisfied. Then, under the assumption of an increasing sequence $\mathcal{A}_n$ with a dense $\bigcup_n \boldsymbol{A}_n$ and a tightness condition [see Koenker and Machado (1999)], we get the analogue of Proposition 5 for the theoretical infeasible portmanteau test statistic:

Proposition 5' [Koenker and Machado (1999)]

$$(\xi^0_{n,T} - HK_n^2)/\sqrt{2HK_n^2} \xrightarrow{d} N(0,1),$$

when T and K_n tend to infinity.

Step 2 Regularity Conditions

The two-step estimator-based statistic $\hat{\xi}_{n,T} = \xi_{n,T}(\tilde{\tilde{\theta}}_{n,T})$ has the standard normal asymptotic distribution provided that: a) the first-step estimator $\tilde{\theta}_{n,T}$ is consistent, asymptotically normally distributed; b) the regularization tuning parameter $\epsilon_{n,T}$ tends to zero at an adequate rate; c) T and K_n tend to infinity at a rate such that K_n^2/T tends to zero.
Under an additional tightness condition, the difference between $a^*_{j,n,T}$ and $a^*_{j,n}, j = 1, ..., K_n$ is asymptotically negligible when K_n tends to infinity.

These conditions ensure that the uncertainty due to the approximation of θ_0 by $\tilde{\theta}_{n,T}$ and of the true orthonormalization by the estimated one are negligible [see e.g., Dovonon and Gospodinov (2024) for a similar argument].

The conditions on the number of transformations are sufficient conditions for the asymptotic normality [see Koenker and Machado (1999)]. They imply that K_n^* has to increase with T, but at a rate slow enough to allow for solving the ill-posed problem due to inversion of the sample autocovariance matrix at lag 0.

The change of asymptotic distribution from $\chi^2(K^2H - dim(\theta))$ for a fixed set of transformations to the normal distribution given above is easy to explain. If the dimension of the parameter is fixed and K_n^* (and/or H_n) tend to infinity when T increases to infinity, the χ^2 distribution with an infinite number of degrees of freedom is well approximated by a Gaussian distribution. Hence, Proposition 5 applies this approximation to a centered and standardized test statistic.

As mentioned earlier, our discussion is focused on the increase of the set $\mathcal{A}_n$ of transformations. It would also be possible to increase the maximum lag H with the number of observations and get a similar result when T, H, K_n tend to infinity at a rate such that

$(HK_n^2)/T$ tends to zero.

The asymptotic results given above are valid for any choice of sequence $\{\mathcal{A}_n\}$ such that $\bigcup_n \mathcal{A}_n$ is dense and identifies the true distribution f_0, and for any ordering of transformations in $\mathcal{A}_n$. However, the choice of sequence $\mathcal{A}_n$ and transformation ordering will have an effect at higher orders and in finite samples.

1.6 Finite Sample Performance of NLSD, GCov Specification and Bootstrap Tests

This section examines the finite sample performance of proposed test statistics in selected causal-noncausal autoregressive processes. We perform simulations to study the size and power of a) the NLSD test (eq. 2.5) to test for the absence of (non)linear serial dependence in time series, b) the GCov specification test (eq. 3.3) in Section 2.6.1, and c) GCov bootstrap test [Section 2.4.1] in Section 2.6.2. Section 2.6.3 shows an example of the GCov specification test with many nonlinear transformations.

1.6.1 Simulation Study

We consider univariate causal-noncausal MAR(r,s) processes with i.i.d. errors from a uniform distribution $\mathcal{U}_{[-1,1]}$, a Laplace distribution with mean zero and variance one, and a t-student distribution with ν degrees of freedom, called t(ν), with mean zero and variance $\nu/(\nu-2)$, $\nu > 2$. We generate these processes in samples of size $T = 100, 200, 500$.

1.6.1.1 Data Generating Process

The simulation method introduced in Gourieroux and Jasiak (2016) is used to generate the MAR(r,s) process (eq. 3.6) where r and s denote the orders of causal and noncausal polynomials, respectively. For $r = 0$ and $s = 1$, we generate MAR(0,1), i.e., the noncausal autoregressive process of order 1:

$$y_t = \psi y_{t+1} + u_t \quad, |\psi| < 1. \tag{6.1}$$

By setting $r = 1$ and $s = 1$, we generate MAR(1,1) processes defined in equation (3.7). It follows from Lanne and Saikkonen (2011) that its unobserved components $v_{1,t}, v_{2,t}$ are defined by:

$$v_{1,t} \equiv (1 - \phi L)y_t \leftrightarrow (1 - \psi L^{-1})v_{1,t} = u_t, \quad v_{2,t} \equiv (1 - \psi L^{-1})y_t \leftrightarrow (1 - \phi L)v_{2,t} = u_t, \tag{6.2}$$

and can be interpreted as the "causal" and "noncausal" components. Then, the MAR(1,1) process has the following deterministic representations based on these unobserved components, used for the simulation and bootstrapping of y_t:

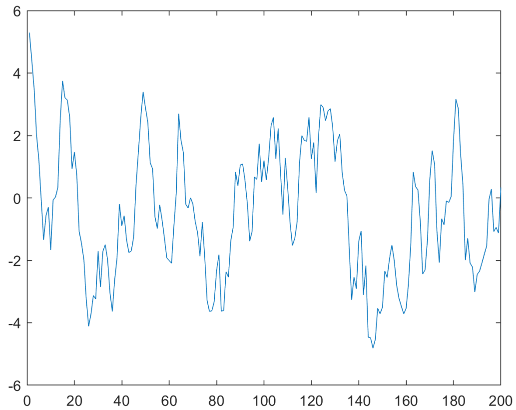
$$i)\, y_t = \frac{1}{1 - \phi\psi}(\phi v_{2,t-1} + v_{1,t}), \quad ii)\, y_t = \frac{1}{1 - \phi\psi}(v_{2,t} + \psi v_{1,t+1}), \tag{6.3}$$

where in i) y_t is a linear function of the first lag of $v_{2,t}$ and of the current value of $v_{1,t}$, and in ii) y_t is a linear function of the current value of $v_{2,t}$ and of the first lag of $v_{1,t}$.

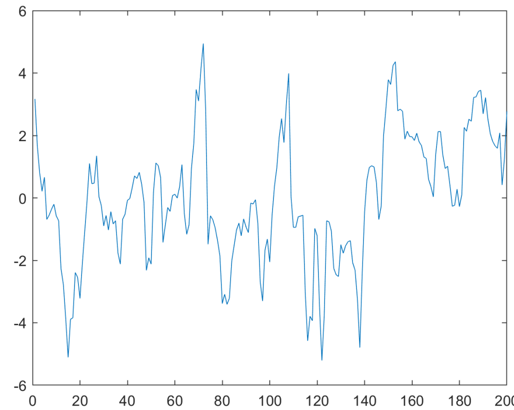
Figure 1.1 shows examples of trajectories of the MAR(1,1) processes simulated with the three error distributions given above and the coefficients $\phi = 0.2$ and $\psi = 0.8$ satisfying the strict stationarity condition. We observe that a large error value creates a spike in the trajectory of MAR(1,1) with explosion and collapse rates determined by coefficients ψ and ϕ, respectively.

1.6.1.2 Nonlinear Transformations

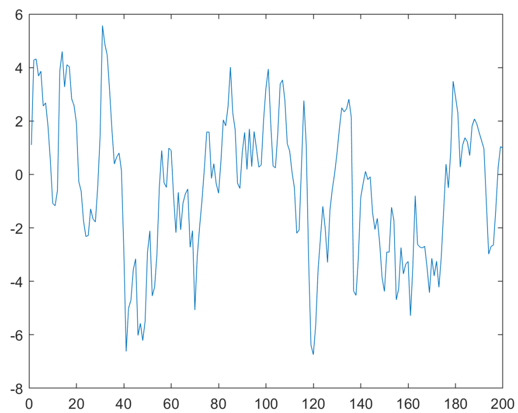
In practice, we choose a finite set of nonlinear autocovariances, depending on the dynamics of the process. For example, we consider the sample autocovariances to capture linear serial dependence in the data, or residuals. To account for nonlinear serial dependence, we use the



(a) $u_t \sim U$



(b) $u_t \sim L$



(c) $u_t \sim t(5)$

Figure 1.1: MAR(1,1) processes with $\phi = 0.2$ and $\psi = 0.8$, and with error densities: L.:Laplace, U.:Uniform, t(5): t-Student with d.f.=5, over a period of $T = 200$

nonlinear autocovariances, which can be selected based on graphically displayed nonlinear autocorrelograms [Gourieroux and Jasiak (2002)]. This approach provides additional insights on the choice of H. A priori, we can set $H \geq p$, where p is the lag length of the model. Specifically, processes with local explosive patterns and/or time-varying volatility are characterized by conditional heteroscedasticity, which can be captured by the autocovariances of squared observations. The third powers can accommodate skewed distributions, common in financial data. Logarithms are useful for capturing nonlinearities and transforming positive price series into variables that take both positive and negative values. For models with Cauchy distributed errors, we may employ fractional transformations, such as the square root of absolute value, since the unconditional expectation of a fractional power of a Cauchy distributed variable exists, while the unconditional expectation of an integer order does not. Section 2.5 shows that combinations of power transformations weighted by exponential functions can also be used and that by increasing the set of associated autocovariance conditions, one can improve the performance of the GCov estimator and test.

1.6.1.3 NLSD Test for Time Series

This section examines the finite-sample size and power of the (non)linear serial dependence (NLSD) test of y_t computed from transformations a of a univariate time series y_t. The null hypothesis: $H_{0,a} = \{\Gamma_{0,a}(h) = 0, \ \forall h = 1, ..., H\}$, is tested against the fixed alternative hypothesis: $H_{1,a} = \{\exists h : \text{such that } \ \Gamma_{0,a}(h) \neq 0\}$. Under this fixed alternative, the process is a MAR(0,1) model with an autoregressive coefficient denoted by γ that varies between 0.1 and 0.9. Thus, the alternative can be written in terms of the parameters as: $H_{1,a} = \{\gamma, f : \Gamma(h; \gamma, f) = 0\}$. For each value of γ and each error density f_0, we simulate the strong white noise series $y_1^s, ..., y_T^s, s = 1, ..., S$ with S=5000 replications and compute the NLSD test statistics, based on $K = 2$ transformations of time series: y_t, y_t^2, and lag $H = 1$. The results are reported in Table 1.1, for the nominal size of the test of 0.05. The first row of Table 1.1 with the zero values of coefficients γ provides the rejection rates corresponding to the size. The remaining rows illustrate the size-adjusted[12] power of the test against the fixed alternative of a MAR(0,1) process with coefficients $\gamma = 0.3, 0.7$. The columns of Table 1.1 show the outcomes for different sample sizes and the Uniform, Laplace, and t(5) error distributions.

The results reported in Table 1.1 show close to the nominal empirical size and good power of the test against fixed alternatives, given each error density. When the sample size increases, the size converges to 0.05 and the power converges to 1. We also observe that the power of the test increases in the values of the autoregressive coefficient ψ of MAR(0,1).

[12] The size adjustment is based on the 95th percentile of the sample distribution of simulated test statistic.

Table 1.1: NLSD Test of the absence of (non)linear dependence against the fixed alternative of MAR(0,1) at 5% significance level: empirical size and power

γ	T=100			T=200			T=500		
	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)
0.0	0.0414	0.0484	0.0512	0.0450	0.0502	0.0518	0.0496	0.0540	0.0480
0.3	0.6014	0.5742	0.5674	0.9202	0.9266	0.9256	1	0.9998	1
0.7	1	1	1	1	1	1	1	1	1

The first row ($\gamma = 0$) shows the size of the test and the remaining rows show the size-adjusted power against fixed alternatives.

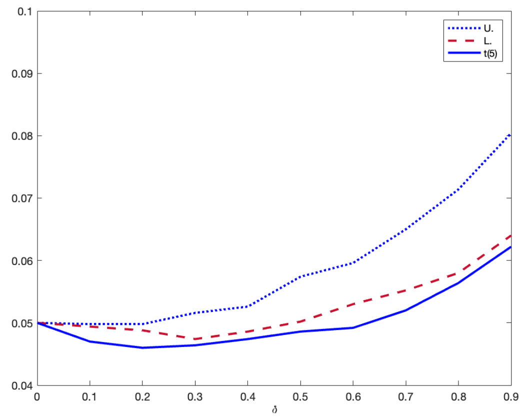
Furthermore, we investigate the power under the local alternatives by generating MAR(0,1) models with autoregressive coefficients equal to $\frac{\delta}{\sqrt{T}}$, where δ varies between 0 and 0.9. The results are provided in Figure 1.2. Since we consider local alternatives, we expect asymptotically the powers to be close to the size for small δ, while for bigger δ, we deviate further away.

We observe that the test has good local power, and the size-adjusted power functions in the neighborhood of the null hypothesis increase fast and nonlinearly in δ. The local asymptotic properties depend on the sample size and the error distribution. When the sample size increases from 100 to 500, we observe a greater improvement for the Uniform and t(5) distributions compared to the Laplace distribution, which has the least heavy tail among these three distributions.

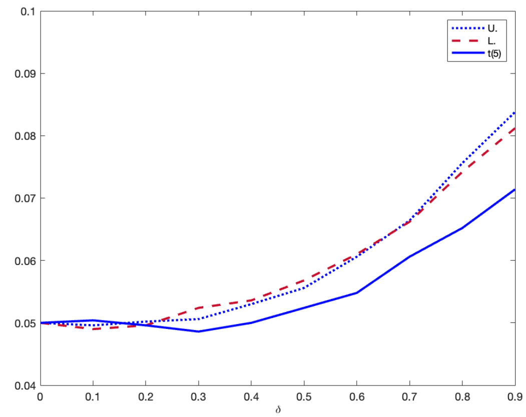
1.6.1.4 GCov Specification Test for Semi-Parametric Models

This section examines the finite-sample size and power of the GCov specification test applied to nonlinear transforms of the residuals of a model. The true model is a MAR(0,1) with errors $u_t = (1 - \psi L^{-1})y_t$. It is estimated by the GCov estimator with lag length H=3. We consider K=2 with the residuals $\hat{u}_{T,t}$ and their squares $\hat{u}_{T,t}^2$ as non-linear transformations, where $\hat{u}_{T,t} = (1 - \hat{\psi}_T L^{-1})y_t$, for the MAR(0,1) process.

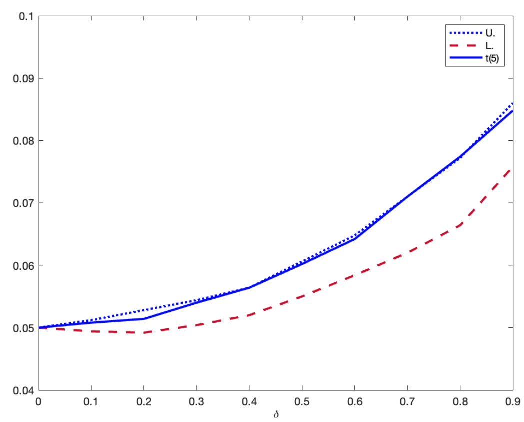
To study the size, we compute the test statistics from the autocovariance matrices of nonlinear transformations of the residuals, denoted by $\Gamma_a(h; ., f)$. The null hypothesis of



(a) $T = 100$



(b) $T = 200$



(c) $T = 500$

Figure 1.2: Local asymptotic power of the NLSD test.

"correct specification" is: $H_{0,a} = \{\psi, f : \Gamma_{0,a}(h; \psi, f) = 0, \ \forall h = 1, ..., H\}$, and it is tested against $H_{0,a} = \{\psi, f : \exists h : \text{such that } \ \Gamma_{0,a}(h; \psi, f) \neq 0\}$.

We generate the noncausal MAR(0,1) processes with the autoregressive coefficients ψ equal to 0.3 and 0.7, using S=5000 replications. The results are reported in Table 1.2 for the nominal size of 0.05. The top two rows of Table 1.2 illustrate the rejection rates corresponding to the empirical size of the GCov specification test applied to the MAR(0,1) model, and the columns display the results for the Uniform, Laplace, and t(5) error distributions and different sample sizes. Table 1.2 shows that the GCov specification test is conservative at $T = 100$ in all cases. When the sample size increases, the size approaches the nominal size

Table 1.2: Empirical size and power of GCov specification test of MAR(0,1) at 5% significance

S./P.	ϕ	ψ	T=100			T=200			T=500		
			Uniform	Laplace	t(5)	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)
S.	0	0.3	0.0224	0.0454	0.0386	0.0348	0.0566	0.0534	0.0406	0.0544	0.0560
	0	0.7	0.0200	0.0400	0.0338	0.0298	0.0528	0.0468	0.0408	0.0552	0.0528
P.	0.8	0.3	0.1016	0.1724	0.1788	0.3404	0.4468	0.4672	0.9092	0.9282	0.9300
	0.8	0.7	0.9680	0.9882	0.9896	1	1	1	1	1	1

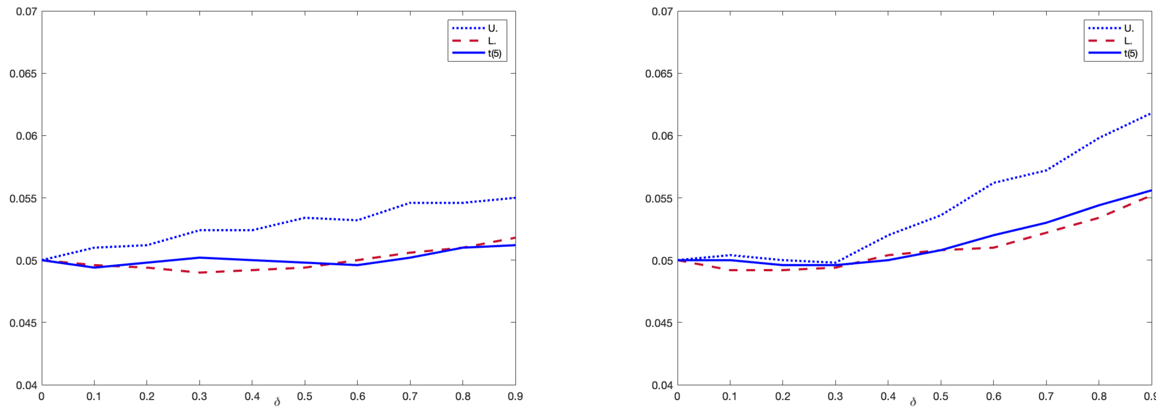
S.: size, P.: power (size-adjusted)

for all distributions. The power increases in ψ, similar to the pattern in Table 2.1, and quickly converges to one.

An interesting way to investigate the power of the GCov specification test is to consider the null hypothesis of MAR(0,1) and deviate from it by adding a causal autoregressive coefficient $\gamma = \phi$. This transforms the MAR(0,1) model under the null hypothesis into a MAR(1,1) model under the alternative. Then, the fixed alternative hypothesis is: $H_{1,a} = \{\psi, \gamma, f : \exists h : \text{such that } \Gamma_{0,a}(h; \psi, \gamma, f) \neq 0\}$.

Rows 3 and 4 of Table 1.2 report the size-adjusted power of the test applied to a noncausal MAR(0,1) model with coefficients $\psi = 0.3$ and $\psi = 0.7$, when an additional causal coefficient $\gamma = \phi = 0.8$ of MAR(1,1) is present under the fixed alternative, for each given error density and sample size reported in the columns. By comparing the results for $\psi = 0.7$ and $\psi = 0.3$, we can conclude that the empirical power increases in ψ for all sample sizes and distributions. Additional results for other values of the causal persistence coefficient $\gamma = \phi$ are given in Table A.2, Appendix C.1.

Figure 2.3 illustrates the power of the GCov specification test against local alternatives, computed from the sample of size $T = 500$ and size-adjusted. The three local power functions of the MAR(0,1) model with Uniform, Laplace and t(5) distributed errors are plotted against $\delta = 0, 0.1, \ldots, 0.9$, where $\phi_T = \gamma_T = \frac{\delta}{\sqrt{T}}$. Panel (a) displays the results for the MAR(0,1) model with $\psi = 0.3$ and panel (b) shows the results for $\psi = 0.7$. We observe that the test has good local power in each case. The best local power is displayed in both panels by the Uniform distributed errors with the heaviest tails. The local power functions increase faster



(a) MAR(0,1) with $\psi = 0.3$ (b) MAR(0,1) with $\psi = 0.7$

Figure 1.3: Local asymptotic power of GCov specification test

in panel (b) for the model with higher noncausal persistence than in panel (a). Overall, the rate of increase of local power functions is lower compared with the local power of the (non)linear dependence NLSD test displayed in Figure 2.2. Hence, larger sample sizes are recommended for the specification test.

1.6.1.5 Infinite Variance Errors

So far, we have considered only error distributions with finite variances. The size and power properties of the GCov specification test of the MAR(0,1) model with Cauchy distributed errors are illustrated in Table A.5, Appendix C.2. We use K=3, with the square roots and logarithms of the absolute values of residuals as nonlinear transformations, and set the maximum lag equal to $H = 3$. We observe that the empirical size is close to the nominal for T=200, and the power quickly converges to one for both values of ψ as the number of observations increases. Compared to Table 1.2, the power is higher for $T = 100$ and the convergence in T is faster.

1.6.2 GCov Bootstrap Test

In the first part of this section, we consider the bootstrap approximation of finite-sample size and power of the GCov specification test based on the GCov estimator. In the second part, we examine the GCov test and GCov bootstrap test, both with the AML estimator.

1.6.2.1 GCov Bootstrap Test with GCov Estimator

The performance of GCov bootstrap test depends on the quality of the bootstrap approximation of finite sample properties of a GCov test. We consider a MAR(0,1) model under the null hypothesis, and a MAR(1,1) under the fixed alternative with the same error distributions and parameter values as those in Section 2.6.1.4. The GCov test statistic is based on $K = 2$ with residuals and their squares, and $H = 3$. The bootstrap with $S = 100$ is replicated 1000 times.

Table A.7, Appendix C.3 shows the bootstrap approximation of finite sample size and power of GCov specification test obtained from rejection rates with a chi-square quantile used as the critical value and two sampling schemes: with and without replacement. By comparing these results with Table 2.2, we observe that the bootstrap provides a reliable approximation of finite-sample size, and the quality of the approximation improves with increasing T. Moreover, we find that sampling with replacement provides more accurate results. An additional insight is obtained for the MAR(0,1) process with t(5)-distributed errors and $\psi = 0.7$. Figure A.1, Appendix C.3 shows the convergence of the distribution of the GCov bootstrap test statistic to the theoretical chi-square distribution when the sample size increases from T=100 to T=500.

1.6.2.2 Parametric Estimation of Causal-Noncausal Models

For diagnostic checking of parametric models, the GCov specification test can be applied to residuals $\hat{u}_{T,t} = g(\tilde{Y}_t, \tilde{\theta}_T)$, where $\tilde{\theta}_T$ is not the GCov estimator.

When a causal-noncausal model is fully parametric, and the errors are assumed to follow a parametric density, the Approximate Maximum Likelihood (ML) estimator can be used.

The Approximate Maximum Likelihood (AML) estimator of univariate MAR(r,s) processes defined in equation (3.6) and introduced by Lanne and Saikkonen (2011) is:

$$(\hat{\Psi},\, \hat{\Phi},\, \hat{\nu}) = Argmax_{\Psi,\Phi,\nu} \sum_{t=r+1}^{T-s} \ln\, f[\Psi(L^{-1})\Phi(L)y_t; \nu],$$

where $f[.;\nu]$ denotes the non-Gaussian probability density function of u_t, such as a t(ν)-student density, for example. Davis and Song (2020) discuss the ML estimator for the multivariate causal-noncausal VAR process given in Example 3, Section 2.3.1, with a parametric error density.

1.6.2.3 GCov Test and GCov Bootstrap Test with AML Estimator

To illustrate the performance of GCov bootstrap test with the AML estimator, we consider the simulated processes with errors from the t(4), t(5) and t(6) distributions. Like in the previous Section, the model under the null hypothesis is a MAR(0,1), and under the fixed alternative is a MAR(1,1). First, we estimate the MAR(0,1) model from the AML estimator based on a t-student log-likelihood function. Next, we plug the AML residuals into the formula of the GCov specification test statistic in eq. (3.3), i.e., we calculate the sample autocovariances from the AML residuals and squared residuals (K=2) and their lags up to H=3.

Table 1.3: **A:** Empirical size and power of GCov test with plug-in AML residuals and **B:** Empirical size and power of GCov bootstrap test of MAR(0,1) with AML, at 5% significance

Panel	S./P.	ϕ	ψ	T=100			T=200			T=500		
				t(4)	t(5)	t(6)	t(4)	t(5)	t(6)	t(4)	t(5)	t(6)
A	S.	0	0.3	0.065	0.061	0.061	0.071	0.050	0.054	0.070	0.051	0.061
		0	0.7	0.064	0.064	0.061	0.072	0.048	0.050	0.072	0.056	0.062
	P.	0.8	0.3	0.431	0.400	0.391	0.735	0.697	0.645	0.988	0.990	0.986
		0.8	0.7	0.994	0.996	0.993	1	1	1	1	1	1
B	S.	0	0.3	0.060	0.061	0.067	0.064	0.048	0.051	0.058	0.045	0.056
		0	0.7	0.061	0.064	0.063	0.066	0.050	0.049	0.053	0.044	0.055
	P.	0.8	0.3	0.430	0.410	0.388	0.704	0.664	0.633	0.982	0.980	0.975
		0.8	0.7	0.991	0.993	0.911	1	1	1	1	1	1

S.: size, P.: power (size-adjusted)

This approach is a one-step procedure proposed as an alternative to two separate Ljung-Box tests applied to the residuals and their squares. The experiment is replicated 1000 times. Then, the finite sample size and power are evaluated from rejection rates based on the theoretical critical value and given in part A of Table 1.3, where we observe a slight size distortion of the test in finite samples. Nevertheless, the test has reasonably good performance.

Instead of this simple procedure, one can use the GCov bootstrap test to ensure better finite sample properties. We examine the size and power of the GCov bootstrap test of MAR(0,1) against MAR(1,1) with $S = 100$ and AML-estimated residuals based on 1000 replications. The results are reported in Table 1.3-B based on sampling with replacement. We find that the size of the GCov bootstrap is close to the nominal level of 0.5 for all sample sizes and error distributions. The power of the GCov bootstrap test is high and increases to 1 with sample size[13]. Moreover, we report the size and power of the bootstrap test based on sampling without replacement in Table A.8 of Appendix C.3, providing additional evidence

[13] The size adjustment is based on the bootstrapped critical value.

based on test size that sampling with replacement proposed in Section 2.4 is preferred.

1.6.2.4 Infinite Variance Errors

Table A.6, Appendix C.2 presents additional results on the GCov bootstrap test of MAR(0,1) with Cauchy error distribution and OLS estimator. The size of our test compares favorably with the size of the bootstrap test of coefficient ψ in MAR(0,1) based on the (unconstrained) OLS estimator and sampling without replacement introduced in Cavaliere et al. (2020), who consider the test of the null hypothesis $\rho = \psi = 0.5$ by unrestricted bootstrap for Stable error distribution with parameters $\alpha = 1.0$ and $\beta = 0.0$ based on the test statistic r_T. The size of our GCov bootstrap test is between 5.3 and 5.5, depending on ψ, which is close to 5.4 reported in Table 2, Cavaliere et al. (2020). The results from sampling with and without replacement are both provided in TableA.6. We observe that, as in Table 2.3, sampling with replacement provides a size closer to the nominal.

1.6.3 GCov Specification Test with Many Transformations

We examine the finite sample performance of the GCov specification test with many transformations applied to the MAR(0,1) processes with t(5) distributed errors, coefficients $\psi = 0.3$ and $\psi = 0.7$ generated in samples of size: $T = 100, 200, 500$. The models are estimated by GCov with $K = 7, 8, 9$ transformations, which are the absolute values of consecutive powers of the residuals weighted by exponential functions with $t = 0.01$, and the maximum lags are $H = 3, 4, 5$. Thus, the number of nonlinear autocovariance conditions used in this study ranges from 48 to 80. The size of the GCov specification test based on rejection rates from 1000 replications is reported in Table 1.4. We observe that the size approaches the nominal value as K, H and T increase. For both values of ψ and all values of H we observe that the size is closer to the nominal for $K = 9$, in comparison to $K = 7, 8$. For $\psi = 0.7$ the nominal size is reached for $K = 7$ and $T = 500$, for all H considered. For $\psi = 0.3$ more transformations $K = 9$ and $T = 500$ are needed, for all H considered.

Table 1.4: Empirical size of GCov test with many transformations at 5% level, $T = 500$.

S./P.	H	$K = 7$		$K = 8$		$K = 9$	
		$\psi = 0.3$	$\psi = 0.7$	$\psi = 0.3$	$\psi = 0.7$	$\psi = 0.3$	$\psi = 0.7$
S.	$H = 3$	0.033	0.051	0.035	0.051	0.054	0.050
	$H = 4$	0.031	0.049	0.034	0.059	0.050	0.056
	$H = 5$	0.031	0.053	0.029	0.052	0.053	0.060

The power functions are illustrated in Appendix C 4. We find that the power of test converges quickly to 1 in sample size T, and the convergence is faster for higher values of ψ.

1.7 Empirical Application

In this section, we apply the GCov specification test to a univariate causal-noncausal model fitted to the series of aluminum prices in U.S. Dollars per metric ton. This approach is motivated by the presence of spikes and bubbles in aluminum prices, and the recent literature on causal-noncausal modeling of commodity prices [Hecq et al. (2016), Fries and Zakoian (2019), Gourieroux and Jasiak (2023)]. First, we apply the (non)linear serial dependence NLSD test to the data, and next use the GCov specification test to examine the goodness of fit of the estimated causal-noncausal processes.

Our sample consists of $T = 228$ monthly average prices recorded between January 2005 and October 2024 and known as the Global price of Aluminum [14]. We detrend the series of prices by regressing it on time (polynomial of degree one). The detrended prices are plotted in Figure 1.4a, where we observe multiple spikes and a sudden drop in aluminum prices during the 2008 recession when the commodity prices fell due to weak demand. In addition, in 2020, we see a spike in the price of aluminum, which is due to the weak supply of commodities at the beginning of the Covid period.

To ensure the identification of causal and noncausal dynamics, we test the data for

[14] International Monetary Fund, Global price of Aluminum [PALUMUSDM], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/PALUMUSDM, December 20, 2024.

normality by applying the Kolmogorov-Smirnov normality test. The test statistic of 0.50 exceeds the critical value of 0.08. Hence, the null hypothesis of the normality of the aluminum price distribution is rejected. Figure A.4 (a) in Appendix C.4 provides the sample density plot of demeaned aluminum price and compares it to the Gaussian density to confirm that the aluminum prices are non-Gaussian.

We test for nonlinear serial dependence in the aluminum prices using the NLSD test introduced in Section 2.2. We compute the test statistic from the series using H=9 and K=2. The value of the NLDS test is 1675.4 and exceeds the critical value of 50.99, showing (non)linear serial dependence in the data. This finding is confirmed by the ACF of the series and their squares in Figures A.5a and A.5b in Appendix C.4.

Next, we use the semi-parametric GCov estimator and explore several specifications of the causal-noncausal MAR(r,s) models for varying autoregressive orders r and s without any distributional assumptions on the errors. In Table 1.5 we report the GCov estimated parameters with H=9 and K=2, where we use the residuals and the logarithm of the second power of the residuals to capture the linear and nonlinear serial dependencies, respectively. By considering nonlinear autocovariances at higher lags, we accommodate possible long-range serial dependence, with the total of 18 transformations. Under the strict stationarity assumption, all models have autoregressive polynomials with roots outside the unit circle. The estimated roots of the polynomials $\hat{\Phi}(L)$ and $\hat{\Psi}(L^{-1})$ are given in the last column of Table 2.4.

We apply the GCov specification test to each model to assess the fit. The test statistics and the associated critical values are given in column 3 of Table 1.5. We find that the GCov specification test does not reject the null hypothesis of correct specification of the MAR(1,1) model, indicating the absence of (non)linear serial dependence in the residuals. The in-sample fitted values of the MAR(1,1) model are shown in Figure 1.4b. Figure A.4 (b) in Appendix C.5 illustrates the non-Gaussian distribution of the residuals. The ACFs of the residuals and their squares given in Figures A.5c and A.5d of Appendix C.4 are not statistically significant, which confirms the results of the GCov specification test. Furthermore, we plot the fitted causal and noncausal components of MAR(1,1), defined in equation (6.2), $\hat{v}_{1,t} = (1 - \hat{\psi}L^{-1})y_t$ and $\hat{v}_{2,t} = (1 - \hat{\phi}L)y_t$ in Figure A.6 of Appendix C.5.

Table 1.5: GCov estimated parameters, GCov specification test with χ^2 critical values at 5% significance level, and roots of $\hat{\Phi}(L^{-1})$ and $\hat{\Psi}(L)$

	ϕ_1	ψ_1	ψ_2	test statistic	$\chi^2_{0.95}$	L_1^{ϕ}	L_1^{ψ}	L_2^{ψ}
MAR(0,1)		0.93*		57.53	49.80	1.07		
MAR(1,1)	**0.41***	**0.87***		**22.32**	**48.60**	**2.43**	**1.14**	
MAR(1,2)	0.62*	0.76*	0.02	22.60	47.4	1.60	1.27	-36.93

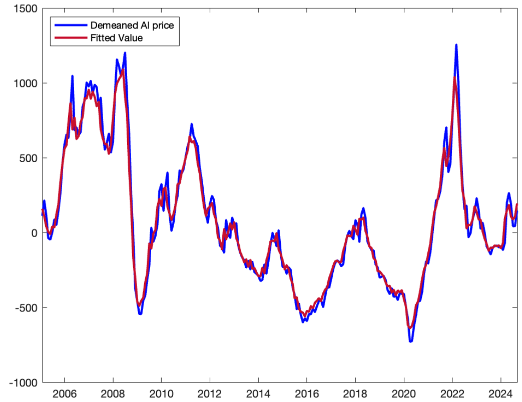
* indicates statistical significance at 5%

In addition, we estimate the MAR(1,1) model using the AML estimator with a log-likelihood function based on the fitted t-student error distribution with 3.9 degrees of freedom[15] and apply the GCov bootstrap test. Since AML suffers from a bi-modality issue [Hecq et al. (2016) and Bec, Nielsen, and Saïdi (2020)], we employ a grid search of initial values as suggested in Bec, Nielsen, and Saïdi (2020) for parameters ϕ and ψ with a grid size of 0.01 between 0 and 1. The AML estimates are $\hat{\phi} = 0.42$ and $\hat{\psi} = 0.90$. The GCov bootstrap test does not reject the null hypothesis of i.i.d. errors, based on comparing the value of the test statistic 42.30 with the critical value of 57.78.

The parameters of MAR(1,1) estimated by the GCov and AML estimators are very close. In both cases, we do not reject the null of correct specification, based on the GCov specification test (Table 1.5) and the GCov bootstrap test, respectively. By comparing the approaches used to estimate the parameters, we can argue that the GCov method provides an estimator and a test with a known asymptotic distribution in one step, while the AML estimator needs great attention to the initial values of optimization and grid search. Moreover, the GCov bootstrap test replaces two separate Liung-Box tests with one testing procedure of a joint hypothesis.

[15] The degrees of freedom are an additional AML parameter estimated in this model.

(a) Detrended aluminum price



(b) MAR(1,1) fitted values

Figure 1.4: Detrended aluminum price, MAR(1,1) fitted values

1.8 Conclusion

This paper examined nonlinear serial dependence testing in non-Gaussian time series and specification testing in semi-parametric models with non-Gaussian i.i.d. errors. We introduced a new test of the null hypothesis of the absence of (non)linear dependence in time series, called the NLSD test. We studied analytically the asymptotic properties of the GCov specification test under the local alternative hypotheses to provide convincing empirical evidence of its potential as a widely applicable diagnostic tool for testing the goodness of fit. We showed that under the local alternatives, the GCov specification test statistic converges to a non-central chi-square distribution with a non-centrality parameter depending on the direction of the alternative. We also compared the GCov specification test with a concentrated and extended CUGMM-based overidentification test.

For finite sample size adjustment of GCov specification tests in dynamic non-Gaussian models estimated by GCov or maximum likelihood, we introduced the GCov bootstrap test and provided regularity conditions ensuring its validity.

We also explored how the asymptotic performance of the GCov specification test can be improved by increasing the set of nonlinear autocovariance conditions. We showed that the nonlinear transformations that help identify the unknown error distribution are power functions weighted by decreasing exponential functions of errors. The GCov estimator based on that enlarged set of nonlinear transformations is asymptotically more efficient, which increases the set of alternatives against which the GCov test has asymptotic power of 1. In addition, by appropriately choosing the nonlinear transformations and orthonormalizing them in the Hilbert space, we obtained an optimally weighted test statistic with an asymptotic normal distribution.

We examined through simulations the finite sample properties of the GCov specification, bootstrap and NLSD tests under the fixed and local alternative hypotheses. Our study shows that these tests perform well in detecting nonlinear serial dependence in mixed causal-noncausal processes and residuals of nonlinear models. For illustration, we applied the NLSD test to the aluminum price and used the GCov specification and bootstrap tests to select the optimal fit of a causal-noncausal MAR model.

Chapter 2

Regularized Generalized Covariance Estimator

Coauthorship Statement

This chapter is based on joint work titled “Regularized Generalized Covariance Estimator" with Francesco Giancaterini, Alain Hecq and Joann Jasiak. All authors contributed equally to the conceptualization, data preparation, empirical modeling, analysis, and manuscript writing.

This is an Original Manuscript of an article published in The Econometrics Journal on 10 Apr 2026, available online: https://doi.org/10.1093/ectj/utag011

2.1 Introduction

Mixed causal and noncausal processes have become increasingly popular for modeling nonlinear and dual patterns in time series, especially in financial contexts such as cryptocurrency rates and commodity prices. These financial variables often exhibit bubbles and other short-lasting nonlinear and explosive patterns [Hencic and Gouriéroux (2015), Lof and Nyberg (2017), Hecq and Voisin (2021), Gourieroux et al. (2021a)]. In this context, it is crucial to recognize that conventional measures of past dependence, often referred to as "causal dependence," may not sufficiently capture the temporal dependence of these variables. This motivates interest in mixed causal and noncausal processes, as explored by, among others, Breidt et al. (1991), Gourieroux and Jasiak (2016), Hecq et al. (2016), Gouriéroux and Zakoïan (2017), Gourieroux and Jasiak (2024), and Truchis et al. (2025).

In the univariate framework, several studies investigated the nonlinear characteristics in financial data, especially bubbles, by using mixed causal and noncausal models [Hencic and Gouriéroux (2015), Fries and Zakoian (2019), Gourieroux et al. (2021a); Gourieroux et al. (2021b); Blasques et al. (2025)]. Noncausal modeling techniques have also been applied to examine processes in a multivariate context [Gourieroux and Jasiak (2017), Gourieroux and Jasiak (2022), Swensen (2022), Moussa and Thomas (2023), and Cubadda et al. (2025)].

While most existing studies focus on multivariate mixed causal-noncausal models with a small number of time series and do not address high-dimensional frameworks, this paper extends the analysis to high-dimensional settings. To this end, it introduces the Regularized Generalized Covariance (RGCov) estimator, which builds upon the Generalized Covariance (GCov) estimator originally proposed by Gourieroux and Jasiak (2017, 2023) and further examined in Jasiak and Neyazi (2023) and Cubadda et al. (2024). The GCov estimator is specifically designed for strictly stationary non-Gaussian processes. It relies on the autocovariances of nonlinear functions of the process, which capture both nonlinear and noncausal serial dependence (Chan et al. (2006)). The estimator minimizes an objective function analogous to a multivariate coefficient of determination, which involves the inverse of the variance-covariance matrix (specifically, the autocovariance at lag zero). In this work, the GCov is favored over parametric estimators, such as the maximum likelihood estimator proposed by Lanne and Saikkonen (2011) and Lanne and Saikkonen (2013), for identifying and estimating mixed causal-noncausal models, as it does not require distributional assumptions on the error term beyond non-Gaussianity. Moreover, GCov can be viewed as a Portmanteau-type test and serves as a consistent, asymptotically normal one-step estimator that is semiparametrically efficient and capable of achieving parametric efficiency when appropriate nonlinear transformations are selected [Gourieroux and Jasiak (2023)].

However, in high-dimensional settings, the inversion of the variance-covariance matrix in the GCov objective function may become ill-conditioned or nearly singular, leading to numerical instability. To address this issue, similar to what Carrasco and Nokho (2022) does with the GMM estimator, the RGCov estimator employs Ridge-type regularization, which inflates small eigenvalues and stabilizes the matrix inversion process. Other types of regularization methods, such as Lasso-type regularization, are not suitable in this context, as they would shrink small eigenvalues to zero, resulting in a singular variance-covariance matrix and making estimation infeasible. It should be noted that the causal-noncausal VAR parameters are not regularized in this paper; instead, their unpenalized estimates are used, for instance, to construct portfolios that are free from nonlinear bubble patterns. This approach differs from methodologies that sparsify VARs in high-dimensional settings (see, e.g., Hecq et al. (2023) and Adamek et al. (2023) for a post-double-selection approach in causal VAR models).

The rest of the paper is organized as follows. Section 3.2 reviews the main results concerning the GCov estimator. Section 3.3 introduces our new RGCov estimator, whose performance is assessed in Section 3.4 through a Monte Carlo study. Section 3.5 estimates a mixed causal-noncausal VAR for stocks included in the Rennix Green Index. Section 3.6 concludes.

2.2 The VAR Model and GCov Estimator

This section describes the causal-noncausal Vector Autoregressive (VAR) model and the semi-parametric GCov estimator, along with the corresponding specification test.

2.2.1 The Model

Let us consider a strictly stationary process $\{Y_t, t \in \mathbb{Z}\}$ of dimension n satisfying $g(\tilde{Y}_t; \theta_0) = u_{0,t}$, where g is a known function, $\tilde{Y}_t = (Y_t, Y_{t-1}, \dots, Y_{t-p})$, p is a given integer, θ_0 is an unknown parameter vector, and $u_{0,t}$'s are i.i.d. with unknown density f_0 and are not assumed to be independent of $\tilde{Y}_{t-1}$. This dynamic is compatible with the semi-parametric model:

$$g(\tilde{Y}_t; \theta) = u_t(\theta), \tag{2.1}$$

where $u_t(\theta)$'s are i.i.d. with density f and are of a known dimension equal to or strictly smaller/greater than n. We assume that the model is correctly specified and that the true values of the parameters θ and f are θ_0 and f_0, respectively, and identifiable; that is, $P_Y(\theta, f) = P_Y(\theta_0, f_0)$ has a unique solution for $\theta = \theta_0, f = f_0$, where P_Y is the distribution of $(Y_t, t \in \mathbb{Z})$.

An example of the process (2.1) is the multivariate causal-noncausal VAR(p) model, defined as:

$$Y_t = \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + u_t(\theta),$$

where $\theta = [\text{vec}(\Phi_1)', \dots, \text{vec}(\Phi_p)']'$[1], and error $(u_t(\theta), t \in \mathbb{Z})$ is a multivariate non-Gaussian i.i.d. process of dimension n with finite fourth-order moments. Suppose that the roots of the characteristic equation $\det(I_d - \Phi_1 z - \dots - \Phi_p z^p) = 0$ are of a modulus that is strictly greater than or less than one. Then, there exists a unique strictly stationary solution with a two-sided moving average (MA(∞)) representation satisfying model (2.1) with:

$$g(\tilde{Y}_t, \theta) = Y_t - \Phi_1 Y_{t-1} - \dots - \Phi_p Y_{t-p} = u_t(\theta).$$

The causal-noncausal VAR(p) model was studied in Gourieroux and Jasiak (2016, 2017),

[1]For any $m \times n$ matrix A whose j-th column is a_j, $j = 1, \dots, n$, $\text{vec}(A)$ denotes the column vector of dimension mn defined as $\text{vec}(A) = (a_1', \dots, a_n')'$, where the prime denotes transposition.

Swensen (2022), Cubadda et al. (2024), and Hall and Jasiak (2024). In this framework, the error term $u_{0,t} = u_t(\theta_0)$ cannot be interpreted as an innovation because it is correlated with the past y's. In addition, even though the function g is linear in the current and lagged values of Y_t, the presence of noncausal components in Y_t implies nonlinear dynamics of Y_t from the calendar time perspective, with $E_0(Y_t|\underline{Y_{t-1}})$ nonlinear in $\underline{Y_{t-1}} = (Y_t, Y_{t-1}, ...)$ and conditional heteroscedasticity $V_0(Y_t|\underline{Y_{t-1}})$ where the expectations are taken with respect to the true density f_0[2].

The representation theorem introduced by Gourieroux and Jasiak (2017) for mixed processes distinguishes between their purely causal and noncausal latent components. For the mixed VAR(1) model:

$$Y_t = \Phi Y_{t-1} + +u_t,$$

with a diagonalizable autoregressive coefficient matrix[3] we can rewrite the matrix Φ as:

$$\Phi = AJA^{-1},$$

where J is a diagonal matrix with eigenvalues of Φ on the diagonal, and A is an invertible matrix consisting of the eigenvectors of Φ. Furthermore, let us denote by n_1 the number of eigenvalues of Φ with a modulus strictly less than 1 and by n_2 the number of eigenvalues with a modulus strictly greater than 1, where $n_2 = n - n_1$. Then, it is possible to express J as follows:

$$J = \begin{bmatrix} J_1 & \mathbf{0} \\ \mathbf{0} & J_2 \end{bmatrix},$$

where J_1 is of dimension $n_1 \times n_1$ and has on its main diagonal the eigenvalues of Φ that lie inside the unit circle, while J_2 is of dimension $n_2 \times n_2$ and has on its diagonal the eigenvalues

[2]Model (2.1) can also be extended to invertible ARMA processes approximated by an AR(p) process of large order p. There are extensions towards noncausal noninvertible ARMA for univariate models, although they are more difficult to identify [see Velasco and Lobato (2018), Hecq, Velasquez (2025)].

[3]Otherwise, we use the real Jordan form of matrix Φ.

of Φ that lie outside the unit circle. Consequently:

$$y_t = A_1 y^*_{1,t} + A_2 y^*_{2,t}, \tag{2.2}$$

$$y^*_{1,t} = J_1 y^*_{1,t-1} + \varepsilon^*_{1,t}, \quad y^*_{2,t} = J_2 y^*_{2,t-1} + \varepsilon^*_{2,t}, \tag{2.3}$$

$$\varepsilon^*_{1,t} = A^1 \varepsilon_t, \quad \varepsilon^*_{2,t} = A^2 \varepsilon_t, \tag{2.4}$$

$$y^*_{1,t} = A^1 y_t, \quad y^*_{2,t} = A^2 y_t, \tag{2.5}$$

where A_1, A_2 represents the blocks in the decomposition of matrix A as : $A = (A_1, A_2)$, and A^1, A^2 represents the blocks in the decomposition of A^{-1} as $A^{-1} = (A_1, A_2)'$.

Since the eigenvalues of J_1 in (2.3) are in modulus less than 1, $y^*_{1,t}$ is defined as the latent purely causal component of y_t. By similar reasoning, $y^*_{2,t}$ captures the noncausal component of the investigated process. In particular, $y^*_{2,t}$ is the locally explosive component of y_t that follows a strictly stationary noncausal (V)AR process.

The presence of noncausal serial dependence can be detected through the analysis of nonlinear autocovariances, which are the autocovariances of nonlinear functions of the observed process. This approach follows Chan et al. (2006), who shows that nonlinear dependence in strictly stationary non-Gaussian time series is revealed through the autocovariances of nonlinear transformations of the series.

Let us consider nonlinear functions that transform the multivariate process $g(Y_t;\theta)$, of dimension n, into a process $v_t(\theta)$ of dimension $K = Jn$, with components $a_j\left[g_i(Y_t;\theta)\right]$ for $i = 1, \ldots, n$ and $j = 1, \ldots, J$. The transformed process is defined as

$$v^a_t(\theta) = \left(a_j\left[g_i(\tilde{Y}_t;\theta)\right]\right)'_{1\leq j\leq J,\,1\leq i\leq n}. \tag{2.6}$$

where the functional components g_i, $i = 1, ..., n$ are known and are serially i.i.d. when $\theta = \theta_0$ and $f = f_0$. The increase in dimension reflects the augmentation by nonlinear differentiable functions of $g(\tilde{Y}_t;\theta)$, such as squares or logarithms. If $g(Y_t;\theta)$ does not have a finite fourth-order moment, it can be replaced by $v_t^a(\theta)$, which ensures the asymptotic normality of the GCov estimator discussed below.

2.2.2 The GCov Estimator

The advantage of the semi-parametric Generalized Covariance (GCov) estimator, introduced in Gourieroux and Jasiak (2017, 2023), is that it does not require any distributional assumptions on the true errors $u_{0,t} = u_t(\theta_0)$ beyond the i.i.d. condition and non-Gaussianity. In addition, it is straightforward to compute and leads to a specification test with a known limiting distribution.

2.2.2.1 The Semi-Parametrically Efficient GCov

Let us suppose a sample of T observations $Y_1, ..., Y_T$. The GCov estimator of θ based on model (2.6) after transformation a is defined as[4]:

$$\hat{\theta}_T = Argmin_\theta \sum_{h=1}^{H} Tr \hat{R}_T^2(h;\theta), \tag{2.7}$$

with

$$\hat{R}_T^2(h;\theta) = \hat{\Gamma}_T(h;\theta)\hat{\Gamma}_T(0;\theta)^{-1}\hat{\Gamma}_T(h;\theta)'\hat{\Gamma}_T(0;\theta)^{-1}, \tag{2.8}$$

where Tr denotes the trace, $\hat{\Gamma}_T(h;\theta)$ is the sample autocovariance of $v_t(\theta)$ and $v_{t-h}(\theta)$, with h denoting the lag. The sample autocovariances are obtained from the process $\{v_t(\theta)\}, t = 1, ..., T$ of dimension $K = Jn$:

$$\hat{\Gamma}_T(h;\theta) = \frac{1}{T}\sum_{t=h}^{T} v_t(\theta)v_{t-h}(\theta)' - \frac{1}{T}\sum_{t=h}^{T-1} v_t(\theta)' \frac{1}{T}\sum_{t=h+1}^{T} v_{t-h}(\theta), h = 0, ..., H. \tag{2.9}$$

The GCov estimator minimizes an objective function, equal to the sum of traces of sample multivariate coefficients of determination computed at different lags h from nonlinear

[4]For ease of exposition, the superscript a on v_t, Γ, ξ, etc., is omitted henceforth.

transforms $v_t(\theta)$. Under the regularity conditions, the GCov estimator is consistent, asymptotically normally distributed, and semi-parametrically efficient. It can achieve parametric efficiency for well-selected transformations [Gourieroux and Jasiak (2023)].

The GCov objective function can also be used to test the specification of the model by verifying the absence of nonlinear and linear serial correlation in the residuals. The null hypothesis is $H_0 = \{\Gamma(h) = 0, \quad h = 1, \ldots, H\}$, and the test statistic is defined as

$$\hat{\xi}_T(H) = T \sum_{h=1}^{H} \mathrm{Tr}\left(\hat{R}_T^2(h; \hat{\theta}_T)\right). \tag{2.10}$$

Since the process $\{Y_t\}$ satisfies $g(\tilde{Y}_t; \theta_0) = u_{0,t}$ with $u_{0,t}$'s serially i.i.d., the test statistic is asymptotically distributed as a chi-square random variable $\chi^2\left(K^2 H - \dim(\theta)\right)$, where $K = Jn$. The specification test at level α is conducted as follows: the null hypothesis H_0 is rejected when $\hat{\xi}_T(H) > \chi^2_{1-\alpha}(K^2 H - dim(\theta))$ and H_0 is not rejected otherwise [Gourieroux and Jasiak (2023)].

For preliminary data analysis, the NLSD test, inspired by the above specification test, was introduced by Jasiak and Neyazi (2023). The NLSD test of the null hypothesis of the absence of nonlinear and linear dependence in a time series relies on a test statistic computed from data transforms $v_t = [a_1(Y_{1,t}), ..., a_1(Y_{n,t}) ...., a_1(Y_{1,t}), ..., a_J(Y_{n,t})]'$ of dimension $Jn = K$. The test statistic is $\hat{\xi}_T(H) = T \sum_{h=1}^{H} Tr \hat{R}_T^2(h)$, with: $\hat{R}_T^2(h) = \hat{\Gamma}_T(h)\hat{\Gamma}_T(0)^{-1}\hat{\Gamma}_T(h)'\hat{\Gamma}_T(0)^{-1}$, where $\hat{\Gamma}_T(h)$ is the sample autocovariance of v_t defined analogously to (2.5) with $v_t(\theta)$ replaced by v_t. Under the serial independence of $\{Y_t\}$, this statistic asymptotically follows a $\chi^2(K^2 H)$ distribution.

2.2.2.2 The Diagonal GCov

The diagonal GCov estimator, introduced by Gourieroux and Jasiak (2017) [see also Cubadda and Hecq (2011)], minimizes the following objective function:

$$\hat{\theta}_T^d = Argmin_\theta \sum_{h=1}^{H} \text{Tr}\left(\hat{R}_{d,T}^2(h;\theta)\right), \tag{2.11}$$

where $\hat{R}_{d,T}^2(h;\theta)$ is defined as in eq. (2.8) with $\hat{\Gamma}_T(0;\theta)$ replaced by the diagonal matrix $\hat{\Gamma}_T^d(0;\theta)$ containing only the diagonal elements of $\hat{\Gamma}_T(0;\theta)$. This estimator is not semi-parametrically efficient, as it is not optimally weighted, and its asymptotic properties are described in Gourieroux and Jasiak (2017).

2.2.3 Nonlinear Transformations

One may choose a high number of nonlinear transformations to improve the efficiency of the GCov estimator and the asymptotic performance of the GCov specification test. The additional nonlinear transformations increase the dimension of the matrix $\hat{\Gamma}_T(0;\theta)$ in (2.8), and inverting the variance matrix of a large dimension can become numerically challenging. Jasiak and Neyazi (2023) discuss how to select the basis of transformations that provides more information on the parameters, in addition to, e.g., linear and quadratic functions of errors u_t. In the causal-noncausal models with extreme risks and local explosive patterns, including bubbles, the error does not necessarily have power moments. Moreover, some of the parameters driving those extreme risks may not be identifiable from the transformations u_t, u_t^2 alone. The following system of generators based on Post's inversion formula can be considered:

$$\mathcal{A} = \{a_{t,p}(u) = |u|^p \exp(-t|u|),\ p \in \mathbb{N}, t \in [0,1]\},$$

with a countable dense subsystem given by:

$$\mathcal{A}_n = \{a_{t_{j,n},p}(u) = |u|^p \exp(-t_{j,n}|u|),\ p \in \mathbb{N}, t_{j,n} \in [0,1],\ j = 1, ..., n\},$$

where $(t_{1,n}, ..., t_{n,n})$ is dense in $[0, 1]$ when n tends to infinity [see, Bierens (1990), page 1448, for a similar approach]. The decreasing exponential functions provide square integrability for the power transforms, ensuring adequate weighting for heavy-tailed data.

2.3 The RGCov Estimator

When n is high but finite, the $K \times K$ matrix $\hat{\Gamma}_T(0; \theta)$ is of high dimension, and its inverse may be difficult to compute numerically. This is, for example, the case of a causal-noncausal VAR process with a large number n of components. In general, the dimension of the parameter $\theta = [vec\Phi_1', ..., vec\Phi_p']'$ in the VAR(p) process is $p \times n^2$. Therefore, it is necessary to use enough non-linear transformations $J > p$ to ensure the identifiability of the model when $H = 1$. Additional nonlinear autocovariance conditions can be obtained by increasing the number of lags H. The three parameters n, J, and p play an important role in determining the dimension of the variance matrix. There are two potential issues with inverting the matrix $\hat{\Gamma}_T(0, \theta)$:

(i) The spectral decomposition of this matrix leads to eigenvalues that are close to zero. Then, the inversion problem concerns the numerical inversion given these close-to-zero eigenvalues. We refer to it as an issue of weak invertibility.

(ii) The available inversion algorithms may not be numerically efficient, even when the eigenvalues are sufficiently different from zero and positive.

The weak invertibility is addressed below by introducing a regularized objective function and defining the RGCOv estimator as its minimizer. The second issue is addressed in subsection 3.3.3, where we show an algorithm for updating the inverse of the regularized variance-covariance matrix with each newly arrived consecutive observation.

2.3.1 Definition

To solve the problem of weak invertibility of the matrix $\hat{\Gamma}_T(0;\theta)$, we introduce the Regularized (RGCov) estimator.

Definition 1: A Regularized (RGCov) estimator of θ_0 in model (2.1) is defined as:

$$\hat{\theta}_T(\delta) = Argmin_\theta L_T(\theta, \delta), \tag{3.1}$$

where the objective function:

$$L_T(\theta, \delta_T) = \sum_{h=1}^{H} Tr \hat{R}_T^2(h; \theta, \delta), \tag{3.2}$$

with

$$\hat{R}_T^2(h;\theta,\delta) = \hat{\Gamma}_T(h;\theta)\hat{\Gamma}_T(0,\theta,\delta)^{-1}\hat{\Gamma}_T(h;\theta)'\hat{\Gamma}_T(0,\theta,\delta)^{-1}, \tag{3.3}$$

depends on the regularized variance matrix:

$$\hat{\Gamma}_T(0,\theta,\delta) = \delta I + \hat{\Gamma}_T(0;\theta), \tag{3.4}$$

where I is the identity matrix of dimension K and $\delta > 0$ is the deterministic shrinkage (tuning) coefficient.

More generally, the RGCov estimator can be defined as $\hat{\theta}_T = \hat{\theta}_T(\delta_T)$, where the deterministic shrinkage parameter δ_T, with $\delta_T > 0$, tends either to zero or to a positive constant $\delta > 0$ as T tends to infinity. This includes, as a special case, $\delta_T = \delta > 0, \forall T$.

The regularization with fixed $\delta > 0$ used in Ridge regression solves the weak invertibility issue, but it leads to an estimator that is not asymptotically efficient. The asymptotic efficiency is recovered by appropriately adjusting the shrinkage parameter as the number of observations T increases.

2.3.2 The Asymptotic Properties

2.3.2.1 Asymptotic Properties of RGCov Estimator

Let us consider the following assumptions:

Assumption 3.1

i) The parameter set Θ is compact with non-empty interior $\overset{o}{\Theta}$;

ii) The model is well-specified, with θ_0 the true value of the parameter;

iii) θ_0 is asymptotically identifiable, i.e. $\Gamma(h;\theta) = 0,\ h = 1, ..., H$ implies $\theta = \theta_0$;

iv) The transformation g is continuous in θ, $\theta_0 \in \overset{o}{\Theta}$ and the transformations $a_j, j = 1, ..., J$ are continuous in u,

v) The process Y_t and errors $u_t(\theta_0) = g(\tilde{Y}_t;\theta_0),\ t = 1, 2, ...$ are strictly stationary, geometrically ergodic with a continuous invariant distribution,

vi) The deterministic sequence of shrinkage paramaters δ_T, $\delta_T > 0\ \forall T$, tends to $\delta \geq 0$, when T tends to infinity and when there $\exists\, \epsilon > 0$ such that $\Gamma(0, \theta_0, \delta)$ is invertible $\forall \delta \in [0, \epsilon]$.

vii) The transformations $g_j(\tilde{Y}_t;\theta),\ j = 1, ..., J$ satisfy the following stochastic equicontinuity condition:

$$sup_{j=1,...,J} |g_j^2(\tilde{Y}_t;\theta) - g_j^2(\tilde{Y}_t;\tilde{\theta})| \leq \mathcal{B}(\tilde{Y}_t) h[d(\theta, \tilde{\theta})], \theta, \tilde{\theta} \in \Theta,\ a.s.$$

where $d(\theta, \tilde{\theta})$ is the distance on the metric parameter space, $h(d)$ is a function that tends to 0 when d tends to 0, and $\mathcal{B}(\tilde{Y}_t)$ is a sequence of non-negative variables, such that $sup_T \frac{1}{T}\sum_{t=1}^{T} E[\mathcal{B}(\tilde{Y}_t)] < \infty$ and $\frac{1}{T}\sum_{t=1}^{T}[\mathcal{B}(\tilde{Y}_t) - E\mathcal{B}(\tilde{Y}_t)] \to 0\, a.s.$ [5].

Assumption 3.2

i) $\theta_0 \in \overset{o}{\Theta}$;

ii) The functions $g(\tilde{Y}, \theta)$ and the functions $a_j(u), j = 1, ..., J$, are twice continously differentiable with respect to θ and u, respectively.

[5] This equicontinuity condition corresponds to Assumption S-LIP in Andrews (1992), p 248, and is a stochastic Lipschitz condition on $g_j(Y;\theta)$.

iii) $g(\tilde{Y};\theta_0)$ and $\frac{\partial g(\tilde{Y};\theta_0)}{\partial \theta'}$ have finite fourth-order moments, $\frac{\partial^2 g_j(Y;\theta_0)}{\partial\theta\partial\theta'}$, $j=1,...,J$ have finite second-order moments.

Let us first consider the general case a) when the sequence of positive shrinkage parameters $\delta_T \to \delta \geq 0$ as T tends to infinity, and then the special case b) when $\delta_T \to 0$.

a) In the general case, the objective function of the RGCov estimator is:

$$L_T(\theta,\delta_T) = \sum_{h=1}^{H} Tr[\hat{\Gamma}_T(h,\theta)\hat{\Gamma}_T(0,\theta,\delta_T)^{-1}\hat{\Gamma}_T(h,\theta)\hat{\Gamma}_T(0,\theta,\delta_T)^{-1}].$$

Proposition 1: Under the regularity conditions 3.1 and 3.2, if $\delta_T \to \delta \geq 0$ when $T \to \infty$:

i) $\hat{\theta}_T = \hat{\theta}_T(\delta_T)$ is a.s. consistent of θ_0

ii) $\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, J(\theta_0,\delta)^{-1} I(\theta_0,\delta) J(\theta_0,\delta)^{-1})$, where matrices $J(\theta_0,\delta), I(\theta_0,\delta)$ are:

$$J(\theta_0,\delta) = 2\sum_{h=1}^{H}\left\{\frac{\partial vec\Gamma(h;\theta_0)'}{\partial\theta}[\Gamma(0;\theta_0,\delta)^{-1}\otimes\Gamma(0;\theta_0,\delta)^{-1}]\frac{\partial vec\Gamma(h;\theta_0)}{\partial\theta'}\right\},$$

and

$$\begin{aligned} I(\theta_0,\delta) &= 4\sum_{h=1}^{H}\left\{\frac{\partial vec\Gamma(h;\theta_0)'}{\partial\theta}\left[\Gamma(0;\theta_0,\delta)^{-1}\otimes\Gamma(0;\theta_0,\delta)^{-1}\right][\Gamma(0;\theta_0)\otimes\Gamma(0;\theta_0)]\right. \\ &\left.\left[\Gamma(0;\theta_0,\delta)^{-1}\otimes\Gamma(0;\theta_0,\delta)^{-1}\right]\frac{\partial vec\Gamma(h;\theta_0)}{\partial\theta'}\right\}. \end{aligned}$$

where the theoretical autocovariances $\Gamma(h,\theta_0)$, $\Gamma(0;\theta_0;\delta) = \delta I + \Gamma(0;\theta_0)$ are computed with respect to the true distribution f_0.

Thus, the speed of convergence of the estimator and its asymptotic distribution do not depend on how δ_T tends to δ. However, its asymptotic variance-covariance matrix depends on δ. Because this estimator is not optimally weighted, it is not semi-parametrically efficient and belongs in the class of covariance estimators considered in Gourieroux and Jasiak (2017) [see also Cubadda and Hecq (2011)].

b) In the special case $\delta_T \to \delta = 0$, the expression of the asymptotic distribution is simplified.

Proposition 2: Under the regularity conditions 3.1 and 3.2, if $\delta_T \to 0$ when $T \to \infty$:

i) $\hat{\theta}_T = \hat{\theta}_T(\delta_T)$ is a.s. consistent with θ_0.

ii) $\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, J(\theta_0)^{-1})$, where $J(\theta_0) = J(\theta_0, 0)/2$ and

$$J(\theta_0, \delta)/2 = I(\theta_0, 0)/4 = \sum_{h=1}^{H} \left\{ \frac{\partial vec\Gamma(h;\theta_0)'}{\partial \theta} [\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}] \frac{\partial vec\Gamma(h;\theta_0)}{\partial \theta'} \right\}.$$

where $\Gamma(0;\theta_0) = \Gamma(0;\theta_0, 0)$.

Proof of Propositions 1 and 2:

The demonstration follows the lines of Gourieroux and Jasiak (2023). We consider the case $\delta_T \to \delta \geq 0$, which includes both cases $\delta_T = \delta, \forall T$, and $\delta_T \to 0$.

Under Assumption 3.1 we have:

Propositions 1,2 i), **Proof of Consistency**: Let $\hat{\theta}_T(\delta_T)$ denote a solution to the RGCov minimization

$$\hat{\theta}_T(\delta_T) = Argmin_\theta \sum_{h=1}^{H} Tr \hat{\Gamma}_T(h,\theta)[\delta_T I + \hat{\Gamma}_T(0,\theta)]^{-1} \hat{\Gamma}_T(h;\theta)'[\delta_T I + \hat{\Gamma}_T(0,\theta)]^{-1},$$

Then, for sufficiently large T, there exists a solution such that $\hat{\theta}_T(\delta_T)$ converges a.s. to θ_0 as T tends to infinity.

Proof: The result is obtained by applying the standard Jennrich argument. The objective function tends asymptotically uniformly to the limiting function $\sum_{h=1}^{H} TrR^2(h;\theta,\delta)$, where $R^2(h;\theta,\delta) = \Gamma(h;\theta)\Gamma(0;\theta,\delta)^{-1}\Gamma(h;\theta)'\Gamma(0;\theta,\delta)^{-1}$, and $\Gamma(h;\theta)$ are the autocovariances computed under the true distribution of (Y_t). Moreover, the convergence is a.s. uniform in θ [6]. It follows that $\hat{\theta}_T(\delta_T)$ exists asymptotically and it a.s. converges to the solution θ_0 of the asymptotic optimization. This is the solution of $\Gamma(h,\theta) = 0, h = 1, ..., H$, since $\Gamma(h,\theta_0) = 0$, $\forall h$ and by the asymptotic identification assumption it is unique. QED

[6] The convergence is uniform except on a universal negligible set independent of θ. The existence of the universal negligible set results from the Lipschitz upper bound in vii) separating the effects of Y and θ.

Therefore, the a.s. consistency does not depend on the limiting value of the regularization parameter.

Propositions 1,2 ii), **Proof of Asymptotic Normality** The following additional assumptions are required:

The demonstration of asymptotic normality proceeds in several steps given below, following Gourieroux and Jasiak (2023), Section 4.2. Steps 1)-2) are common for the proofs of normality in Propositions 1 and 2 ii).

1) First-order derivative of the theoretical objective function.

Following Gourieroux and Jasiak (2023), and considering $h = 1$ for simplicity, we have:

$$\begin{aligned} d\,Tr\,R^2(1;\theta,\delta) &= 2Tr\,[\Gamma(0;\theta,\delta)^{-1}\Gamma(1;\theta)'\Gamma(0;\theta,\delta)^{-1}d\Gamma(1;\theta)] \\ &- Tr\,\left[R^2(1;\theta,\delta)\Gamma(0;\theta,\delta)^{-1} + \Gamma(0;\theta,\delta)^{-1}R^2(1;\theta,\delta)]d\Gamma(0;\theta,\delta)\right] \end{aligned} \quad (3.5)$$

Let us denote: $L_T(\theta,\delta) = \sum_{h=1}^{H} Tr R_T^2(h,\theta,\delta)$, for ease of exposition.

2) First-order conditions (FOC)

The first-order conditions for the minimization of the objective function are:

$$\frac{\partial L_T(\hat{\theta}_T,\delta_T)}{\partial\theta_j} = \sum_{h=1}^{H} \frac{\partial Tr\,R_T^2(h;\hat{\theta}_T,\delta_T)}{\partial\theta_j} = 0,\ j = 1,...,J = dim\theta.$$

They can be re-written by using the formula above for any h. We get:

$$\sum_{h=1}^{H}\{2\,Tr[\hat{\Gamma}_T(0;\hat{\theta}_T,\delta_T)^{-1}\hat{\Gamma}_T(h;\hat{\theta}_T)'\hat{\Gamma}_T(0;\hat{\theta}_T,\delta_T)^{-1}\frac{\partial\hat{\Gamma}_T(h;\hat{\theta}_T)}{\partial\theta_j}]$$
$$-Tr\{\,[R_T^2(h;\hat{\theta}_T,\delta_T)\hat{\Gamma}_T(0;\hat{\theta}_T,\delta_T)^{-1} + \hat{\Gamma}_T(0;\hat{\theta}_T,\delta_T)^{-1}R_T^2(h,\hat{\theta}_T,\delta_T)]\frac{\partial\hat{\Gamma}_T(0;\hat{\theta}_T,\delta_T)}{\partial\theta_j}\}\} = 0,$$

$j = 1,...,J = dim\theta$, where $\hat{\theta}_T = \hat{\theta}_T(\delta_T)$.

3) Expansion of the FOC

Let us now perform an expansion of the FOC in a neighborhood of the limiting values $\hat{\theta}_T \approx \theta_0$ and $\delta_T \approx \delta$. We get:

$$\sqrt{T}\frac{\partial L_T(\theta_0, \delta)}{\partial \theta} = \frac{\partial^2 L_T(\theta_0, \delta)}{\partial \theta \partial \theta'}\sqrt{T}(\hat{\theta}_T - \theta_0) + \sqrt{T}\frac{\partial^2 L_T(\theta_0, \delta)}{\partial \theta \partial \delta'}(\delta_T - \delta) + o_p(1),$$

where $o_p(1)$ is negligible in probability. Next, we get:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = \left(-\frac{\partial^2 L_T(\theta_0, \delta)}{\partial \theta \partial \theta'}\right)^{-1} \sqrt{T}\frac{\partial L_T(\theta_0, \delta)}{\partial \theta} - (\frac{\partial^2 L_T(\theta_0, \delta)}{\partial \theta \partial \theta'})^{-1}(\sqrt{T}\frac{\partial^2 L_T(\theta_0, \delta)}{\partial \theta \partial \delta'})(\delta_T - \delta) + o_p(1)$$

The above expansion takes into account the orders of the different derivatives. We know that the estimated autocovariances $\hat{\Gamma}_T(h, \theta)$ and variances $\hat{\Gamma}_T(0, \theta, \delta)$ are consistent of $\Gamma(h, \theta_0)$ and $\Gamma(0, \theta_0, \delta)$. Moreover, the autocovariances $\hat{\Gamma}_T(h, \theta_0)$ tend to zero at speed $1/\sqrt{T}$ and are asymptotically normally distributed. Thus, the decomposition is written to get the convergence to an invertible limit of $\frac{\partial^2 L_T(\theta_0,\delta)}{\partial \theta \partial \theta'}$, and convergence in distribution of $\sqrt{T}\frac{\partial L_T(\theta_0,\delta)}{\partial \theta}$ and $\sqrt{T}\frac{\partial^2 L_T(\theta_0,\delta)}{\partial \theta \partial \theta'}$ to a Gaussian vector and matrix, respectively.

In particular, we observe that the second term of the right-hand side is negligible if $\delta_T \to \delta$. Therefore, this expansion is equivalent to:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = \left(-\frac{\partial^2 L_T(\theta_0, \delta)}{\partial \theta \partial \theta'}\right)^{-1} \sqrt{T}\frac{\partial L_T(\theta_0, \delta)}{\partial \theta} + o_p(1).$$

We deduce that:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, J(\theta_0, \delta)^{-1} I(\theta_0, \delta) J(\theta_0, \delta)^{-1}),$$

where $J(\theta_0, \delta) = \lim_{T\to\infty} \frac{\partial^2 L_T(\theta_0,\delta)}{\partial \theta \partial \theta'}$ and $I(\theta_0, \delta) = Var\left[\sqrt{T}\frac{\partial L_T(\theta_0,\delta)}{\partial \theta}\right]$.

To prove Proposition 2 ii), we observe that when $\delta = 0$, the limiting behaviour is the same as for the GCov estimator, which has been derived in Gourieroux and Jasiak (2023). We have:

$$I(\theta_0, 0) = J(\theta_0, 0)$$

$$\text{and } \sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, J(\theta_0, 0)^{-1}),$$

where $J(\theta_0, 0) = 2\sum_{h=1}^{H}\left\{\frac{\partial vec\Gamma(h;\theta_0)'}{\partial\theta}[\Gamma(0;\theta_0,0)^{-1} \otimes \Gamma(0;\theta_0,0)^{-1}]\frac{\partial vec\Gamma(h;\theta_0)}{\partial\theta'}\right\} = J(\theta_0)$.

To prove Proposition 1 ii), we consider $\delta > 0$, for which the expressions of matrices $I(\theta_0, \delta)$ and $J(\theta_0, \delta)$ are derived below.

4) Expressions of matrices $J(\theta_0, \delta)$ and $I(\theta_0, \delta)$ when $\delta_T \to \delta > 0$.

Following the proof in Gourieroux and Jasiak (2023), Appendix A.2.2 of Supplementary Material, we can compute the second-order derivative of the objective function at the limiting values and deduce that:

$$-\frac{\partial^2 L_T(\theta_0, \delta)}{\partial\theta\partial\theta'} \overset{T\to\infty}{\to} J(\theta_0, \delta),$$

with

$$J(\theta_0, \delta) = 2\sum_{h=1}^{H}\left\{\frac{\partial vec\Gamma(h;\theta_0)'}{\partial\theta}[\Gamma(0;\theta_0,\delta)^{-1} \otimes \Gamma(0;\theta_0,\delta)^{-1}]\frac{\partial vec\Gamma(h;\theta_0)}{\partial\theta'}\right\},$$

Therefore, we have:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = [J(\theta_0, \delta)]^{-1}\sqrt{T}\frac{\partial L_T(\theta_0, \delta)}{\partial\theta} + o_p(1).$$

Let us now focus on the limiting score. Since at the limit $R^2(h, \theta_0, \delta) = 0, h = 1, ..., H$, we have asymptotically:

$$\begin{aligned}
\sqrt{T}\frac{\partial L_T(\theta_0, \delta)}{\partial\theta_j} &= 2\sum_{h=1}^{H} Tr\left[\Gamma(0;\theta_0,\delta)^{-1}\sqrt{T}\hat{\Gamma}_T(h;\theta_0)'\Gamma(0;\theta_0,\delta)^{-1}\frac{\partial\Gamma(h;\theta_0)}{\partial\theta_j}\right] + o_p(1) \\
&= 2\sum_{h=1}^{H} Tr\left[\Gamma(0;\theta_0,\delta)^{-1}\frac{\partial\Gamma(h;\theta_0)}{\partial\theta_j}\Gamma(0;\theta_0,\delta)^{-1}\sqrt{T}\hat{\Gamma}_T(h;\theta_0)\right] + o_p(1) \\
&= 2\sum_{h=1}^{H} Tr\left\{\frac{\partial vec\Gamma(h;\theta_0)'}{\partial\theta_j}\Gamma(0;\theta_0,\delta)^{-1} \otimes \Gamma(0;\theta_0,\delta)^{-1}vec[\sqrt{T}\hat{\Gamma}_T(h;\theta_0)]\right\} + o_p(1).
\end{aligned}$$

We know that under the serial independence condition, the vectors $vec[\sqrt{T}\hat{\Gamma}_T(h, \theta_0)]$, $h = 1, ..., H$, are asymptotically independent and have the same Gaussian distribution $N[0, \Gamma(0, \theta_0)\otimes$

$\Gamma(0,\theta_0)]$ [Chitturi (1976), Hannan (1976))]:

$$vec[\sqrt{T}\hat{\Gamma}'_T(h,\theta_0)] \sim N[0, \Gamma(0,\theta_0) \otimes \Gamma(0,\theta_0)], \; h = 1, ..., H$$

Therefore $\sqrt{T}\frac{\partial L_T(\theta_0,\delta)}{\partial\theta}$ is asymptotically normally distributed, with mean zero and asymptotic variance:

$$\begin{aligned} V_{asy}[\sqrt{T}\frac{\partial L_T(\theta_0,\delta)}{\partial\theta}] &= 4\sum_{h=1}^{H}\left\{\frac{\partial vec\Gamma(h;\theta_0)'}{\partial\theta}\left[\Gamma(0;\theta_0,\delta)^{-1}\otimes\Gamma(0;\theta_0,\delta)^{-1}\right][\Gamma(0;\theta_0)\otimes\Gamma(0;\theta_0)]\right. \\ & \left.\left[\Gamma(0;\theta_0,\delta)^{-1}\otimes\Gamma(0;\theta_0,\delta)^{-1}\right]\frac{\partial vec\Gamma(h;\theta_0)}{\partial\theta'}\right\} = I(\theta_0,\delta) \quad (a.2). \end{aligned}$$

Since the asymptotic properties do not depend on the way δ_T tends to 0, these properties are the same as those of the GCov estimator. Then, the RGCov is asymptotically semi-parametrically efficient [Gourieroux and Jasiak (2023)]. In practice, we estimate the asymptotic variance by replacing θ_0 by $\hat{\theta}_T$, matrices $\Gamma(h;\theta_0)$ by $\hat{\Gamma}_T(h;\hat{\theta}_T)$ and $\Gamma(0;\theta_0,\delta)$ by $\hat{\Gamma}_T(0;\hat{\theta}_T,\delta)$. The partial derivative can be either approximated numerically, or computed knowing that for a known function $a(.)$, we have $\frac{\partial}{\partial\theta_k}\Gamma(h,\theta) = Cov[\frac{\partial}{\partial\theta_k}a(Y_t,\theta), a(Y_{t-h},\theta)] + Cov[a(Y_t,\theta), \frac{\partial}{\partial\theta_k}a(Y_{t-h},\theta)], \; \forall k$.

We observe that the expressions of $J(\theta_0,\delta)$ and $I(\theta_0,\delta)$ (up to factor 4) are similar to the expressions appearing in the variance of a Generalized Least Squares (GLS) estimator in a regression model with a weighting matrix Ω^{-1} when the true weights are Ω_0, say, that is:

$$(X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}\Omega_0\Omega^{-1}X)(X'\Omega^{-1}X)^{-1},$$

where $X = \frac{\partial vec\Gamma'(h,\theta_0)}{\partial\theta}$, $\Omega = \Gamma(0,\theta_0,\delta)\otimes\Gamma(0,\theta_0,\delta)$, and $\Omega_0 = \Gamma(0,\theta_0)\otimes\Gamma(0,\theta_0)$.

2.3.2.2 Asymptotic Properties of RGCov Specification and RNLSD Test Statistics

The regularization with δ_T can be used to build test statistics for testing the model specification or for testing the absence of linear and nonlinear dependence in the data.

We can use the RGCov estimator $\hat{\theta}_T = \hat{\theta}_T(\delta_T)$ in the formulas for the test statistics to test the fit of the model by extending the GCov test introduced in Gourieroux and Jasiak (2023). Let us define the residual-based RGCov test statistic for testing the model specification:

$$\hat{\xi}_T(H, \delta_T) = T \sum_{h=1}^{H} Tr \hat{R}_T^2(h, \hat{\theta}_T(\delta_T), \delta_T). \tag{3.6}$$

When $dim(\theta) > 0$, the test statistic (3.5) tests the specification of the semi-parametric model, which is applied to the residuals of the model estimated by the RGCov estimator. The case $dim(\theta) = 0$ corresponds to the regularized extension of the NLSD test [Jasiak and Neyazi (2023)]. The Regularized NLSD (RNLSD) test statistic tests the null hypothesis of the absence of linear and nonlinear dependence in the data themselves: $H_0 = (\Gamma(h) = 0,\ h = 1, ..., H)$.

a) Let us again consider first the general case $\delta_T \to \delta \geq 0$.

Proposition 3: If $\delta_T \to \delta \geq 0$, when $T \to \infty$, then under the independence of $u_{0,t}'s$, the RGCov specification test statistic is asymptotically distributed as a positive combination of independent chi-square variables:

$$\hat{\xi}_T(H, \delta_T) \xrightarrow{d} \sum_{l=1}^{L} \lambda_l Z_l,$$

where $L = K^2H - dim(\theta)$, Z_l, $l = 1, ..., L$ are independent $Z_l \sim \chi^2(1)$ and $\lambda_l \forall l$ are the non-zero eigenvalues of $diag[\Omega(\theta_0)^{1/2}\Omega(\theta_0, \delta)^{-1/2}](Id - P)diag[\Omega(\theta_0, \delta)^{-1/2}\Omega(\theta_0)^{1/2}]$ where $\Omega(\theta_0, \delta) = \Gamma(0, \theta_0, \delta) \otimes \Gamma(0, \theta_0, \delta)$, $\Omega(\theta_0) = \Gamma(0, \theta_0) \otimes \Gamma(0, \theta_0)$,
$P = diag[\Omega(\theta_0, \delta)^{-1/2}]\frac{\partial \gamma(\theta_0)}{\partial \theta'} J(\theta_0, \delta)^{-1} \frac{\partial \gamma'(\theta_0)}{\partial \theta} diag[\Omega(\theta_0, \delta)^{-1/2}]$, $\gamma(h, \theta_0) = vec\Gamma(h, \theta_0), \forall h$ and

$\gamma(\theta_0) = [\gamma'(1,\theta_0), ..., \gamma'(H,\theta_0)]'$.

For the Regularized NLSD (RNLSD) test statistic with $dim(\theta) = 0$, we have:

$$\hat{\xi}_T(H, \delta_T) \xrightarrow{d} \sum_{l=1}^{K^2} \lambda_l^* Z_l^*,$$

where Z_l^*, $l = 1, ..., K^2$ are independent $\chi^2(H)$ variables and $\lambda_l^*, l = 1, ..., K^2$ are the eigenvalues of $[\Gamma(0)^{-1/2}\Gamma(0,\delta)\Gamma(0)^{-1/2}] \otimes [\Gamma(0)^{-1/2}\Gamma(0,\delta)\Gamma(0)^{-1/2}]$. The properties of the tensor (Kronecker) product imply that λ_l^* are obtained by considering all products $\mu_j\mu_k$, where μ_k, $k = 1, ..., K$ are the eigenvalues of $\Gamma(0)^{-1/2}\Gamma(0,\delta)\Gamma(0)^{-1/2}$.

Proof: In the case $dim(\theta) \geq 1$, the derivation of the asymptotic distribution of the test statistic needs to account for the estimator $\hat{\theta}_T$, creating dependence between the components of the objective function at the optimum. Then, it is preferable to derive the results from the vectorized representation of multivariate autocovariances: $\hat{\gamma}_T(h,\theta) = vec\hat{\Gamma}_T(h,\theta), \forall h$, with $\hat{\gamma}_T(\theta) = [\hat{\gamma}_T'(1,\theta), ..., \hat{\gamma}_T'(H,\theta_0)]'$.

Then, the objective function becomes:

$$\begin{aligned} L_T(\theta) &= \sum_{h=1}^{H} [\hat{\gamma}_T(h,\theta)'\hat{\Omega}_T(\theta,\delta_T)^{-1}\hat{\gamma}_T(h,\theta)] \\ &= \hat{\gamma}_T'(\theta) diag[\hat{\Omega}_T(\theta,\delta_T)^{-1}]\hat{\gamma}_T(\theta) \end{aligned}$$

where $\hat{\Omega}_T(\theta,\delta_T) = \hat{\Gamma}_T(0,\theta,\delta_T) \otimes \hat{\Gamma}_T(0,\theta,\delta_T)$.

Similarly, when the autocovariances are replaced by their true values, we define: $\gamma(h,\theta_0) = vec\Gamma(h,\theta_0)$, $\gamma(\theta_0) = [\gamma'(1,\theta_0), ..., \gamma'(H,\theta_0)]'$ and $\Omega(\theta_0,\delta) = \Gamma(0,\theta_0,\delta) \otimes \Gamma(0,\theta_0,\delta)$, $\Omega(\theta_0) = \Gamma(0,\theta_0) \otimes \Gamma(0,\theta_0)$.

1. First-order conditions

The FOC are:

$$\frac{\partial L_T(\hat{\theta}_T)}{\partial \theta} = 0.$$

They are asymptotically equivalent to:

$$\hat{\gamma}'_T(\hat{\theta}_T)\, diag[\Omega(\theta_0,\delta)^{-1}]\, \frac{\partial\gamma(\theta_0)}{\partial\theta'} = o_p(1/\sqrt{T}),$$

or,

$$[\hat{\gamma}_T(\theta_0) + \frac{\partial\gamma(\theta_0)}{\partial\theta'}(\hat{\theta}_T - \theta_0)]'\, diag[\Omega(\theta_0,\delta)^{-1}]\, \frac{\partial\gamma(\theta_0)}{\partial\theta'} = o_p(1/\sqrt{T}),$$

yielding

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = -J(\theta_0,\delta)^{-1}\frac{\partial\gamma'(\theta_0)}{\partial\theta} diag[\Omega(\theta_0,\delta)^{-1}]\, \sqrt{T}\hat{\gamma}_T(\theta_0) + o_p(1),$$

with $J(\theta_0,\delta) = \frac{\partial\gamma'(\theta_0)}{\partial\theta}\, diag[\Omega(\theta_0,\delta)^{-1}]\, \frac{\partial\gamma(\theta_0)}{\partial\theta'}$.

Asymptotically, we have:

$$\sqrt{T}\hat{\gamma}_T(\theta_0) = diag[\Omega(\theta_0,0)^{1/2}]Z_T,$$

where Z_T is a standard normal vector of length $K^2H = dim(\gamma)$.

2) Expansion of the test statistic at the optimum.

We have:

$$\begin{aligned}
&TL_T(\hat{\theta}_T) = T\hat{\gamma}_T(\hat{\theta}_T)'diag[\hat{\Omega}_T(\hat{\theta}_T,\delta_T)^{-1}]\hat{\gamma}_T(\hat{\theta}_T) = \\
&[\sqrt{T}\hat{\gamma}_T(\theta_0) + \frac{\partial\gamma(\theta_0)}{\partial\theta'}\sqrt{T}(\hat{\theta}_T - \theta_0)]'diag[\Omega(\theta_0,\delta)^{-1}][\sqrt{T}\hat{\gamma}_T(\theta_0) + \frac{\partial\gamma(\theta_0)}{\partial\theta'}\sqrt{T}(\hat{\theta}_T - \theta_0)] \\
&+o_P(1).
\end{aligned}$$

Then,

$$\begin{aligned}
&\sqrt{T}\hat{\gamma}_T(\theta_0) + \frac{\partial\gamma(\theta_0)}{\partial\theta'}\sqrt{T}(\hat{\theta}_T - \theta_0) \\
\approx\ & \left\{Id - \frac{\partial\gamma(\theta_0)}{\partial\theta'}J(\theta_0,\delta)^{-1}\frac{\partial\gamma'(\theta_0)}{\partial\theta}diag[\Omega(\theta_0,\delta)^{-1}]\right\}\sqrt{T}\hat{\gamma}_T(\theta_0).
\end{aligned}$$

This transformation corresponds to a projection with respect to the inner product associated with $diag[\Omega(\theta_0,\delta)]$. We have:

$$TL_T(\hat{\theta}_T) = Z_T'\, diag[\Omega(\theta_0)^{1/2}\Omega(\theta_0,\delta)^{-1/2}](Id - P)diag[\Omega(\theta_0)^{1/2}\Omega(\theta_0,\delta)^{-1/2}]\, Z_T$$

with $P = diag[\Omega(\theta_0,\delta)^{-1/2}]\frac{\partial\gamma(\theta_0)}{\partial\theta'}J(\theta_0,\delta)^{-1}\frac{\partial\gamma'(\theta_0)}{\partial\theta}diag[\Omega(\theta_0,\delta)^{-1/2}]$.

Therefore, we deduce that the matrix:

$$diag[\Omega(\theta_0)^{1/2}\Omega(\theta_0,\delta)^{-1/2}](Id - P)diag[\Omega(\theta_0,\delta)^{-1/2}\Omega(\theta_0)^{1/2}]$$

has a rank equal to $dim(\gamma) - dim(\theta)$. It follows that the asymptotic distribution of the test statistic is a mixture of chi-square of degree 1 distributions:

$$TL_T(\hat{\theta}_T) \xrightarrow{d} \sum_{j=1}^{L}\lambda_j Z_j^2, \text{ with indepenent } Z_j^2 \sim \chi^2(1) \text{ distributed},$$

where $L = dim(\gamma) - dim(\theta)$ and $\lambda_j, j = 1, ..., L$ the non-zero eigenvalues of matrix

$$diag[\Omega(\theta_0)^{1/2}\Omega(\theta_0,\delta)^{-1/2}](Id - P)diag[\Omega(\theta_0,\delta)^{-1/2}\Omega(\theta_0)^{1/2}].$$

b) In the special case $\delta_T \to \delta = 0$, the specification test statistic (3.5) has a chi-square asymptotic distribution as given below.

Proposition 4: If $\delta_T \to 0$ when $T \to \infty$, then under the independence of $u_{0,t}'s$, the RGCov test statistic (3.5) has a chi-square distribution with the degree of freedom equal to

$K^2H - dim(\theta)$ for $dim(\theta) \geq 0$.

Proof: It follows from Proposition 3 and the asymptotic equivalence of the GCov and RGCov estimators for $\delta_T \to \delta = 0$.

When the degree of freedom is greater than 30, under the null hypothesis of independence of $u_t's$, the following function of the test statistic (3.5) is evaluated for a given H and δ, and the degree of freedom is denoted by $\nu = K^2H - dim(\theta)$ with $dim(\theta) \geq 0$:

$$\zeta(\hat{\xi}_T, \nu) = \sqrt{2\,\hat{\xi}_T} - \sqrt{2\nu - 1},$$

is asymptotically normally distributed:

$$\zeta(\hat{\xi}_T, \nu) \sim N(0, 1).$$

Hence, the null hypothesis can be alternatively tested using the function ζ and the asymptotically valid critical values of the standard normal. This approach is particularly useful for models of large dimensions n, or when a high number of nonlinear transformations J are considered.

2.3.3 Efficient Inverse Updating

The implementation of the RGCov approach requires the inversion of matrices of high dimension, equal either to the total number K of moments when computing $\hat{\Gamma}_T(0, \theta, \delta)$, or the number of parameters, when estimating the asymptotic variance-covariance matrix of the estimators. For numerical efficiency, the algorithms based on the Sherman-Morrison formula [Sherman and Morrison (1949, 1950)] can be used[7].

We first recall this formula and next explain its implementation.

Lemma 1 [Sherman and Morrison (1949), Sherman and Morrison (1950)]: If A is a

[7] The approach based on the Sherman-Morrison formula is cheaper than the computation of the inverse from either a spectral decomposition, or a Cholesky decomposition. Note also that these decompositions are not unique.

symmetric positive definite matrix, then

$$(A + xx')^{-1} = A^{-1} - \frac{A^{-1}xx'A^{-1}}{1 + x'A^{-1}x}$$

Proof: We have

$(A + xx')(A^{-1} - \frac{A^{-1}xx'A^{-1}}{1+x'A^{-1}x}) = AA^{-1} + xx'A^{-1}(1 - \frac{1}{1+x'A^{-1}x} - \frac{x'A^{-1}x}{1+x'A^{-1}x}) = I.$

The result follows. QED

The above Lemma can be used to compute recursively the matrix $C_T = [\rho_1 I + \rho_2 \sum_{t=1}^{T} x_t x_t']^{-1}$ from the sample of $t = 1, ..., T$ observations.

Corollary 1:

$$C_T = C_{T-1} - \frac{\rho_2 C_{T-1} x_T x_T' C_{T-1}}{1 + \rho_2 x_T' C_{T-1} x_T},\ T \geq 2.$$

Proof: We can apply Lemma 1 with $A = \rho_1 I + \rho_2 \sum_{t=1}^{T-1} x_t x_t'$ and $\sqrt{\rho_2} x_t$ for x. The result follows by observing that $A^{-1} = C_{T-1}$. QED

The starting value C_1 in this recursion is given in the next corollary:

Corollary 2:

$$C_1 = (\rho_1 I + \rho_2 x_1 x_1')^{-1} = \frac{1}{\rho_1} I - \frac{\rho_2}{\rho_1^2} \frac{x_1 x_1'}{1 + \frac{\rho_2}{\rho_1} x_1' x_1}.$$

Proof: The result follows by Lemma 1 applied with $A = \rho_1 I$ and $x = \sqrt{\rho_2} x_1$. QED

The recursive approach given in the above corollaries can be used to invert the matrix:

$$\hat{\Gamma}_T(0, \theta, \delta_T) = \delta_T I + \frac{1}{T} \sum_{t=1}^{T} [v_t(\theta) - \frac{1}{T} \sum_{t=1}^{T} v_t(\theta)][v_t(\theta) - \frac{1}{T} \sum_{t=1}^{T} v_t(\theta)]',$$

for given values of T, δ_T and θ. The recursive formulas are then applied with $\rho_1 = \delta_T$, $\rho_2 = \frac{1}{T}$, and $x_t = v_t(\theta) - \frac{1}{T} \sum_{t=1}^{T} v_t(\theta)$.

This inversion approach can be used when the objective function is maximized by an algorithm that requires evaluating the value of the objective function at $\hat{\theta}_T^{(j)}$, where $\hat{\theta}_T^{(j)}$ is the approximation of θ at iteration j of the algorithm.

The above corollaries can also be used to numerically compute the asymptotically efficient variance-covariance matrix of the RGCov estimator when the inverse of the Hessian matrix is approximated by the inverse of the outer product of scores.

2.4 Monte Carlo Experiment

This section evaluates the RGCov, GCov, and diagonal GCov estimators through a simulation study of a VAR(1) model with t(4) distributed errors with mean zero and a scale matrix equal to the identity matrix. We consider two scenarios where the high-dimensional $K \times K$ matrix $\Gamma(0, \theta)$ with $K = 30$ is driven either 1) by a large number of components n, or 2) by a large number of nonlinear transformations J. Below, we illustrate scenario 1) many components and a few transformations $n = 15$, $J = 2$. This design gives $K \times K = nJ \times nJ = 30 \times 30$ as a dimension of $\Gamma(0, \theta)$. In subsection 3.4.2, we consider a few components and many transformations: $n = 3$, $J = 10$. This design also gives $K \times K = nJ \times nJ = 30 \times 30$ as dimension of $\Gamma(0, \theta)$. The choices of n and J are crucial because they determine the dimension of the matrix to be inverted. The choice of H is based on the range of serial dependence in the process and has no impact on the dimension of $\Gamma(0, \theta)$, i.e., the matrix to be inverted. We set $H = 10$ in our simulation experiment illustrated below, as in the Application Section 3.5 later in the text.

2.4.1 Many Variables

This subsection evaluates the performance of the RGCov, GCov, and diagonal GCov estimators through two simulation studies, focusing on scenarios in which the $K \times K$ matrix $\Gamma(0)$ is high-dimensional. In the first simulation study, high dimensionality arises from the number of variables: we set $n = 15$, $H = 10$, $J = 2$, hence implying $K = n \times J = 30$. The second simulation study, discussed below in Section 2.4.1.1, instead considers a setting where the number of linear and nonlinear transformations drives high dimensionality. Here, we report the squared bias, MSE, and variance values for the three estimators obtained from the Monte Carlo experiment.

2.4.1.1 High Dimensionality Driven by the Number of Transformations

We evaluate here the performance of the three estimators RGCov, GCOv and Diagonal GCov in scenarios where the high dimensionality of $\Gamma(0, \theta)$ is driven by J, the number of transformations.

We assess the performance of the three estimators by analyzing bias, estimated variance, and mean squared error (MSE). In particular, to measure the bias of the coefficient in the i-th row and j-th column of the autoregressive matrix, we use the following expression:

$$\widehat{\text{Bias}}(\hat{\Phi}_{ij})^2 = \frac{1}{M}\sum_{m=1}^{M}\left(\hat{\Phi}_{ij}^{(m)} - \Phi_{ij,0}\right)^2 = \frac{1}{M}\sum_{m=1}^{M}\left[\widehat{\text{Bias}}\left(\hat{\Phi}_{ij}^{(m)}\right)\right]^2 \tag{4.1}$$

where $i, j = 1, \dots, n$; $\hat{\Phi}_{ij}^{(m)}$ denotes the estimated coefficient in the i-th row and j-th column of the autoregressive matrix obtained from the m-th replication; $\overline{\Phi}_{ij}$ represents the mean value of this coefficient across all M replications, and $\Phi_{ij,0}$ is its true population coefficient. To measure the estimated variance of the coefficient in the i-th row and j-th column of the autoregressive matrix, we use:

$$\widehat{\text{Var}}(\hat{\Phi}_{ij}) = \frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{\Phi}_{ij}^{(m)} - \overline{\Phi}_{ij}\right)^2. \tag{4.2}$$

The estimated MSE for each coefficient in the matrix Φ is given by:

$$\widehat{\text{MSE}}(\hat{\Phi}_{ij}) = \widehat{\text{Var}}(\hat{\Phi}_{ij}) + \left[\widehat{\text{Bias}}(\hat{\Phi}_{ij})\right]^2, \tag{4.3}$$

where $\widehat{\text{Bias}}(\hat{\Phi}_{ij})$ and $\text{Var}(\hat{\Phi}_{ij})$ are defined in (4.1) and (4.2), respectively. Due to the high dimensionality of Φ, reporting bias, variance, and MSE for each coefficient is impractical. Instead, we summarize the performance of the estimators—GCov, diagonal GCov, and RGCov—by averaging all the coefficients of the matrix obtained from (4.1), (4.2), and (4.3).

We consider a 3-dimensional mixed VAR(1) model as the DGP characterized by two real eigenvalues inside the unit circle, $\lambda_1 = \{0.20, 0.41\}$, and one real eigenvalue outside the unit circle, $\lambda_2 = 1.5$. We assume that the distribution of the error term is a multivariate Student-t with degrees of freedom $\nu = 4$ and a scale matrix equal to the identity matrix. For all three estimators, we set $H = 2$ and $J = 10$, resulting in a 30×30 matrix $\Gamma(0, \theta)$ to be inverted, which is of the same dimension as in Section 3.4.

The use of nonlinear transformations in the GCov is essential to correctly capture the noncausal component and identify the coefficients that result in an error term i.i.d. [8]. We follow Cubadda et al. (2024) and consider the following transformations: linear transformation $a_1\Big[g_i(y_t;\theta)\Big] = u_{i,t}$, quadratic transformation $a_2\Big[g_i(y_t;\theta)\Big] = u_{i,t}^2$, cubic transformation $a_3\Big[g_i(y_t;\theta)\Big] = u_{i,t}^3$, sign transformation $a_4\Big[g_i(y_t;\theta)\Big] = \text{sign}(u_{i,t})$, absolute value transformation $a_5\Big[g_i(y_t;\theta)\Big] = |u_{i,t}|$, cubic absolute transformation $a_6\Big[g_i(y_t;\theta)\Big] = |u_{i,t}|^3$, logarithmic transformation $a_7\Big[g_i(y_t;\theta)\Big] = \log(|u_{i,t}|)$, squared logarithmic transformation $a_8\Big[g_i(y_t;\theta)\Big] = \log(|u_{i,t}|)^2$, cubic logarithmic transformation $a_9\Big[g_i(y_t;\theta)\Big] = \log(|u_{i,t}|)^3$, and square root absolute transformation $a_{10}\Big[g_i(y_t;\theta)\Big] = |u_{i,t}|^{1/2}$. These transformations are designed to capture a variety of different dynamic features in the data. The linear transformation preserves the original data structure, while the quadratic and cubic transformations capture higher-order dependencies typical in time series with volatility clustering. Transformations such as the absolute value and sign transformations are used to further model volatility and extreme fluctuations. In particular, these transformations separate dynamic volatility from other effects, such as bid-ask bounce, by highlighting the magnitude and directionality of fluctuations [see Gourieroux and Jasiak (2023)]. The cubic absolute transformation emphasizes large values, which are crucial for capturing extreme events, while logarithmic transformations compress large values to handle diminishing returns and im-

[8] Examples of informative transformations are provided in [Gourieroux and Jasiak (2023), Cubadda et al. (2024), **?**].

prove robustness to outliers. Finally, the square-root absolute transformation helps reduce the impact of large values or extreme fluctuations in the data, making it useful for stabilizing variance.

Table 2.1: *Squared bias, variance, and MSE (each multiplied by T) for RGCov (under the settings $\delta_t = \delta$ and $\delta_t = \eta/T$), GCov, and diagonal GCov, computed across different values of T, δ, and η. Each metric is calculated element-wise and summarized by averaging over all entries of the corresponding matrices. Bold values of δ and η indicate the optimal shrinkage parameters for each metric. The high dimensionality results from the number of transformations J.*

T	**RGcov: Case $\delta_t = \delta$**				**RGcov: Case $\delta_t = \eta/T$**				**GCov**			**Diagonal GCov**		
	δ	Sq. Bias	Var	MSE	η	Sq. Bias	Var	MSE	Sq. Bias	Var	MSE	Sq. Bias	Var	MSE
200	0				0				298.33	11456.0	10909.0	0.0114	7.8286	7.7303
	0.5	0.0406	**2.0184**	**2.0569**	100	0.0406	**2.0184**	**2.0569**						
	1.0	0.0534	2.4622	2.5132	200	0.0534	2.4622	2.5132						
	1.5	**0.0246**	2.1004	2.1228	300	**0.0246**	2.1004	2.1228						
	2	0.0471	2.2012	2.2461	400	0.0471	2.2012	2.2461						
	2.5	0.0392	2.1887	2.2257	500	0.0392	2.1887	2.2257						
	3.0	0.0460	2.5999	2.6433	600	0.0460	2.5999	2.6433						
	3.5	0.0515	2.7336	2.7824	700	0.0515	2.7336	2.7824						
	4.0	0.0448	2.7083	2.7504	800	0.0448	2.7083	2.7504						
500	0				0				741.86	28230.0	28828.0	0.0066	1.3381	1.3420
	0.5	0.0258	2.8180	2.841	100	**0.0092**	**1.9084**	**1.9156**						
	1.0	**0.0229**	**2.5949**	**2.6152**	200	0.0178	2.0530	2.0688						
	1.5	0.0487	2.7858	2.8317	300	0.0238	2.6414	2.6625						
	2.0	0.0398	2.9939	3.0307	400	0.0206	3.2274	3.2448						
	2.5	0.0390	3.3423	3.3780	500	0.0229	2.5949	2.6152						
	3.0	0.0384	4.0999	4.1342	600	0.0178	3.0441	3.0588						
	3.5	0.0526	3.3909	3.4401	700	0.0348	2.8151	2.8471						
	4.0	0.0298	3.3338	3.3602	800	0.0324	2.7725	2.8021						
800	0				0				1495.6	45531.0.0	46981.0	0.0060	1.0012	1.0042
	0.5	0.0144	**1.9306**	**1.9431**	100	0.0086	2.2771	2.2812						
	1.0	0.0179	2.2508	2.2664	200	**0.0075**	1.8028	1.8085						
	1.5	0.0180	2.5149	2.5304	300	0.0191	**1.6805**	**1.6980**						
	2.0	0.0296	2.2751	2.3024	400	0.0144	1.9306	1.9431						
	2.5	0.0266	2.7658	2.7896	500	0.0098	1.8621	1.8701						
	3.0	**0.0176**	2.6159	2.6308	600	0.0141	2.3707	2.3824						
	3.5	0.0154	3.1610	3.1733	700	0.0163	2.3822	2.3961						
	4.0	0.0356	3.2324	3.2648	800	0.0179	2.2508	2.2664						

The results are based on $M = 1000$ replications with increasing sample sizes $T = (200, 500, 800)$. We estimate the model using the three aforementioned estimators and compute the element-wise squared bias, variance, and MSE, averaged over the corresponding matrix. Table 2.1 presents the results. It should be noted that the performance of the RGCov estimator is examined here under the two shrinkage settings: $\delta_T = \delta$ for the case $\delta_T \to \delta$ as T increases, and $\delta_T = \eta/T$ when considering the special case $\delta_T \to 0$. For the case $\delta_T = \eta/T$, it should be noted that as the sample size increases, a larger η is required to minimize bias. This pattern underscores the importance of selecting a smaller η in small samples to control bias. At the same time, larger samples require a larger η to prevent δ_T from becoming too small and losing its regularization effect. As in the case of $\delta_T = \delta$, variance and MSE respond differently to shrinkage than bias: while large shrinkage coefficients introduce bias by distorting estimates, they also stabilize the estimator by reducing variance and MSE.

The results clearly show the good performance of the RGCov estimator and reveal that the diagonal GCov performs well when the high dimensionality of the matrix $\Gamma(0, \theta)$ is driven by the number of nonlinear transformations J, particularly in terms of variance and MSE, especially when larger sample sizes ($T = 500, 800$) are considered. Moreover, a small shrinkage coefficient is required for RGCov not only to minimize bias but also to jointly optimize bias and MSE for $\delta_T = \delta$ and $\delta_T = \eta/T$.

We also compute the eigenvalues of the estimated autoregressive matrix in each replication to examine the percentage of correct identifications of the true DGP, characterized by one eigenvalue outside and two inside the unit circle. Table 2.2 presents the results.
We observe that the RGcov performs equally well as the diagonal GCov in terms of correct identification percentage.

Table 2.2: *Identification frequencies (in %) for the RGCov and diagonal Gcov (D. GCov) methods in correctly identifying the true process.*

Large J **:** $\boldsymbol{\delta_T = \delta}$		$\boldsymbol{\delta = 0}$	$\boldsymbol{\delta = 0.5}$	$\boldsymbol{\delta = 1}$	$\boldsymbol{\delta = 1.5}$	$\boldsymbol{\delta = 2}$	$\boldsymbol{\delta = 2.5}$	$\boldsymbol{\delta = 3}$	$\boldsymbol{\delta = 3.5}$	$\boldsymbol{\delta = 4}$	**D. GCov**
	$\boldsymbol{T = 200}$	14.0	86.1	85.9	**86.3**	85.7	84.9	85.6	85.3	86.0	89.9
	$\boldsymbol{T = 500}$	11.2	92.8	**92.9**	91.3	90.9	90.9	91.1	91.5	91.2	97.3
	$\boldsymbol{T = 800}$	11.0	**97.7**	97.3	97.0	96.5	96.8	96.8	96.5	96.2	98.8
Large J **:** $\boldsymbol{\delta_T = \frac{\eta}{T}}$		$\boldsymbol{\eta = 0}$	$\boldsymbol{\eta = 100}$	$\boldsymbol{\eta = 200}$	$\boldsymbol{\eta = 300}$	$\boldsymbol{\eta = 400}$	$\boldsymbol{\eta = 500}$	$\boldsymbol{\eta = 600}$	$\boldsymbol{\eta = 700}$	$\boldsymbol{\eta = 800}$	
	$\boldsymbol{T = 200}$	12.3	85.9	86.0	**86.8**	85.7	85.2	85.3	85.3	86.8	
	$\boldsymbol{T = 500}$	11.2	**93.8**	93.2	93.4	93.5	92.9	93.1	92.0	92.8	
	$\boldsymbol{T = 800}$	11.0	98.0	**98.4**	98.2	97.7	97.7	98.0	97.3	97.3	

2.4.2 Many Transformation

The data-generating process (DGP) is a 15-dimensional mixed VAR(1) model with 14 real eigenvalues inside the unit circle, λ_1={-0.437, -0.374, -0.360, -0.263, -0.248, -0.201, -0.162, -0.105, -0.004, 0.320, 0.277, 0.200, 0.162, 0.164}, and one real eigenvalue outside the unit circle, $\lambda_2 = 1.5$. We assume that the distribution of the error term is a multivariate Student-t with degrees of freedom $\nu = 4$ and a scale matrix equal to the identity matrix.

For all three estimators, we set $H = 10$, $J = 2$, and consider two power transformations: $a_1\Big(g_i(y_t;\theta)\Big) = u_{i,t}, a_2\Big(g_i(y_t;\theta)\Big) = u_{i,t}^2, \forall\, i, t$. By combining the linear function a_1 with the quadratic function a_2, the estimator captures both linear dynamics and volatility in the underlying data. Inverting $\Gamma(0,\theta)$ of dimension 30×30 in the GCov estimator is computationally challenging due to the curse of dimensionality. To address this issue, we investigate whether the RGCov or the diagonal GCov lead to meaningful improvements. It should be noted that the performance of the RGCov estimator is examined here under the two shrinkage settings: $\delta_T = \delta$ for the case $\delta_T \to \delta$ as T increases, and $\delta_T = \eta/T$ when considering the special case $\delta_T \to 0$.

Table 2.3: *Squared bias, variance, and MSE (each multiplied by T) for RGCov (under the settings $\delta_t = \delta$ and $\delta_t = \eta/T$), GCov, and diagonal GCov, computed across different values of T, δ, and η. Each metric is calculated element-wise and summarized by averaging over all entries of the corresponding matrices. Bold values of δ and η indicate the optimal shrinkage parameters for each metric. The high dimensionality arises from the number of variables n.*

T	**RGcov: Case $\boldsymbol{\delta_t = \delta}$**				**RGcov: Case $\boldsymbol{\delta_t = \eta/T}$**				**GCov**			**Diagonal GCov**		
	$\boldsymbol{\delta}$	Sq. Bias	Var	MSE	$\boldsymbol{\eta}$	Sq. Bias	Var	MSE	Sq. Bias	Var	MSE	Sq. Bias	Var	MSE
200	0				0				1.6216	933.44	934.12	5.1312	4004.7	4005.9
	0.5	**0.44531**	8.4826	8.9194	100	**0.44531**	8.4826	8.9194						
	1.0	0.49845	2.8199	3.3155	200	0.49845	2.8199	3.3155						
	1.5	0.52071	2.3038	2.8127	300	0.52071	2.3038	2.8127						
	2.0	0.53406	2.1684	2.7003	400	0.53406	2.1684	2.7003						
	2.5	0.54640	2.1291	**2.6734**	500	0.5464	2.1291	**2.6734**						
	3.0	0.55709	**2.1274**	2.6824	600	0.55709	**2.1274**	2.6824						
	3.5	0.56579	2.1416	2.7052	700	0.56579	2.1416	2.7052						
	4.0	0.57476	2.1613	2.7339	800	0.57476	2.1613	2.7339						
500	0				0				2.0467	453.6194	455.2125	27.905	10360.0	10377.0
	0.5	**0.86669**	71.398	72.193	100	0.84646	418.61	419.04						
	1.0	1.0846	14.389	15.459	200	0.80938	122.21	122.90						
	1.5	1.1777	6.6127	7.7839	300	0.91176	41.927	42.796						
	2.0	1.2338	3.5697	4.800	400	1.0409	20.584	21.604						
	2.5	1.2771	3.4921	4.7657	500	1.0846	14.389	15.459						
	3.0	1.3138	3.960	4.8064	600	1.1319	8.5237	9.6471						
	3.5	1.3436	3.5220	4.8621	700	1.1641	6.7842	7.9415						
	4.0	1.3692	3.5627	4.9283	800	1.1887	3.8086	4.9934						
800	0				0				13.596	2735.8	2746.7	76.411	12702.0	12766.0
	0.5	**1.0099**	192.62	193.44	100	3.8121	1431.7	1434.1						
	1.0	1.5179	37.066	38.547	200	1.1511	629.95	630.47						
	1.5	1.7452	11.694	13.428	300	0.95566	358.29	358.89						
	2.0	1.8354	6.4035	8.2325	400	1.0099	192.62	193.44						
	2.5	1.9065	4.2248	6.127	500	1.2217	101.1	102.23						
	3.0	1.9698	4.3037	6.2692	600	1.3123	71.074	72.315						
	3.5	2.0268	4.4117	6.434	700	1.4465	38.16	39.568						
	4.0	2.0759	4.505	6.5763	800	1.5179	37.066	38.547						

We perform $M = 1000$ replications with increasing sample sizes $T = (200, 500, 800)$. In each of the 1000 Monte Carlo replications, we simulate and estimate the mixed VAR(1) model of dimension 15. To compare the three aforementioned estimators, we compute the element-wise squared bias, variance, and MSE, averaged over the corresponding matrix. Table 2.3 presents the results. We observe that the RGcov performs better than the diagonal GCov and GCov in terms of bias and MSE.

2.5 Application

This section analyzes the monthly share price series of green energy companies included in the RENIXX index. The RENIXX (Renewable Energy Industrial Index) tracks the global renewable energy market, encompassing sectors such as wind, solar, bioenergy, geothermal, hydropower, electric mobility, and fuel cells. It consists of 30 companies, each of which derives more than 50% of its revenue from these sectors (see `www.iwr.de/renixx`). Companies are selected based on their float-adjusted market capitalization (i.e., the market capitalization of freely traded shares).

Giancaterini et al. (2025) identifies speculative bubbles in the RENIXX index during three major crises: the 2008 financial collapse, the COVID-19 pandemic, and the onset of the Ukraine war. Based on this analysis, we select 12 constituent stocks from the index that both offer the most extensive historical data and display signs of bubble behavior during all three episodes. These stocks are: Ballard Power Systems ($PO0_F$), Boralex Inc. Class A ($B3H_F$), Canadian Solar Inc. ($L5A_F$), Encavis AG (ECV_DE), First Solar Inc. ($F3A_F$), FuelCell Energy Inc. ($FEY2_F$), Innergex Renewable Energy Inc. ($X3IX_F$), Nordex SE ($NDX1_F$), Ormat Technologies Inc. (HNM_F), Plug Power Inc. ($PLUN_F$), VERBIO Vereinigte BioEnergie AG (VBF_F), and Verbund AG (VER_VI); see `https://www.renewable-energy-industry.com/stocks` for details. We then employ these series to construct both bubble-hedging and bubble-exploiting portfolios. The sample covers the period from January 2008 to January 2024, yielding $T = 192$ monthly observations.

Each price series is detrended using the cubic spline method.[see Hall and Jasiak (2024)] and normalized by its standard deviation. The scaled price series is shown in Figure 2.1. We fit a mixed-VAR(1) model of dimension 12 using both the GCov and RGCov estimators for comparison. For both estimators, we consider four nonlinear power transformations of errors up to order four, while the number of lags is set to $H = 10$ to accommodate the unknown serial dependence in the data. In this framework, the dimension of the covariance matrix $\Gamma(0; h)$ is 48×48.

For the RGCov estimator, we consider fixed values of δ ranging from 0.1 to 0.5 in increments of 0.1. To select the optimal value, we implemented a four-fold cross-validation procedure. Specifically, we divide the sample into four folds, using three folds to estimate the model and the remaining one to compute the in-sample MSE. This process is repeated so that each fold serves once as the validation set. We then select the value of δ that minimizes the average MSE across the four validation sets, that is, $\delta^* = 0.3$. As shown in Table 2.6, this choice not only minimizes the prediction error in the sample but also produces a more stable solution, with eigenvalues further away from the unit circle.

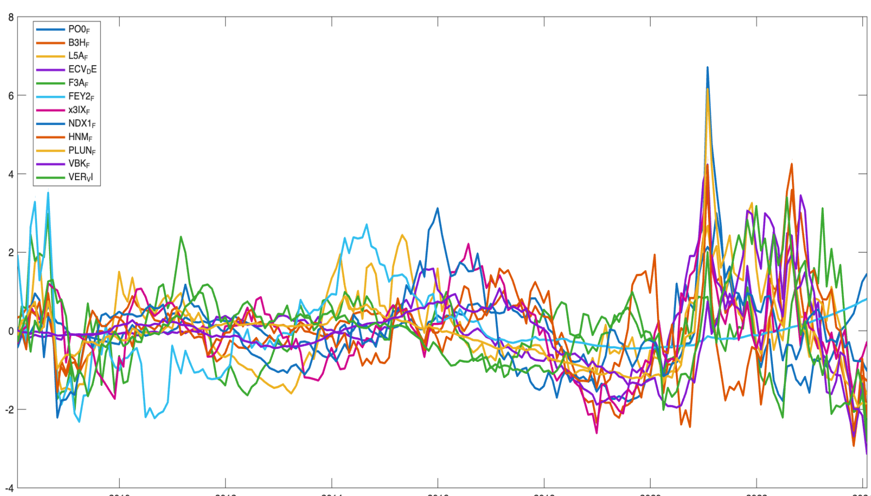


Figure 2.1: *Detrended and Scaled Monthly Prices of 12 Rennix Components*

We observe that the GCov estimator yields eigenvalues of the autoregressive matrix with moduli close to one, and the test statistic computed from the GCov estimates is much higher than that obtained with the RGCov estimator, rejecting the null hypothesis of strong white-noise errors. Another difference lies in the identification, as GCov produces 11 eigenvalues of modulus greater than one, whereas RGCov correctly yields 1, regardless of the regularization parameter. To select the regularization parameter for RGCov, we choose δ^*, which maximizes the distance of all the eigenvalues from one. We exclude $\delta = 0.1$, as it results in an eigenvalue close to zero. Based on this criterion, we select $\delta^* = 0.3$ for subsequent comparisons with GCov. Table 3.5 reports the ten lowest eigenvalues of $\hat{\Gamma}(0)$ and $\hat{\Gamma}(0, \delta^*)$ for $\delta^* = 0.3$. Figure 3.2 displays the residuals. We find that the residuals estimated from RGCov exhibit much less variation compared to those from GCov.

According to the Representation Theorem introduced in Gourieroux and Jasiak (2017), the mixed VAR process can be decomposed into latent causal and noncausal components. The causal component, linked to eigenvalues inside the unit circle, reflects the stable part of the system, while the noncausal component, associated with eigenvalues outside the unit circle, captures locally explosive behavior. In the mixed VAR, which is neither purely causal nor purely noncausal, the latent structure implies the presence of both common causal and noncausal components. In contrast, purely causal or purely noncausal processes are entirely driven by their respective dynamics, assuming a full-rank autoregressive matrix. The causal component can thus be interpreted as a stabilizing linear combination of the 12 stock price series, analogous to "bubble cointegration" [Hall and Jasiak (2024)][9]. Figure 3.2 illustrates the extracted causal and noncausal components of the VAR(1) model estimated from RGCov with $\delta^* = 0.3$ and used for portfolio management in the next section.

We now propose a portfolio management strategy based on causal and noncausal decomposition obtained from the RGCov estimated VAR. In the causal–noncausal VAR models, bubbles naturally emerge as features of a strictly stationary multivariate process. As shown by Hall and Jasiak (2024), linear combinations of green stocks that eliminate common bubbles can be estimated from the causal components. Thus, the causal components that eliminate bubbles can be viewed as 'common features' in the sense of Engle and Kozicki

[9]See also Cubadda et al. (2023) and Cubadda et al. (2019) for analyses of comovements in multiplicative VAR models.

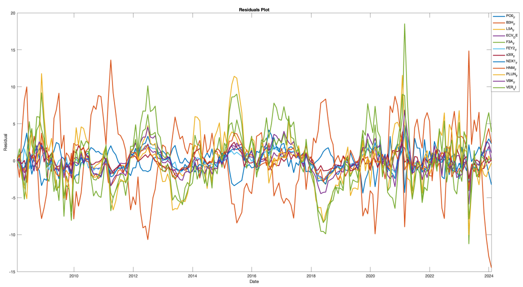
(1993), and they represent the bubble cointegration discussed in Hall and Jasiak (2024). In this setting, the causal component defines stable, bubble-free portfolios, while the noncausal component captures bubble dynamics, implying riskier but potentially more profitable investment strategies. We illustrate this by analyzing the causal and noncausal components of a mixed VAR(1) model estimated on twelve green stocks from the Renixx index. Portfolios are constructed using constant weights based on the estimated matrices A^1 and A^2 of coefficients of each component, and their cumulative returns are compared with those of the Renixx index. Specifically, we compute the returns from non-detrended prices, with weights given in Table 2.4.

We consider an investor who begins with an initial investment of $V_1 = 100$ in green stocks at time $t = 1$ and sells the holdings at the end of the month for $V_2 = V_1 + r_2$, where r_2 is the return for the first month. At time $t = 2$, the investor reinvests the full amount, V_2. This strategy continues, so that at any time $t = 1, \ldots, T$, the investment is given by $V_t = V_1 + \sum_{h=2}^{t} r_h$, with monthly liquidation and reinvestment. Negative coefficients in the allocations based on the estimated A^1 and A^2 matrices are interpreted as short positions. The initial investment V_1 is maintained constant in all portfolios. From the allocation matrices A^1 and A^2, we construct 12 portfolios: 11 causal and 1 noncausal, with a_{ij} denoting the weight of the asset j^{th} in the portfolio i^{th}. The values of portfolio weights are provided in Table 2.4.

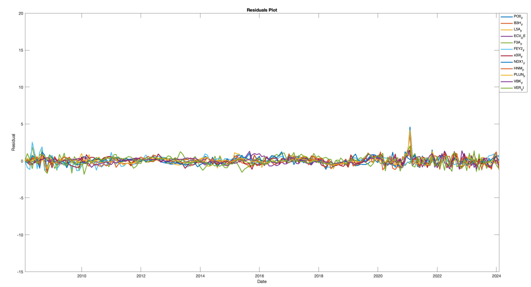
Figure 2.3 displays the cumulative returns on the noncausal portfolio, the best-performing causal portfolio among the 11 alternatives, and the Renixx index. Both proposed portfolios consistently outperform Renixx in terms of cumulative return. During bubble periods, the causal portfolio achieves higher returns, ensuring more stable performance in the long run. This portfolio analysis is based on the knowledge of true mixed-VAR(1) identification of the series; proposing portfolio management based on out-of-sample forecasts is beyond the scope of this paper.

Figure 2.2 displays the residuals obtained using GCov and RGCov and also shows the causal and non-causal components of the VAR(1) model estimated by RGCov with $\delta = 0.3$.

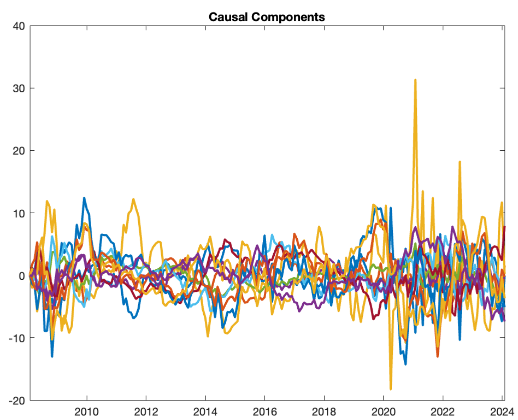
(a) GCov



(b) RGCov



(c) Causal components



(d) Non-causal component

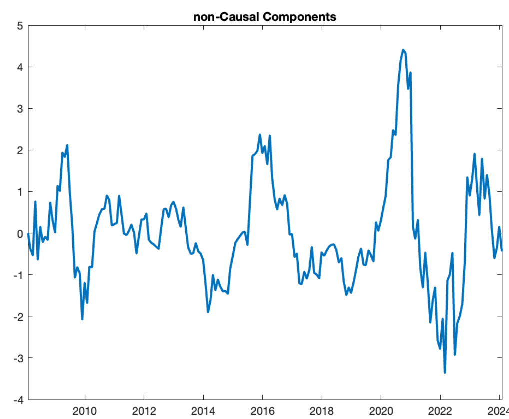


Figure 2.2: *Top row: residuals of the mixed VAR(1). Bottom row: causal and non-causal components recovered by RGCov ($\delta = 0.3$).*

We can construct 12 causal and noncausal, i.e., stable portfolios from the allocations given in A^1 and A^2. We consider a_{ij} indicates i^{th} portfolio and j^{th} asset. According to the budget constraints on V_t in each period, the weights are equal to $w_{ijt} = s_{it} * a_{ij}$ where s_{it} is:

$$s_{it} = \frac{V_{it}}{\sum_{j=1}^{12} |a_{ij}| * p_{j,t}},$$

where V_{it} is the value of the investment at time t in portfolio i and p_{jt} is the price of asset j at time t. To estimate the return of each portfolio, we have

$$r_{i,t+1} = \sum_{j=1}^{12} w_{ijt} * ln(p_{j,t+1}) - \sum_{j=1}^{12} w_{ijt} * ln(p_{j,t}).$$

$$= s_{it} \sum_{j=1}^{12} a_{ij}[(ln(p_{j,t+1}) - ln(p_{j,t})].$$

where the prices are not detrended. Consequently, the cumulative sums of returns are the summation of the returns over time. Since Renixx is an index of green stocks with positive weights, we calculate the returns on Renixx as

$$r_{Renixx,t+1} = s_{Renixx,t} * [(ln(p_{Renixx,t+1}) - ln(p_{Renixx,t})],$$

where

$$s_{Renixx,t} = \frac{V_{it}}{p_{Renixx,t}}.$$

We report the weights of the best causal portfolio (a_{5j}) and the noncausal portfolio (a_{12j}) in Table 2.4, and their cumulative returns are plotted in Figure 2.3.

Table 2.4: *Weights for portfolios*

	PO0_F	B3H_F	L5A_F	ECV_DE	F3A_F	FEY2_F	x3IX_F	NDX1_F	HNM_F	PLUN_F	VBK_F	VER_VI
noncausal	-0.1901	-0.7808	-0.2620	1.3694	0.5606	-0.1462	0.6871	0.1525	-0.2348	-0.2820	-0.4354	-0.6480
causal	-0.2485	-0.8805	0.2279	0.3961	0.4299	-1.2187	1.0462	-0.7652	0.3013	0.4975	-0.3092	0.2998

Table 2.5: *The ten lowest eigenvalues of* $\hat{\Gamma}(0)$ *estimated by GCov and* $\hat{\Gamma}(0, \delta^*)$ *with* $\delta^* = 0.3$ *estimated by RGCov.*

	Sorted eigenvalues of $\hat{\Gamma}(0)$ and $\hat{\Gamma}(0, \delta^*)$									
GCov	0.0	0.0	0.0	0.0	0.0	0.0	0.0001	0.0004	0.0006	0.0012
RGCov	0.0006	0.0740	0.2952	0.6434	0.8501	1.1095	1.2171	1.2737	1.5902	1.6305

Table 2.6: *Sorted moduli of eigenvalues of $\hat{\Phi}$ and the goodness-of-fit test statistic (ξ) for different values of δ.*

	GCov($\delta = 0$)	**RGCov** ($\delta = 0.1$)	**RGCov** ($\delta = 0.2$)	**RGCov** ($\delta = 0.3$)	**RGCov** ($\delta = 0.4$)	**RGCov** ($\delta = 0.5$)
Sorted moduli of eigenvalues of $\hat{\Phi}$						
	0.91	0.05	0.24	0.29	0.29	0.29
	0.91	0.27	0.24	0.29	0.29	0.29
	0.98	0.68	0.56	0.39	0.41	0.42
	1.02	0.68	0.64	0.70	0.70	0.70
	1.02	0.76	0.64	0.70	0.70	0.70
	1.10	0.76	0.81	0.79	0.77	0.77
	1.10	0.88	0.86	0.79	0.77	0.77
	1.50	0.88	0.86	0.90	0.93	0.94
	2.25	0.89	0.89	0.95	0.95	0.96
	2.25	0.89	0.89	0.96	0.97	0.96
	4.88	0.90	1.00	0.96	0.97	0.97
	4.88	1.06	1.08	1.08	1.08	1.08
ξ:	20 838.21	4898.71	3076.49	2219.52	1718.64	1392.39

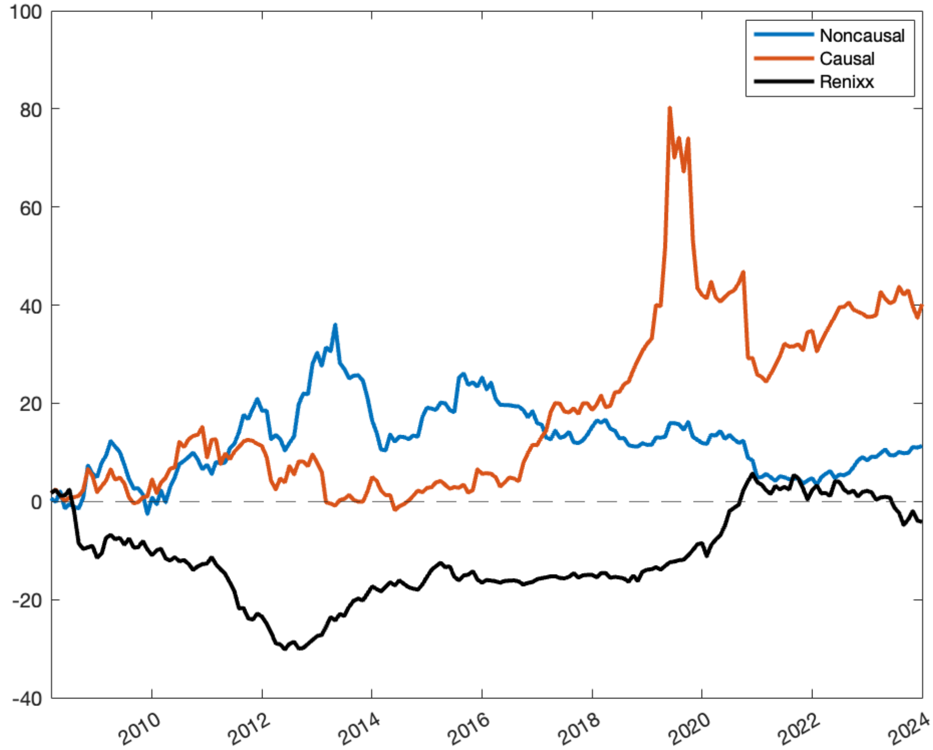


Figure 2.3: *The graph shows the cumulative returns over time for three series: the noncausal portfolio, the best-performing causal portfolio selected among 11 alternatives, and the RENIXX index*

2.6 Conclusion

This paper introduces a regularized GCov estimator to address the invertibility issue of the variance matrix in the objective function of the GCov estimator, which arises due to high-dimensional multivariate variables or a large number of nonlinear transformations. We consider the Ridge-type regularization of the variance matrix in the objective function and show that RGCov is consistent and has an asymptotic normal distribution. Moreover, we provide conditions for the RGCov estimator to reach the semiparametric efficiency bound. We also introduce the RGCov specification test based on estimated residuals and the RNLSD test for the absence of linear and nonlinear serial dependence in time series. We show that both tests have an asymptotic mixture of chi-square distributions when the regularization parameter is converging to a nonzero constant and an asymptotic chi-square distribution with known degrees of freedom when the regularization parameter is converging to zero. We explore the finite-sample properties of the RGCov estimator through simulation studies. In particular, the Monte Carlo analysis examines two sources of high dimensionality: a large number of variables and a large number of nonlinear transformations. When dimensionality is driven by the number of variables, the RGCov estimator substantially improves upon both GCov and its diagonal version, delivering smaller bias, variance, and MSE. When dimensionality is instead driven by a large number of nonlinear transformations, diagonal GCov remains competitive for moderate and large samples, as it bypasses the inversion of the full covariance matrix by construction. In this configuration, RGCov still stabilizes estimation and improves on GCov, but does not uniformly dominate the diagonal estimator.

In the empirical application, we use the RGCov estimator to estimate a 12-dimensional mixed VAR model that includes green energy stocks from the Renixx index. The green energy stocks display bubbles, evidenced in Giancaterini et al. (2025) during extreme events such as the 2008 financial crisis, the COVID-19 pandemic, and the Russian-Ukrainian war. In this paper, we use the green energy stocks to construct portfolios that both ride bubbles and hedge against bubble risk. We show that such portfolios outperform the Renixx index in terms of cumulative returns.

Chapter 3

Shrinkage Regularization for Nonlinear Serial Dependence Test

Coauthorship Statement

This chapter is based on joint work titled “Shrinkage Regularization for Nonlinear Serial Dependence Test" with Francesco Giancaterini, Alain Hecq, and Joann Jasiak, and it is work in progress. All authors contributed equally to the conceptualization, data preparation, empirical modeling, analysis, and manuscript writing.

3.1 Introduction

The nonlinear serial dependence test (NLSD hereafter) introduced by Jasiak and Neyazi (2023) is a portmanteau-type test based on the autocovariances of nonlinear functions of strictly stationary time series with non-Gaussian distributions. This approach follows from Chan et al. (2006) and has been further developed by Gourieroux and Jasiak (2023) for inference on mixed causal-noncausal models from the GCov estimator and the GCov-based specification test. In the frequency domain, a nonlinear serial dependence test was introduced by Escanciano and Lobato (2009).

The NLSD test can be used to identify nonlinear and noncausal dynamics in both univariate and multivariate time series. The test statistic is computed from nonlinear transforms $a(.)$ of a univariate or multivariate non-Gaussian process $\{X_t,\}, t = 1, ..., T$ of dimension N. The transformed vector X_t^a is of dimension higher than N, because it is augmented by K nonlinear functions of X_t, such as the squares, absolute values or logarithms, for example, satisfying the regularity conditions given in Gourieroux and Jasiak (2023). For example, let us consider K nonlinear functions $a_1, ..., a_K$ of a univariate X_t with $N = 1$. These functions transform X_t into a multivariate process of dimension $NK = K$ with components $a_k(X_t)$ as follows:

$$X_t^a = \begin{pmatrix} a_1(X_t) \\ \vdots \\ a_K(X_t) \end{pmatrix}, \tag{4.1}$$

where $a_1(X_t) = X_t$ is the vector time series itself to capture the linear dependence. Note that if X_t has no finite second-order moments, then it can be replaced by a transformed multivariate process X_t^a with finite moments. Then, from a sample of T observations on X_t, we can compute the sample autocovariance matrix of the transformations $a_k(X_t), k = 1, ..., K$:

$$\hat{\Gamma}_T^a(h) = \frac{1}{T} \sum_{t=h+1}^{T} X_t^a X_{t-h}^{a'} - \frac{1}{T} \sum_{t=1}^{T-h} X_t^a \frac{1}{T} \sum_{t=h+1}^{T} X_{t-h}^{a'}. \tag{4.2}$$

and the sample variance matrix $\hat{\Gamma}_T^a(0)$. Let us consider the following set of null hypotheses: $H_{0,a} = (\Gamma^a(h) = 0,\ h = 1, ..., H)$. To test for the absence of linear and nonlinear serial dependence, we compute the test statistic:

$$\hat{\xi}_T^a(H) = T \sum_{h=1}^{H} Tr \hat{R}_a^2(h), \tag{4.3}$$

where

$$\hat{R}_a^2(h) = \hat{\Gamma}_T^a(h) \hat{\Gamma}_T^a(0)^{-1} \hat{\Gamma}_T^a(h)' \hat{\Gamma}_T^a(0)^{-1}.$$

Under the null hypothesis of independence, the NLSD test statistic $\hat{\xi}_T^a(H)$ follows asymptotically a $\chi^2((NK)^2H)$ distribution [1].

When N or K or both are large then the matrix $\hat{\Gamma}_T^a(0)$ is of high dimension and its inverse may be difficult to compute. Two approaches in the literature address this issue. The first one consists in replacing the sample variance matrix by a matrix containing only its diagonal elements, and was proposed by Gourieroux and Jasiak (2017). The test statistics computed under this approach does not have an asymptotic chi-square distribution under the null of independence. The second is the Ridge regularization (RNLSD) test proposed by **?**, which relies on a Ridge-type regularization of the NLSD test and yields a RNLSD test statistic with an asymptotic chi-square distribution under the null of independence. **?** suggest the cross-validation for selecting the optimal regularization parameter. In this paper, we consider the shrinkage approach of Ledoit and Wolf (2004) for estimating the sample covariance matrix, when it suffers from the curse of dimensionality. The advantage of this approach is that we can estimate the shrinkage parameter in a single step, directly from the sample. The shrinkage Regularized NLSD (SR-NLSD hereafter) test has an asymptotic chi-square distribution with known degrees of freedom under the null of independence and relies on consistent estimators of the shrinkage parameters.

3.2 Linear Shrinkage Estimator of Ledoit and Wolf

Ledoit and Wolf (2004) propose a general asymptotic framework, which is reviewed below using their generic notation and allows $p = N$ to tend to infinity at a rate such that p/T stays bounded. They consider X, which is a $p \times T$ matrix of T i.i.d observations on random variables of dimension p with mean zero and variance matrix Σ. The Frobenius norm is defined as $||A|| = \sqrt{tr(AA')/p}$, based on the quadratic form of $||.||^2$ and the inner product

[1] Jasiak and Neyazi (2023) show that for an increasing set of well-selected nonlinear transformations the null hypothesis of the absence of (non)linear dependence is close to the independence hypothesis, which is the implicit null hypothesis used for the derivation of the asymptotic properties of the test.

$< A_1, A_2 >= tr(A_1 A_2')/p$. The sample covariance matrix is denoted by $S = XX'/T$. The goal is to find the linear combination of identity matrix I and sample covariance matrix S, which are the components of a regularized variance estimator $\Sigma^* = \rho_1 I + \rho_2 S$ which minimize $E[||\Sigma^* - \Sigma||^2]$, the expected distance between the regularized variance estimator and true variance and ρ_1, ρ_2 are the tuning parameters. Ledoit and Wolf (2004) identify four scalars that play important roles:

$$\mu =< \Sigma, I >, \;\; \alpha^2 = ||\Sigma - \mu I||^2, \;\; \beta^2 = E[||S - I||^2], \;\; \delta^2 = E[||S - \mu I||^2],$$

where $||.||$ denotes the Frobenius norm. For β^2 and δ^2 to be finite we need to assume X has a finite fourth moment. Theorem 2.1 of Ledoit and Wolf (2004) shows that the solution to the optimization problem of

$$\begin{aligned} \underset{\rho_1,\varphi_2}{Min} \quad & E[||\Sigma^* - \Sigma||^2] \\ \text{s.t.} \quad & \Sigma^* = \rho_1 I + \rho_2 S \end{aligned} \tag{4.1}$$

where ρ_1 and ρ_2 are non-random regularization parameters is

$$\Sigma^* = \frac{\beta^2}{\delta^2}\mu I + \frac{\alpha^2}{\delta^2} S, \tag{4.2}$$

and the objective function is:

$$E[||\Sigma^* - \Sigma||^2] = \frac{\alpha^2 \beta^2}{\delta^2}. \tag{4.3}$$

Ledoit and Wolf (2004) propose a consistent estimator of Σ^*, which is independent of the population variance Σ and based on a sample of T observations on i.i.d. variables of dimension p arranged in matrix X_T. They consider the spectral decomposition of the variance matrix into eigenvectors and eigenvalues: $\Sigma_T = G_T \Lambda_T G_T'$ where G_T is the rotation matrix whose each column is the eigenvectors of Σ_T and Λ_T is a diagonal matrix whose elements are the eigenvalues of Σ_T. Consider $Y_T = G_T' X_T$ and let $(y_{11}^T, ..., y_{p1}^T)'$ be the first column of Y_T.

To define the consistent estimator of Σ^*, we need the following assumptions:

Assumption 1. There exists a constant κ_2 independent of the number of observations such that

$$\frac{1}{p}\sum_{i=1}^{p} E[(y_{i1}^T)^8] \leq \kappa_2.$$

Assumption 2.

$$\lim_{T\to\infty} \frac{(p)^2}{T^2} \times \frac{\sum_{(i,j,k,l)\in Q_T}(Cov[y_{i1}^T y_{j1}^T, y_{k1}^T y_{l1}^T])^2}{Card(Q_T)} = 0,$$

Where Q_T is a set of all quadruples of distinct integers of 1 to p and $Card(Q_T)$ gives the number of elements in the set Q_T.

Theorem 3.2, Ledoit and Wolf (2004):

Suppose there exists a constant $\kappa_1 \in [0, \infty)$ independent of the number of observations T such that $\frac{p}{T} \to \kappa_1$. Then, under Assumptions 1 and 2, the estimator

$$S_T^* = \frac{b_T^2}{d_T^2} m_T I + \frac{a_T^2}{d_T^2} S_T, \tag{4.4}$$

is a consistent estimator of Σ_T^*,where:

$$m_T =< S_T, I >, d_T^2 = ||S_T - m_T I||^2, a_T^2 = d_T^2 - b_T^2, b_T^2 = min(d_T^2, \bar{b}_T^2),$$

and:

$$\bar{b}_T^2 = \frac{1}{T^2}\sum_{i=1}^{T} ||X_i X_i' - S_T||^2$$

where X_i is the ith column of X_T of dimension $p \times 1$.

Note that the proposed estimator S_T^* is a consistent estimator of Σ_T^* and not of Σ. From the above theorem, it follows that S_T^* is consistent estimator of Σ if

$$\frac{p}{T} \to 0$$

Then, $m_T^2 + \text{Var}^2\left(\frac{1}{p}\sum_{i=1}^{p}(y_{i1})^2\right) \rightarrow 0$, and the tunning parameter estimators $\hat{\rho}_{1,T} = \frac{b_T^2}{d_T^2} m_T$ and $\hat{\rho}_{2,T} = \frac{a_T^2}{d_T^2}$ tend to deterministic limits $\hat{\rho}_{1,T} \rightarrow \rho_1 = 0$ and $\hat{\rho}_{1,T} \rightarrow \rho_1 = 1$ [Pourahmadi (2013) Corollary 4.1].

3.3 Shrinkage Regularization of the NLSD Test

Under the null hypothesis of independence, the framework of Ledoit and Wolf (2004) can be applied to regularize the NLSD test statistic.

Let us assume that all components of the multivariate process $\{X_t^a\}, t = 1, ...T$ have mean zero or are demeaned. Then, the consecutive observations can be arranged into a matrix X_T^a of dimension $p \times T$ where $p = NK$, which replaces matrix X_T from Section 3.2. Then, under the null hypothesis of independence, X_T^a contains i.i.d observations of dimension p on random variables with mean zero and population variance matrix $\Gamma^a(0)$.
Following the approach of Ledoit and Wolf (2004) and their their generic notation we define:

$$\Gamma^{a^*}(0) = \rho_1 I + \rho_2 \Gamma^a(0).$$

Then, based on a sample of size T, we obtain a consistent estimator of $\Gamma^{a^*}(0)$ and of the tuning parameters:

$$\hat{\Gamma}_T^{a^*}(0) = \hat{\rho}_{1,T} I + \hat{\rho}_{2,T} \hat{\Gamma}_T^a(0) = \frac{b_T^2}{d_T^2} m_T I + \frac{a_T^2}{d_T^2} \hat{\Gamma}_T^a(0), \tag{4.1}$$

where:

$$m_T =< \hat{\Gamma}_T^a(0), I >, d_T^2 = ||\hat{\Gamma}_T^a(0) - m_T I||^2, a_T^2 = d_T^2 - b_T^2, b_T^2 = min(d_T^2, \bar{b}_T^2),$$

and:

$$\bar{b}_T^2 = \frac{1}{T^2} \sum_{i=1}^{T} ||X_i^a X_i^{a'} - \hat{\Gamma}_T^a(0)||^2$$

with X_i being a $p \times 1$ vector.

Definition 1: The shrinkage-regularized NLSD test statistics (SR-NLSD) is

$$\hat{\xi}^a_{SR}(H) = T \sum_{h=1}^{H} Tr \hat{R}^2_{SR}(h), \tag{4.2}$$

where

$$\hat{R}^2_{SR}(h) = \hat{\Gamma}^a_T(h) \hat{\Gamma}^{a^*}_T(0)^{-1} \hat{\Gamma}^a_T(h)' \hat{\Gamma}^{a^*}_T(0)^{-1}$$

Recall that Theorem 3.2 of Ledoit and Wolf (2004) implies that $\hat{\Gamma}^{a^*}_T(0)$ is consistent estimator of $\Gamma(0)$ if $\frac{p}{T} \to 0$, which includes the case of $p = NK$ large and constant and $T \to \infty$, considered in **?**. In that context, the autocovariance estimator maintains its standard asymptotic properties established by e.g. Chitturi (1976). Then, the tunning parameter estimators $\hat{\rho}_{1,T} = \frac{b^2_T}{d^2_T} m_T$ and $\hat{\rho}_{2,T} = \frac{a^2_T}{d^2_T}$ defined above tend to deterministic limits $\hat{\rho}_{1,T} \to \rho_1 = 0$ and $\hat{\rho}_{2,T} \to \rho_2 = 1$

Under the Assumptions 1, 2 of Ledoit and Wolf (2004), Assumption A.2 of **?**, we have the following result:

Proposition 1:

If Assumptions 1-2 and A.2 are satisfied and if $\hat{\rho}_{1,T} \to 0$ and $\hat{\rho}_{2,T} \to 1$ when p is large and $T \to \infty$ so that $\frac{p}{T} \to 0$, then under the null hypothesis of independence, the SR-NLSD test statistic $\hat{\xi}^a_{SR}(H)$ has an asymptotic chi-square distribution with a degree of freedom equal to $p^2 H$.

Proof: Based on Theorem 3.2 of Ledoit and Wolf (2004) we know that $\hat{\Gamma}^{a^*}_T(0)$ is a consistent estimator of $\Gamma^{a^*}(0) = \rho_1 I + \rho_2 \Gamma^a(0)$ which itself is a consistent estimator of $\Gamma^a(0)$ when $\frac{p}{T} \to 0$. Therefore, the asymptotic distribution follows from the asymptotic equivalence of the SR-NLSD and NLSD test statistics.

The test consists in rejecting the null hypothesis of the absence of linear and nonlinear dependence if $\hat{\xi}^a_{SR}(H) > \chi^2_{0.95}(p^2 H)$.

3.4 Simulation Studies

We investigate the empirical size of the NLSD and SR-NLSD tests. We consider two experiments. First, we generate i.i.d observations X_t for $t = 1, ..T$, where $T = 100, 200, ..., 1000$ and $N = 2, 4, ..., 20$ from a Student's t-distribution with degrees of freedom equal to 4, 7 and 10 and with variance equal to I_N. We consider $H = 1$ and $K = 2$ (level and power of the time series). Each square of Figure 4.1 plots is obtained by 1000 replications in the Monte Carlo, and in each experiment, we test the absence of linear or nonlinear dependence in the time series. We report the rate at which we reject the null hypothesis in both cases. In the second experiment, we use the same setup, but instead of increasing the number of variables, we increase the number of transformations as $K = 2, 4, ..., 20$, which includes the powers of time series up to the K and fixed the number of variables $N = 2$. Figure 4.2 illustrates the empirical size of the NLSD and SR-NLSD tests for the second experiment.

Figures 1 and 2 show that NLSD performs poorly in terms of empirical size in a high-dimensional setting with many variables or transformations. However, SR-NLSD provides an empirical size close to the nominal size. Comparing the SR-NLSD in many variables and many transformations, it is relatively more conservative in the experiment with many transformations.

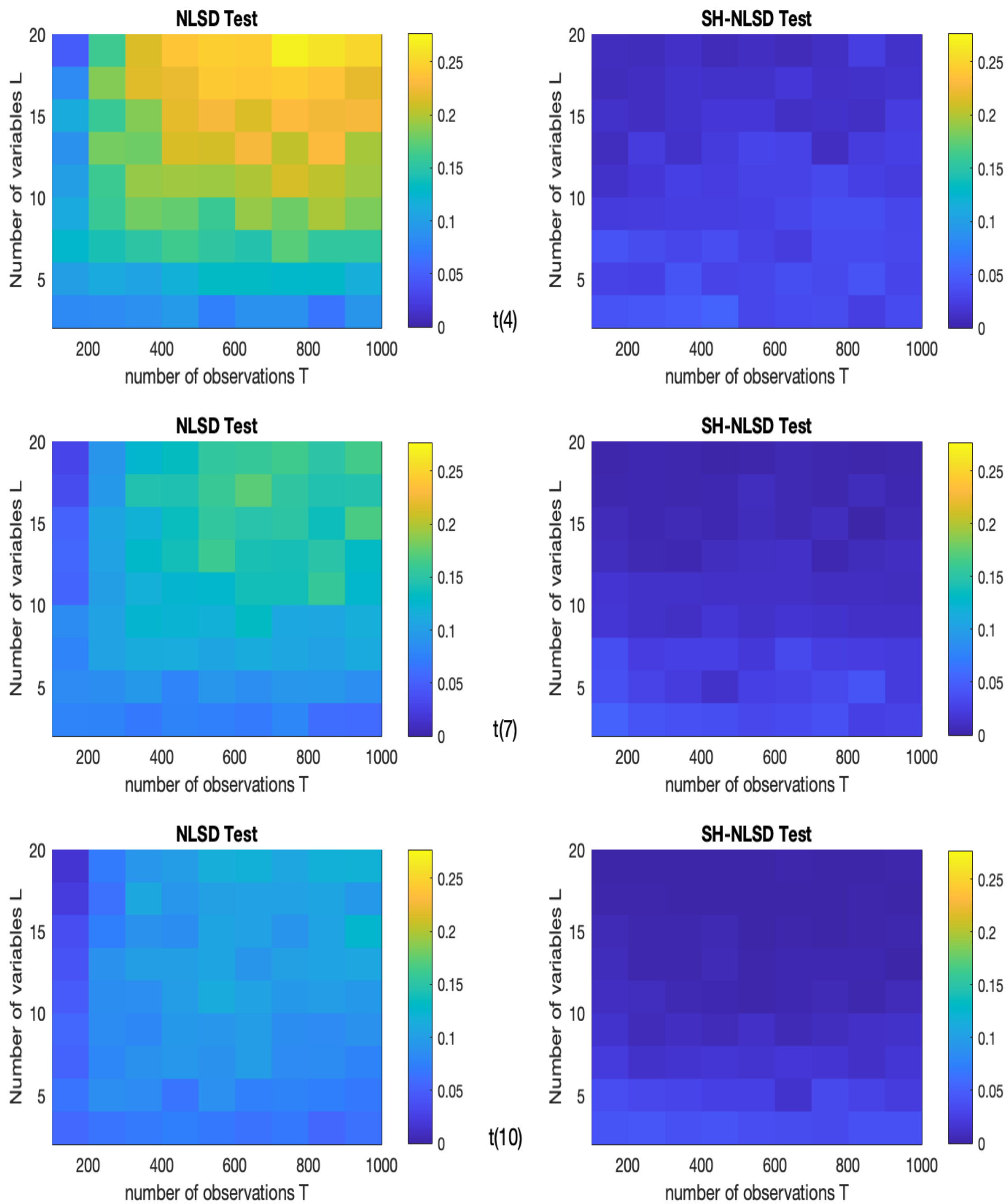


Figure 3.1: Empirical Size of NLSD test and SR-NLSD test for many variables N

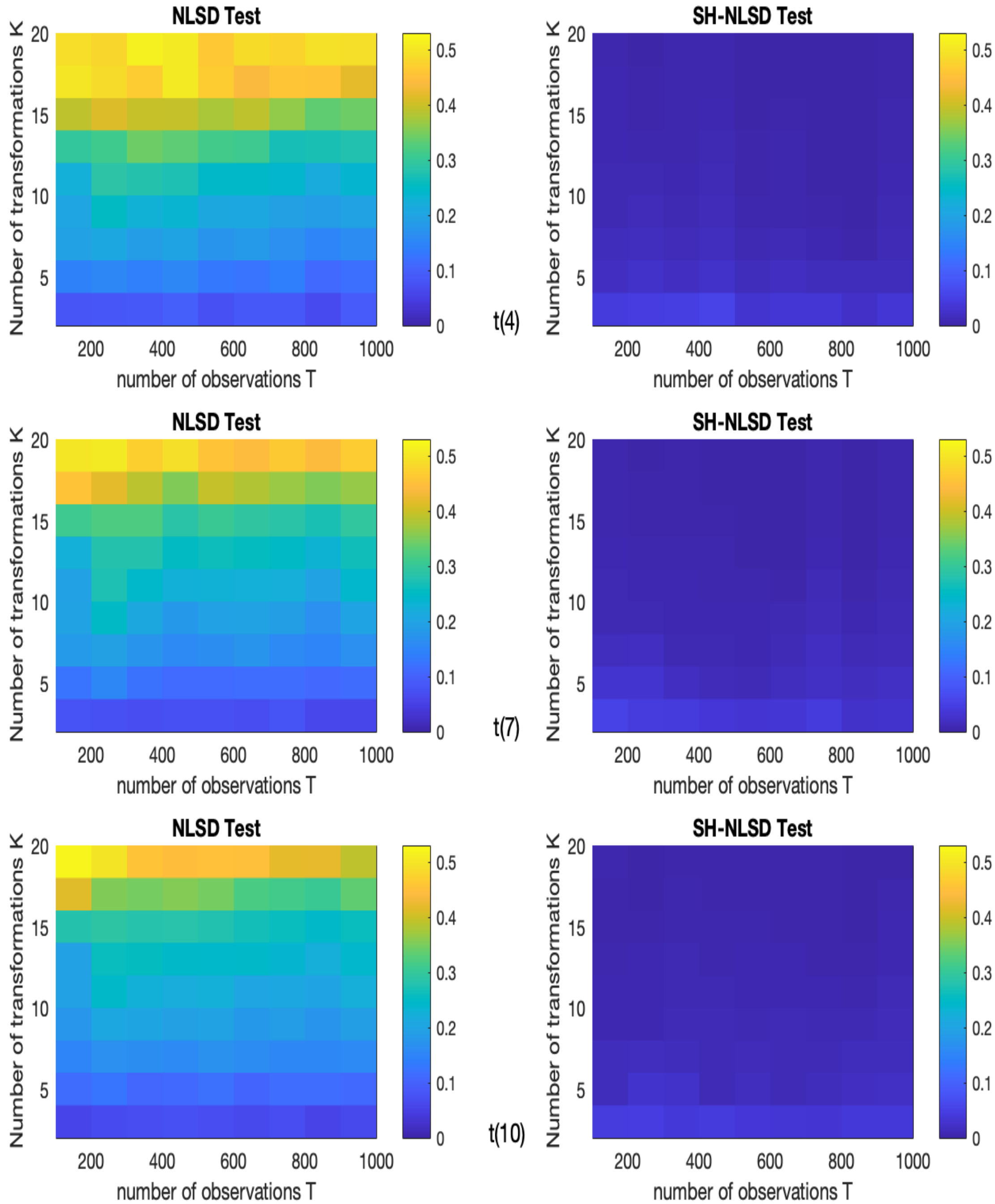


Figure 3.2: Empirical Size of NLSD test and SR-NLSD test, many transformations K

Conclusion

This dissertation develops new methods for estimation and hypothesis testing in strictly stationary non-Gaussian nonlinear time series models. The central insight is that nonlinear autocovariances provide a powerful and flexible mechanism for detecting serial dependence, evaluating model specification, and identifying dynamic structures that remain hidden under purely linear diagnostics. By building upon and extending the Generalized Covariance framework, the thesis establishes a unified approach to both testing and estimation in semi-parametric settings.

The first essay introduces the NLSD and GCov specification tests, which jointly assess linear and nonlinear serial dependence in observed data and model residuals. Theoretical results establish their asymptotic distributions under the null hypothesis and characterize their local power under drifting alternatives. These tests offer a practical and theoretically grounded alternative to multi-step residual diagnostics, particularly in models with independent non-Gaussian errors and mixed causal-noncausal dynamics.

The second essay addresses the challenges of high dimensionality by proposing the Regularized Generalized Covariance estimator. Through shrinkage regularization, the RGCov estimator improves stability and finite-sample performance when the number of variables or nonlinear transformations is large. The derived asymptotic theory and simulation evidence demonstrate that regularization can substantially reduce mean squared error and improve identification accuracy in complex multivariate systems.

The third essay extends shrinkage techniques to hypothesis testing, improving the empirical performance of portmanteau-type statistics in high-dimensional settings. By stabilizing covariance matrix estimation, the shrinkage-regularized test achieves better size control while maintaining strong power properties.

Beyond their theoretical contributions, the proposed methods have practical implications for applied econometrics. They are particularly suited to financial and macroeconomic time series characterized by heavy tails, asymmetries, and locally explosive behavior. The empirical applications illustrate how nonlinear transformations reveal dependence structures that standard linear methods overlook, emphasizing the importance of robust diagnostic tools in modern dynamic modeling.

Several avenues for future research emerge naturally from this work. Extensions to bootstrap-based inference under broader forms of misspecification, adaptive selection of nonlinear transformations, and integration with high-dimensional model selection techniques offer promising directions. Moreover, applying generalized covariance methods to structural identification problems in macroeconomics or systemic risk measurement in finance may yield further insights.

In sum, this dissertation contributes to the econometric analysis of nonlinear and non-Gaussian time series by providing theoretically sound, computationally feasible, and practically relevant tools for estimation and hypothesis testing. By expanding the role of nonlinear autocovariances in dynamic modeling, it advances our understanding of dependence structures in complex economic and financial data.

Bibliography

Adamek, R., Smeekes, S., and Wilms, I. (2023). Lasso inference for high-dimensional time series. *Journal of Econometrics*, 235(2):1114–1143.

Anderson, T. W. (1999). Asymptotic theory for canonical correlation analysis. *Journal of Multivariate Analysis*, 70(1):1–29.

Anderson, T. W. (2002). Canonical correlation analysis and reduced rank regression in autoregressive models. *The Annals of Statistics*, 30(4):1134–1154.

Andrews, D. W. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric theory*, 4(3):458–467.

Andrews, D. W. (1992). Generic uniform convergence. *Econometric theory*, 8(2):241–257.

Bickel, P. (1982). On adaptive estimation. *Annals of Statistics*, 10:647–671.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica: Journal of the Econometric Society*, pages 1443–1458.

Blasques, F., Koopman, S. J., Mingoli, G., and Telg, S. (2025). A novel test for the presence of local explosive dynamics. *Journal of Time Series Analysis*.

Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526.

Breidt, F. J., Davis, R. A., Lii, K.-S., and Rosenblatt, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*, 36(2):175–198.

Brown, B. M. (1971). Martingale central limit theorems. *The Annals of Mathematical Statistics*, pages 59–66.

Carrasco, M. and Nokho, C. (2022). Hansen-Jagannathan distance with many assets. *Available at SSRN 5078000*.

Cavaliere, G., Nielsen, H. B., and Rahbek, A. (2020). Bootstrapping noncausal autoregressions: with applications to explosive bubble modeling. *Journal of Business & Economic Statistics*, 38(1):55–67.

Chan, K.-S., Ho, L.-H., and Tong, H. (2006). A note on time-reversibility of multivariate linear processes. *Biometrika*, 93(1):221–227.

Chen, J., Gao, Z., and Tsay, R. S. (2025). Forward variable selection in ultra-high dimensional linear regression using gram-schmidt orthogonalization. *arXiv preprint arXiv:2507.04668*.

Chitturi, R. V. (1974). Distribution of residual autocorrelations in multiple autoregressive schemes. *Journal of the American Statistical Association*, 69(348):928–934.

Chitturi, R. V. (1976). Distribution of multivariate white noise autocorrelations. *Journal of the American Statistical Association*, 71(353):223–226.

Chu, B. (2023). A distance-based test of independence between two multivariate time series. *Journal of Multivariate Analysis*, 195:105151.

Cubadda, G., Giancaterini, F., and Grassi, S. (2025). Sequential monte carlo for noncausal processes. *arXiv preprint arXiv:2501.03945*.

Cubadda, G., Giancaterini, F., Hecq, A., and Jasiak, J. (2024). Optimization of the generalized covariance estimator in noncausal processes. *Statistics and Computing*, 34(4):127.

Cubadda, G. and Hecq, A. (2011). Testing for common autocorrelation in data-rich environments. *Journal of Forecasting*, 30(3):325–335.

Cubadda, G., Hecq, A., and Telg, S. (2019). Detecting co-movements in non-causal time series. *Oxford Bulletin of Economics and Statistics*, 81(3):697–715.

Cubadda, G., Hecq, A., and Voisin, E. (2023). Detecting common bubbles in multivariate mixed causal–noncausal models. *Econometrics*, 11(1):9.

Davidson, R. and MacKinnon, J. G. (2006). The power of bootstrap and asymptotic tests. *Journal of Econometrics*, 133(2):421–441.

Davis, R., Matsui, M., Mikosch, T., and Wan, P. (2018). Applications of distance correlation to time series. *Bernoulli*, 24(4A):3087–3116.

Davis, R. A. and Song, L. (2020). Noncausal vector ar processes with application to economic time series. *Journal of Econometrics*, 216(1):246–267.

De Gooijer, J. G. (2023). On portmanteau-type tests for nonlinear multivariate time series. *Journal of Multivariate Analysis*, 195:105157.

De Jong, R. M. (1998). Weak laws of large numbers for dependent random variables. *Annales d'Économie et de Statistique*, pages 209–225.

Dovonon, P. and Gonçalves, S. (2017). Bootstrapping the gmm overidentification test under first-order underidentification. *Journal of Econometrics*, 201(1):43–71.

Dovonon, P. and Gospodinov, N. (2024). Specification testing for conditional moment restrictions under local identification failure. *Quantitative Economics*, 15(3):849–891.

Dovonon, P., Hall, A. R., and Kleibergen, F. (2020). Inference in second-order identified models. *Journal of Econometrics*, 218(2):346–372.

Dvoretzky, A. (1972). Asymptotic normality for sums of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, volume 6, pages 513–536. University of California Press.

Engle, R. F. and Kozicki, S. (1993). Testing for common features. *Journal of Business & Economic Statistics*, 11(4):369–380.

Escanciano, J. C. (2007). Model checks using residual marked empirical processes. *Statistica Sinica*, 17(1):115–138.

Escanciano, J. C. and Lobato, I. N. (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics*, 151(2):140–149.

Feller, W. (1991). *An introduction to probability theory and its applications, Volume 2*, volume 2. John Wiley & Sons.

Fokianos, K. and Pitsillou, M. (2018). Testing independence for multivariate time series via the autodistance correlation matrix. *Biometrika*, 105:337–352.

Fries, S. and Zakoian, J.-M. (2019). Mixed causal-noncausal ar processes and the modelling of explosive bubbles. *Econometric Theory*, 35(6):1234–1270.

Giancaterini, F., Hecq, A., Jasiak, J., and Neyazi, A. M. (2025). Bubble detection with application to green bubbles: A noncausal approach. *arXiv preprint arXiv:2505.14911*.

Gourieroux, C., Hencic, A., and Jasiak, J. (2021a). Forecast performance and bubble analysis in noncausal mar (1, 1) processes. *Journal of Forecasting*, 40(2):301–326.

Gourieroux, C. and Jasiak, J. (2002). Nonlinear autocorrelograms: an application to inter-trade durations. *Journal of Time Series Analysis*, 23(2):127–154.

Gourieroux, C. and Jasiak, J. (2005). Nonlinear innovations and impulse responses with application to var sensitivity. *Annales d'Economie et de Statistique*, pages 1–31.

Gourieroux, C. and Jasiak, J. (2016). Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis*, 37(3):405–430.

Gourieroux, C. and Jasiak, J. (2017). Noncausal vector autoregressive process: Representation, identification and semi-parametric estimation. *Journal of Econometrics*, 200(1):118–134.

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

Gourieroux, C. and Jasiak, J. (2023). Generalized covariance estimator. *Journal of Business & Economic Statistics*, 41(4):1315–1327.

Gourieroux, C. and Jasiak, J. (2024). Generalized covariance-based inference for models set-identified from independence restrictions. *Journal of Time Series Analysis*.

Gourieroux, C., Jasiak, J., and Tong, M. (2021b). Convolution-based filtering and forecasting: An application to wti crude oil prices. *Journal of Forecasting*, 40(7):1230–1244.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017). Statistical inference for independent component analysis: Application to structural var models. *Journal of Econometrics*, 196(1):111–126.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2020). Identification and estimation in non-fundamental structural varma models. *The Review of Economic Studies*, 87(4):1915–1953.

Gouriéroux, C. and Zakoïan, J.-M. (2017). Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(3):737–756.

Guay, A. (2021). Identification of structural vector autoregressions through higher unconditional moments. *Journal of Econometrics*, 225(1):27–46.

Hall, M. K. and Jasiak, J. (2024). Modelling common bubbles in cryptocurrency prices. *Economic Modelling*, 139:106782.

Hannan, E. J. (1976). The asymptotic distribution of serial covariances. *The Annals of Statistics*, 4(2):396–399.

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

Hecq, A., Margaritella, L., and Smeekes, S. (2023). Granger causality testing in high-dimensional vars: A post-double-selection procedure. *Journal of Financial Econometrics*, 21(3):915–958.

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

Hencic, A. and Gouriéroux, C. (2015). Noncausal autoregressive model in application to bitcoin/usd exchange rates. *Econometrics of Risk*, 583:17–40.

Hotelling, H. (1992). Relations between two sets of variates. In *Breakthroughs in statistics: methodology and distribution*, pages 162–190. Springer.

Inoue, A. and Shintani, M. (2006). Bootstrapping gmm estimators for time series. *Journal of Econometrics*, 133(2):531–555.

Jasiak, J. and Neyazi, A. M. (2023). Gcov-based portmanteau test. *arXiv preprint arXiv:2312.05373*.

Keweloh, S. A. (2021). A generalized method of moments estimator for structural vector autoregressions based on higher moments. *Journal of Business & Economic Statistics*, 39(3):772–782.

Koenker, R. and Machado, J. A. (1999). Gmm inference when the number of moment conditions is large. *Journal of Econometrics*, 93(2):327–344.

Kundu, S., Majumdar, S., and Mukherjee, K. (2000). Central limit theorems revisited. *Statistics & probability letters*, 47(3):265–275.

Lanne, M. and Luoto, J. (2021). Gmm estimation of non-gaussian structural vector autoregression. *Journal of Business & Economic Statistics*, 39(1):69–81.

Lanne, M., Meitz, M., and Saikkonen, P. (2017). Identification and estimation of non-gaussian structural vector autoregressions. *Journal of Econometrics*, 196(2):288–304.

Lanne, M. and Saikkonen, P. (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, 3(3):Article 2.

Lanne, M. and Saikkonen, P. (2013). Noncausal vector autoregression. *Econometric Theory*, 29(3):447–481.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411.

Leucht, A. and Neumann, M. H. (2009). Consistency of general bootstrap methods for degenerate u-type and v-type statistics. *Journal of Multivariate Analysis*, 100(8):1622–1633.

Lof, M. and Nyberg, H. (2017). Noncausality and the commodity currency hypothesis. *Energy Economics*, 65:424–433.

Moussa, Z. and Thomas, A. (2023). Identifying oil supply news shocks and their effects on the global oil market.

Newey, W. K. (1988). Adaptive estimation of regression models via moment restrictions. *Journal of Econometrics*, 38(3):301–339.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, pages 1161–1167.

Post, E. L. (1930). Generalized differentiation. *Transactions of the American Mathematical Society*, 32(4):723–781.

Pourahmadi, M. (2013). *High-dimensional covariance estimation: with high-dimensional data*, volume 882. John Wiley & Sons.

Robinson, P. M. (1973). Generalized canonical analysis for time series. *Journal of Multivariate Analysis*, 3(2):141–160.

Sherman, J. and Morrison, W. J. (1949). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. In *Annals of Mathematical Statistics*, volume 20, pages 317–317.

Sherman, J. and Morrison, W. J. (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127.

Swensen, A. R. (2022). On causal and non-causal cointegrated vector autoregressive time series. *Journal of Time Series Analysis*, 43(2):178–196.

Truchis, G. d., Fries, S., and Thomas, A. (2025). Forecasting extreme trajectories using seminorm representations.

Velasco, C. and Lobato, I. N. (2018). Frequency domain minimum distance inference for possibly noninvertible and noncausal arma models. *The Annals of Statistics*, 46(2):555–579.

Wan, P. and Davis, R. A. (2022). Goodness-of-fit testing for time series models via distance covariance. *Journal of Econometrics*, 227:4–24.

Wooldridge, J. M. and White, H. (1988). Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric theory*, 4(2):210–230.

Appendix A

Technical Appendix for Paper 1

Appendix A

The following notation is used:

m - dimension of Y_t

u_t is of dimension $J = dim(g)$

K - dimension of transformations a

$a(u_t) = g_a$ is of dimension K

Γ or (Γ^a) is of dimension $K \times K$

$dim(\theta)$ - dimension of θ

$dim(\gamma) = 1$, γ is a scalar

Id is the Identity matrix

Asymptotic Behavior of the Portmanteau Statistic Under the Independence Hypothesis

This Appendix reviews the results that already exist in the literature and are used in the proofs of new results in Appendix B.

A.1 Asymptotic Behavior of Sample Autoregressive Coefficients

Suppose that process (Y_t) is strictly stationary and follows a VAR(1) model:

$$Y_t = \alpha + BY_{t-1} + u_t, \tag{A.1}$$

where u_t is a square integrable strong white noise, $E(u_t) = 0$, $V(u_t) = \Sigma$, where Σ is invertible and the coefficient matrix B has no eigenvalues of modulus 1. This VAR model is a SUR model with identical regressors $X_t = Y_{t-1}$ in all equations. In this case, the OLS estimators applied equation by equation are equal to the GLS estimator of B [1]. The estimator $\hat{B}' = \hat{\Gamma}(0)^{-1}\hat{\Gamma}(1)'$ is asymptotically normally distributed:

$$\sqrt{T}[vec(\hat{B}') - vecB'] \approx N[0, \Sigma \otimes \Gamma(0)^{-1}].$$

Under the null hypothesis: $H_0 = (\Gamma(1) = 0) = (B = 0)$, we have $\Sigma = \Gamma(0)$ and

$$\sqrt{T}vec(B') \sim N(0, \Gamma(0) \otimes [\Gamma(0)^{-1}]).$$

? where the $\otimes$ denotes the Kronecker product [see Chitturi (1974), eq. (1.13)].

A.2 Portmanteau Statistic as a Lagrange Multiplier test

It can be shown that the Lagrange Multiplier test statistic for testing $H_0 = (\Gamma(1) = 0) = (B = 0)$ is:

$$\hat{\xi}_T(1) = T\ Tr[\hat{\Gamma}(1)'\hat{\Gamma}(0)^{-1}\hat{\Gamma}(1)\hat{\Gamma}(0)^{-1}] = T\ Tr\hat{R}^2(1). \tag{A.2}$$

[See, e.g., Gourieroux and Jasiak (2023), Supplementary Material].

A.3 Asymptotic Behavior of Sample Autocovariance

The asymptotic distribution of $\sqrt{T}vec[\hat{\Gamma}(1)' - \Gamma(1)']$ for model (A.1) is given in Gourieroux and Jasiak (2023), Supplementary Material [see alsoChitturi (1976), Hannan (1976)]. We have:

[1]In this Appendix, the index T of the estimators is omitted to simplify the notation.

$$\sqrt{T}[\hat{\Gamma}(1)' - \Gamma(1)'] = \hat{\Gamma}(0)\sqrt{T}[\hat{B}' - B'] = \Gamma(0)\sqrt{T}[\hat{B}' - B'] + o_p(1),$$

and

$$vec[\sqrt{T}[\hat{\Gamma}(1)' - \Gamma(1)']] = vec[\Gamma(0)\sqrt{T}[\hat{B}' - B']] = [Id \otimes \Gamma(0)]vec(\sqrt{T}[\hat{B}' - B']) + o_p(1).$$

Under the null hypothesis $H_0 := (\Gamma(1) = 0) = (B = 0)$ of independently and identically distributed (i.i.d.) process (Y_t) with finite fourth order moment:

$$\begin{aligned} vec[\sqrt{T}[\hat{\Gamma}(1)' - \Gamma(1)']] &\sim N[0, [Id \otimes \Gamma(0)][\Gamma(0) \otimes \Gamma(0)^{-1}][Id \otimes \Gamma(0)]] \\ &= N[0, \Gamma(0) \otimes \Gamma(0)]. \end{aligned}$$

It follows that under this null hypothesis, the statistic (A.2) has asymptotically a chi-square distribution $\chi^2(K^2)$, where $K = m$ is the dimension of (Y_t).

A.4 Statistic Based on Several Autocovariances

The interpretation as a SUR regression can be extended to any lag H. Then, under the stationarity assumption, the VAR model becomes:

$$Y_t = \alpha + B_1 Y_{t-1} + \cdots + B_H Y_{t-H} + u_t, \tag{A.3}$$

where (u_t) is a square integrable strong white noise and the companion matrix of autoregressive coefficients has no eigenvalues of modulus 1. Under the null hypothesis of the independence of Y_t, or equivalently under $H_0 = \{B_1 = \cdots = B_H = 0\}$, the explanatory variables are orthogonal, and the OLS estimators of $B_1, ..., B_H$ are such that $\hat{B}_h$ coincides with the OLS estimator in the simple SUR model $Y_t = \alpha_h + B_h Y_{t-h} + v_t$. It follows that, under this null hypothesis, the estimators $\sqrt{T}\hat{B}_h,\ h = 1, ..., H$ are independent, normally distributed with the same distribution $N(0, \Gamma(0) \otimes \Gamma(0))$. Then, the test statistics:

$$\hat{\xi}_T(H) \approx T \sum_{h=1}^{H} vec[\sqrt{T}\hat{\Gamma}(h)']'[\hat{\Gamma}_0(0)^{-1} \otimes \hat{\Gamma}_0(0)^{-1}]vec[\sqrt{T}\hat{\Gamma}(h)'] \tag{A.4}$$

follows asymptotically the chi-square distribution $\chi^2(K^2H)$, where $K = m$ is the dimension of (Y_t).

Appendix B

Asymptotic Distribution in the Semi-Parametric Framework

This Appendix provides the regularity conditions and proofs of Propositions 1, 2, B1, and B2.

B.1 The Law of Large Numbers (LLN) for Triangular Arrays

As pointed out in Section 3.3.2, the proof of the consistency of estimated autocovariances and of the GCov estimator under the local alternatives is similar to the proof under the null hypothesis of independence. The difference is in the use of the LLN for empirical autocovariances of a triangular array of observations, uniform in θ.

Below, we provide a sufficient set of regularity conditions.

Regularity Conditions for LLN uniform in θ.

1. Conditions on the true nonlinear dynamics

i) The observations satisfy the model:

$$g^*(\tilde{Y}_{T,t}; \theta_T, \gamma_T) = u_t, \tag{B.1}$$

where the u_t's are i.i.d. with pdf f_0.

ii) The function g^* is invertible with respect to $Y_{T,t}$; then we can write:

$$Y_{T,t} = h(u_t, Y_{T,t-1}, ..., Y_{T,t-p}; \theta_T, \gamma_T). \tag{B.2}$$

iii) For each given T, $(Y_{T,t})$ with $t = 1, ..., T$, is a strictly stationary and ergodic solution of the autoregressive equation (B.2).

2. Conditions on the parameters

Suppose that the parameter space is $\Theta \times C$, where $\theta \in \Theta \subset \boldsymbol{R}^{dim(\theta)}$ and $\gamma \in C \subset \boldsymbol{R}$. We assume that:

i) Θ and C are compact sets with non-empty interiors.

ii) θ_0 is in the interior of Θ and 0 is in the interior of C.

iii) $\theta_T = \theta_0 + \mu/\sqrt{T},\ \gamma_T = \nu/\sqrt{T}$.

iv) The combined parameter vector θ, γ is identifiable in a neighborhood of $\theta_0, 0$.

In particular, for T sufficiently large, θ_T, γ_T are in the interior of Θ and C, respectively.

3. Regularity conditions on the function g^*

i) The functions $g^*_k(y; \theta, \gamma),\ k = 1, ..., K$ are continuously differentiable on the interior $\Theta \times C$.

ii) Let us define: $G^*_k(\tilde{y}) = Max_{(\theta,\gamma)\in\Theta\times C}[g^*_k(\tilde{y}, \theta, \gamma)]^2$. We assume $E_0 G^*_k(\tilde{Y}) < \infty,\ k = 1, ..., K$ where E_0 denotes the expectation computed for the process $(\tilde{Y}_t)$ associated with the "asymptotic" parameter values $(\theta_0, 0)$.

iii) Let us denote by $\mathcal{B}(\tilde{y})$ a uniform Lipschitz coefficient for functions $g^*_j(\tilde{y}; \theta, \gamma),\ j = 1, ..., J,\ g^{*2}_j(\tilde{y}; \theta, \gamma),\ j = 1, ..., J$ and for $g^*_j(\tilde{y}; \theta, \gamma),\ g^*_j(\tilde{y}_{-h}; \theta, \gamma),\ j, k = 1, ..., J, h = 1, ..., H$. In this expression $\tilde{y}$ denotes the trajectory of the process and $\tilde{y}_{-h}$ denotes this trajectory lagged by h. It is assumed that $sup_T \frac{1}{T} E|\mathcal{B}(Y_{T,t})] < \infty$ [Gourieroux and Jasiak (2023)], where the expectation is taken with respect to the distribution of process $Y_{T.} = (Y_{Tt})$, with $t = 1, ..., T$.

4. Condition of Near Epoch Dependence [De Jong (1998)]

The functions of the triangular array of random variables $Y_{T,t}, t \leq T, T \geq 1$ are L_2-NED (near epoch dependent), i.e., for $\nu(r) \geq 0$ and $c_{Tt} \geq 0$ and for all $r \geq 0$ and $t \geq 1$

$\sup_{\theta\in\Theta} E[g_j(\tilde{Y}_{T,t}, \theta) - E(g_j(\tilde{Y}_{T,t}, \theta)|Y_{T,t-r}, ..., Y_{T,t+r})]^2 \leq c_{T,t}\varphi(r)$

$\sup_{\theta\in\Theta} E[g^2_j(\tilde{Y}_{T,t}, \theta) - E(g^2_j(\tilde{Y}_{T,t}, \theta)|Y_{T,t-r}, ..., Y_{T,t+r})]^2 \leq c_{T,t}\varphi(r)$

$\sup_{\theta\in\Theta} E[g_j(\tilde{Y}_{T,t},\theta)g_k(\tilde{Y}_{T,t-h},\theta) - E(g_j(\tilde{Y}_{T,t},\theta)|Y_{T,t-r},...,Y_{T,t+r})E(g_k(\tilde{Y}_{T,t-h},\theta)|Y_{T,t-r},...,Y_{T,t+r})]^2$
$\leq c_{T,t}\varphi(r)$
for all $j,k = 1,...,K$, $h = 1,...,H$, $\varphi(r) \to 0$ as $r \to \infty$ and $\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} c_{T,t} < \infty$.

From Assumption 4, it follows that the functions $g(Y_{Tt})$ are uniformly integrable mixingales [De Jong (1998)]. Because the NED condition implies a mixingale condition, the weak LLN of Theorem 2,Andrews (1988) can be applied. Then, Theorem 4 of Andrews (1992) implies the uniform weak LLN (U-WLLN).

To summarize, we get the following Proposition:

Proposition B1:

Under the regularity conditions 1 to 4, we have:

$$plim_{T\to\infty}\hat{\Gamma}_T(h;\theta) = \Gamma_0(h;\theta)$$

uniformly in $\theta \in \Theta$ for $h = 1,...,H$, where $\Gamma_0(h;\theta)$ is evaluated at θ_0, γ_0 and f_0 is the true pdf of the error.

When the functions g are distinguished from their transforms g_a, then conditions 1 and 2 concern g and conditions 3 and 4 concern g_a.

B.2 Central Limit Theorem (CLT) for Triangular Array

We need to introduce additional regularity conditions to justify the expansion (3.13) and the asymptotic normality of the sample autocovariances $\sqrt{T}\hat{\Gamma}(h;\theta_T,\gamma_T,f_0)$ computed under a sequence of local alternatives. To obtain the corresponding CLT, we use the conditional Lindeberg-Feller conditions for martingale difference triangular array [Dvoretzky (1972), Brown (1971)] extended to the multivariate case [Kundu et al. (2000), Th. 1.3]. To apply these conditions, we first define triangular filtration and the appropriate martingales. We denote by $\mathcal{F}_{T,t}$ the information generated by the array $Y_{T,\tau},\ \tau \leq t$. Then, we consider the transformations $g_j^*(\tilde{Y}_{T,t};\theta_0,0)$, $g_j^*(\tilde{Y}_{T,t};\theta_0,0)g_k^*(\tilde{Y}_{T,t-h};\theta_0,0)$, $j,k = 1,...,K$, $h = 1,...,H$.

They can be written as a vector $G^*(\tilde{Y}_{T,t};\theta_0,0)$, say. Next, we transform this vector into a multivariate martingale difference array by considering:

$$X_{T,t} = \frac{1}{\sqrt{T}}\{G(\tilde{Y}_{T,t};\theta_0,0) - E_0[G(\tilde{Y}_{T,t};\theta_0,0)|\mathcal{F}_{T,t-1}]\}.$$

The additional regularity conditions are the following:

Regularity Conditions for the CLT

i) The multivariate martingale difference array $X_{T,t}$ has finite second-order moments.

ii) For any vector b of the same dimension as $X_{T,t}$, there exists a matrix Ω such that:

$$\sum_{t=1}^{T} E[(b'X_{T,t})^2|\mathcal{F}_{T,t-1}] \xrightarrow{P} b'\Omega b.$$

iii) Conditional Lindeberg-Feller condition:

$$\sum_{t=1}^{T} E\{(b'X_{T,t})^2 \boldsymbol{1}_{|b'X_{T,t}|>\epsilon}|\mathcal{F}_{T,t-1}\} \xrightarrow{P} 0, \text{ for any } b \text{ and } \epsilon > 0.$$

These regularity conditions ensure that the sum $S_T = \sum_{t=1}^{T} X_{T,t}$ tends in distribution to the multivariate Gaussian distribution $N(0,\Omega)$. Then we get the asymptotic normality of the estimated autocovariances under the sequence of local alternatives by applying the Slutsky Theorem.

Proposition B2: Under the sequence of local alternatives and the regularity conditions 1-5, the vectors $vec[\sqrt{T}\hat{\Gamma}_T(h;\theta_T,\gamma_T,f_0)]$ are asymptotically independent, normally distributed with mean $\Delta(h;\theta_0,f_0,\mu,\nu)$ defined in (3.10) and variance-covariance matrix $\Gamma_0(0,\theta_0) \otimes \Gamma_0(0,\theta_0)$.

Thus, the behavior of the estimated autocovariances differs from the behavior under the null by the presence of the asymptotic bias measured by $\Delta(h;\theta_0,f_0,\mu,\nu)$.

We have introduced a set of regularity conditions to derive the asymptotic behavior of the estimated autocovariances. Let us now explain why this set of conditions is also sufficient to derive the asymptotic behavior of the GCov estimator and of the GCov specification test statistic. First, we review the standard expansions under the null hypothesis. Next, we derive their analogues under the sequence of local alternatives, before applying the CLT to the estimated autocovariances of a triangular array.

B.3 First-order Expansion of the GCov Estimator under the Null Hypothesis

Below we recall the results under the null hypothesis derived in Gourieroux and Jasiak (2023). Let us consider $H = 1$ for ease of exposition. The first-order conditions of the GCov estimator are

$$\frac{\partial Tr\hat{R}^2(1;\theta_j)}{\partial \theta_j} = 0, \quad j = 1, ..., J = dim(\theta),$$

Let us define:

$$A(\theta_0) = 2\frac{\partial vec\Gamma(1;\theta_0)'}{\partial \theta}[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}],$$

and

$$J(\theta_0) = -2\frac{\partial vec\Gamma(1;\theta_0)'}{\partial \theta}[\Gamma(0;\theta_0)^{-1} \otimes \Gamma(0;\theta_0)^{-1}]\frac{\partial vec\Gamma(1;\theta_0)}{\partial \theta}.$$

The first-order Taylor series expansion of the GCov estimator is:

$$\sqrt{T}(\hat{\theta_T} - \theta_0) = J(\theta_0)^{-1}A(\theta_0)vec[\sqrt{T}\hat{\Gamma}_T(1;\theta_0)'] + o_p(1), \tag{B.3}$$

B.4 Expansion of the Portmanteau Statistic under the Null Hypothesis

The expansion of the test statistic under the null hypothesis is:

$$\hat{\xi}_T(H) = \sum_{h=1}^{H} vec[\sqrt{T}\hat{\Gamma}_T(h,\theta_0,f_0)']'\Pi(h;\theta_0,f_0)vec[\sqrt{T}\hat{\Gamma}_T(h,\theta_0,f_0)'] + o_p(1), \tag{B.4}$$

[See Gourieroux and Jasiak, "Generalized Covariance Estimator Supplemental Material"

(2023), equation (a.11)], where:

$$\begin{aligned}\Pi(h;\theta_0,f_0) &= [\Gamma_0(0,\theta_0,f_0)^{-1}\otimes\Gamma_0(0,\theta_0,f_0)^{-1}] - [\Gamma_0(0,\theta_0,f_0)^{-1}\otimes\Gamma_0(0,\theta_0,f_0)^{-1}]\frac{\partial vec\Gamma(h,\theta_0,f_0)}{\partial\theta'} \\ &\left\{\frac{\partial vec\Gamma(h,\theta_0,f_0)'}{\partial\theta}[\Gamma_0(0,\theta_0,f_0)^{-1}\otimes\Gamma_0(0,\theta_0,f_0)^{-1}]\frac{\partial vec\Gamma(h,\theta_0,f_0)}{\partial\theta}\right\}^{-1} \\ &\times\frac{\partial vec\Gamma(h,\theta_0,f_0)'}{\partial\theta'}[\Gamma_0(0,\theta_0,f_0)^{-1}\otimes\Gamma_0(0,\theta_0,f_0)^{-1}].\end{aligned}$$

Matrix $\Pi(h;\theta_0,f_0)$ satisfies for all $h=1,...,H$ the condition:

$$\Pi(h;\theta_0,f_0)V_{asy}[\sqrt{T}\hat{\Gamma}_T(h,\theta_0)']\Pi(h;\theta_0,f_0)=\Pi(h;\theta_0,f_0),$$

where $V_{asy}[\sqrt{T}\hat{\Gamma}_T(h,\theta_0,f_0)']=[\Gamma_0(0,\theta_0,f_0)\otimes\Gamma_0(0,\theta_0,f_0)]$.

This condition means that the matrix $\Pi(h;\theta_0,f_0)$ has an interpretation in terms of an orthogonal projector. Therefore, under the null hypothesis, the quadratic form (A.10), where the $vec(\sqrt{T}\hat{\Gamma}_T(h;\theta_0,f_0))$ are independent identically distributed, still follows a chi-square distribution with a reduced degree of freedom.

Appendix C

This Appendix provides additional results based on the simulations and empirical application.

C.1 Simulations for NLSD and GCov Test

Table A.1 presents the comprehensive results for the empirical size and power of NLSD.

Table A.1: NLSD Test of the absence of (non)linear dependence against the fixed alternative of MAR(0,1) at 5% significance level: size and power

γ	T=100			T=200			T=500		
	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)
0.0	0.0414	0.0484	0.0512	0.0450	0.0502	0.0518	0.0496	0.0540	0.0480
0.1	0.0878	0.0706	0.0684	0.1502	0.1360	0.1220	0.3760	0.3390	0.3572
0.2	0.2682	0.2202	0.2240	0.5618	0.5390	0.5322	0.9586	0.9546	0.9570
0.3	0.6014	0.5742	0.5674	0.9202	0.9266	0.9256	1	0.9998	1
0.4	0.8754	0.8814	0.8754	0.9966	0.9988	0.9976	1	1	1
0.5	0.9796	0.9822	0.9856	1	1	1	1	1	1
0.6	0.998	0.9986	0.9998	1	1	1	1	1	1
0.7	1	1	1	1	1	1	1	1	1
0.8	1	1	1	1	1	1	1	1	1
0.9	1	1	1	1	1	1	1	1	1

The first row ($\gamma = 0$) shows the size of the test and the remaining rows show the size-adjusted power against fixed alternatives.

Table A.2 illustrates the power of the GCov specification test for the values of γ between 0.1 and 0.9 under the fixed alternative.

Table A.2: Power of GCov specification test of MAR(0,1) against the fixed alternative of MAR(1,1) at 5% significance level

ψ	$\gamma = \phi$	T=100			T=200			T=500		
		Uniform	Laplace	t(5)	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)
0.3	0.1	0.0226	0.0528	0.0446	0.0394	0.0672	0.0616	0.0572	0.0758	0.0752
	0.2	0.0276	0.0626	0.0568	0.0720	0.0956	0.0980	0.1986	0.1624	0.1880
	0.3	0.0416	0.0850	0.0866	0.1364	0.1612	0.1748	0.5244	0.3898	0.4162
	0.4	0.0568	0.1080	0.1242	0.1796	0.2286	0.2394	0.6136	0.5280	0.5590
	0.5	0.0604	0.1238	0.1346	0.2174	0.2568	0.2696	0.8000	0.6812	0.7002
	0.6	0.0642	0.1276	0.1332	0.2504	0.3070	0.3216	0.8286	0.7898	0.7938
	0.7	0.0746	0.1460	0.1518	0.2796	0.3614	0.3796	0.8722	0.8644	0.8634
	0.8	0.1016	0.1724	0.1788	0.3404	0.4468	0.4672	0.9092	0.9282	0.9300
	0.9	0.2100	0.2478	0.2552	0.5486	0.6180	0.6174	0.9804	0.9846	0.9862
0.7	0.1	0.0182	0.0500	0.0378	0.0374	0.0702	0.0664	0.1174	0.1224	0.1140
	0.2	0.0296	0.0718	0.0628	0.1012	0.1456	0.1454	0.5346	0.4464	0.4602
	0.3	0.0656	0.1430	0.1526	0.2854	0.3712	0.3764	0.8908	0.8630	0.8654
	0.4	0.1578	0.3120	0.3098	0.5718	0.7040	0.6852	0.9926	0.9950	0.9908
	0.5	0.3392	0.5700	0.5490	0.8318	0.9316	0.9196	1	1	0.9996
	0.6	0.6034	0.8144	0.7848	0.9690	0.9914	0.9898	1	1	1
	0.7	0.8528	0.9424	0.9370	0.9990	0.9996	1	1	1	1
	0.8	0.9680	0.9882	0.9896	1	1	1	1	1	1
	0.9	0.9970	0.9980	0.9990	1	1	1	1	1	1

The following Tables A.3 and A.4 provide additional simulation results on the NLSD and GCov specification tests, respectively. Table A.3 provides the results on the size of the NLSD test for local alternatives and δ increasing from 0 to 0.9. Table A.4 shows the size of the GCov specification test for fixed alternatives and different values of ψ.

Table A.3: Test of the absence of (non)linear dependence (MAR(0,0)) against local MAR(0,1) alternatives at 5% significance level: size and size-adjusted power

$\gamma_T = \frac{\delta}{\sqrt{T}}$	T=100			T=200			T=500		
	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)
$\frac{0}{\sqrt{T}}$	0.0414	0.0484	0.0512	0.0450	0.0502	0.0518	0.0496	0.0540	0.0480
$\frac{0.1}{\sqrt{T}}$	0.0498	0.0494	0.0470	0.0496	0.0490	0.0504	0.0512	0.0494	0.0508
$\frac{0.2}{\sqrt{T}}$	0.0498	0.0488	0.0460	0.0502	0.0496	0.0496	0.0528	0.0492	0.0514
$\frac{0.3}{\sqrt{T}}$	0.0516	0.0474	0.0464	0.0506	0.0524	0.0486	0.0544	0.0504	0.0540
$\frac{0.4}{\sqrt{T}}$	0.0526	0.0486	0.0474	0.0530	0.0536	0.0500	0.0564	0.0520	0.0564
$\frac{0.5}{\sqrt{T}}$	0.0574	0.0502	0.0486	0.0556	0.0568	0.0524	0.0606	0.0550	0.0602
$\frac{0.6}{\sqrt{T}}$	0.0596	0.0530	0.0492	0.0606	0.0610	0.0548	0.0648	0.0584	0.0642
$\frac{0.7}{\sqrt{T}}$	0.0650	0.0552	0.0520	0.0664	0.0662	0.0606	0.0710	0.0620	0.0710
$\frac{0.8}{\sqrt{T}}$	0.0714	0.0580	0.0564	0.0756	0.0742	0.0652	0.0772	0.0664	0.0774
$\frac{0.9}{\sqrt{T}}$	0.0804	0.0640	0.0622	0.0838	0.0812	0.0714	0.0860	0.0758	0.0848

The first row ($\psi = 0$) shows the size of the test and the remaining rows show the power against local alternatives with $\gamma_T = \psi_T = \frac{\delta}{\sqrt{T}}$.

Table A.4: Size of GCov specification test of MAR(0,1) at 5% significance level

ψ	T=100			T=200			T=500		
	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)
0.1	0.0212	0.0404	0.0378	0.0342	0.0538	0.0514	0.0408	0.0538	0.0544
0.2	0.0216	0.0414	0.0390	0.0344	0.0548	0.0534	0.0414	0.0558	0.0560
0.3	0.0224	0.0454	0.0386	0.0348	0.0566	0.0534	0.0406	0.0544	0.0560
0.4	0.0214	0.0460	0.0390	0.0344	0.0560	0.0522	0.0414	0.0548	0.0540
0.5	0.0194	0.0450	0.0382	0.0326	0.0550	0.0518	0.0414	0.0548	0.0540
0.6	0.0196	0.0428	0.0348	0.0318	0.0536	0.0492	0.0426	0.0560	0.0544
0.7	0.0200	0.0400	0.0338	0.0298	0.0528	0.0468	0.0408	0.0552	0.0528
0.8	0.0202	0.0392	0.0316	0.0296	0.0512	0.0454	0.0396	0.0532	0.0500
0.9	0.0178	0.0384	0.0310	0.0278	0.0508	0.0428	0.0350	0.0494	0.0498

C.2 GCov Specification Test for Processes with Cauchy Error Distribution

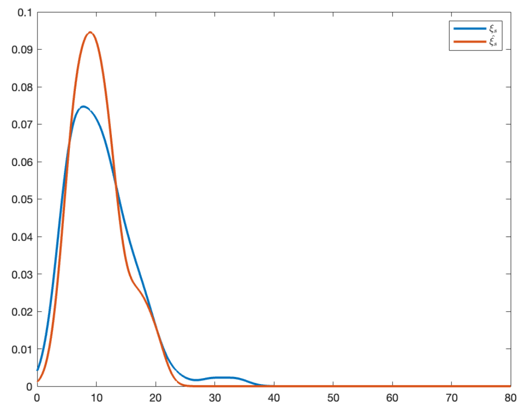
Table A.5: Size and power of GCov specification test of the MAR(0,1) with Cauchy error distribution at 5% significance level

S./P.	ϕ	ψ	T				
			100	200	300	400	500
S.		0.3	0.0306	0.0468	0.0472	0.0552	0.0538
		0.7	0.0342	0.0440	0.0470	0.0558	0.0518
P.	0.8	0.3	0.7526	0.9724	0.9990	1	1
	0.8	0.7	0.9994	1	1	1	1

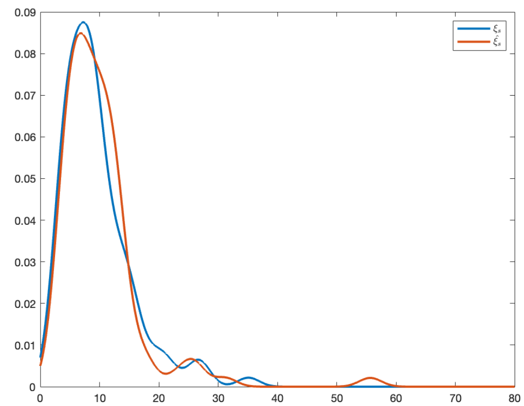
Table A.6: Size and power of bootstrap test with and without replacement, MAR(0,1) with Cauchy error distribution at 5% significance level with estimation by OLS

S./P.	ϕ	ψ	T=100		T=200		T=500	
			with	without	with	without	with	without
S.		0.3	0.077	0.080	0.056	0.078	0.050	0.073
		0.7	0.075	0.091	0.054	0.087	0.048	0.077
P.	0.8	0.3	0.808	0.802	0.880	0.915	0.986	0.988
	0.8	0.7	0.999	1.000	1.000	1	1	1

C.3 Distribution of GCov bootstrap test statistic



(a) $T = 100$



(b) $T = 500$

Figure A.1: Comparison of the distributions of ξ_s and $\hat{\xi}_s$ for MAR(0,1) with t(5) error distribution and $\psi = 0.7$

Table A.7: Bootstrap approximation size and power of GCov specification based on GCov estimation of MAR(0,1), at 5% significance level **A:** without replacement and **B:** with replacement

Panel	S./P.	ϕ	ψ	T=100			T=200			T=500		
				Uniform	Laplace	t(5)	Uniform	Laplace	t(5)	Uniform	Laplace	t(5)
A	S.	0	0.3	0.051	0.067	0.064	0.060	0.056	0.053	0.063	0.061	0.046
		0	0.7	0.052	0.066	0.066	0.062	0.054	0.053	0.064	0.064	0.046
	P.	0.8	0.3	0.227	0.361	0.365	0.454	0.529	0.566	0.946	0.936	0.941
		0.8	0.7	0.982	0.990	0.998	1	1	1	1	1	1
B	S.	0	0.3	0.052	0.068	0.060	0.062	0.060	0.047	0.056	0.065	0.046
		0	0.7	0.058	0.066	0.062	0.054	0.064	0.050	0.051	0.066	0.042
	P.	0.8	0.3	0.219	0.349	0.375	0.456	0.540	0.589	0.936	0.936	0.950
		0.8	0.7	0.985	0.989	0.996	1	1	1	1	1	1

S.: size, P.: power

Table A.8: Size and power of GCov bootstrap test without replacement based on AML estimation of MAR(0,1), at 5% significance level

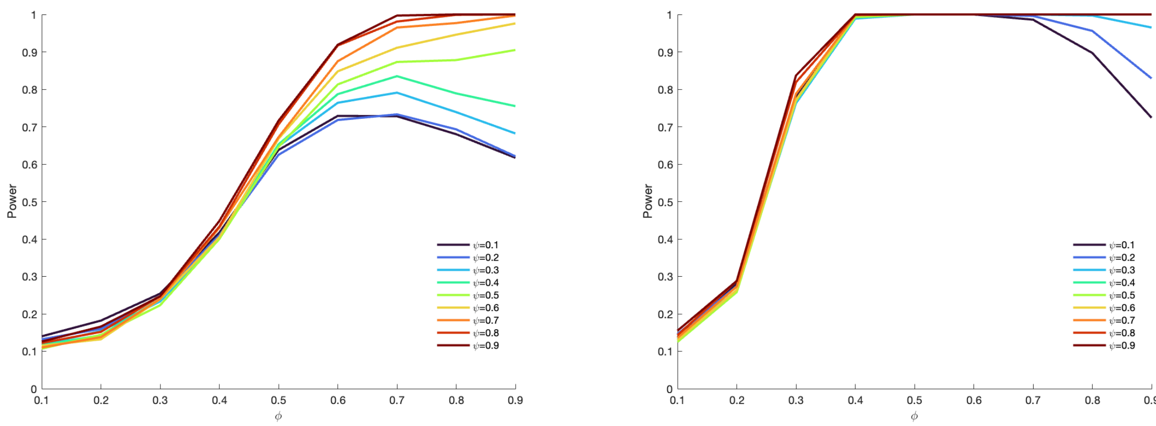
S./P.	ϕ	ψ	T=100			T=200			T=500		
			t(4)	t(5)	t(6)	t(4)	t(5)	t(6)	t(4)	t(5)	t(6)
S.	0	0.3	0.061	0.062	0.063	0.059	0.046	0.049	0.062	0.046	0.058
	0	0.7	0.063	0.063	0.066	0.060	0.047	0.056	0.062	0.049	0.060
P.	0.8	0.3	0.438	0.412	0.399	0.706	0.671	0.633	0.982	0.986	0.980
	0.8	0.7	0.991	0.992	0.993	1	1	1	1	1	1

S.: size, P.: power

Table A.9: Empirical size and power of GCov test with many transformations at 5% significance level. For power, we generate MAR(1,1) with $\phi = 0.8$.

S./P.	H	T	$K=7$		$K=8$		$K=9$	
			$\psi=0.3$	$\psi=0.7$	$\psi=0.3$	$\psi=0.7$	$\psi=0.3$	$\psi=0.7$
S.	$H=3$	$T=100$	0.007	0.012	0.028	0.031	0.001	0.002
		$T=200$	0.012	0.022	0.030	0.042	0.029	0.024
		$T=500$	0.033	0.051	0.035	0.051	0.054	0.050
	$H=4$	$T=100$	0.013	0.006	0.027	0.029	0.002	0.006
		$T=200$	0.010	0.016	0.030	0.028	0.021	0.019
		$T=500$	0.031	0.049	0.034	0.059	0.050	0.056
	$H=5$	$T=100$	0.009	0.006	0.020	0.023	0.002	0.002
		$T=200$	0.010	0.019	0.038	0.025	0.026	0.025
		$T=500$	0.031	0.053	0.029	0.052	0.053	0.060
P.	$H=3$	$T=100$	0.549	0.503	0.201	0.356	0.139	0.213
		$T=200$	0.703	0.771	0.231	0.318	0.454	0.665
		$T=500$	0.703	0.897	0.535	0.529	0.567	0.877
	$H=4$	$T=100$	0.498	0.467	0.198	0.327	0.141	0.193
		$T=200$	0.695	0.752	0.215	0.307	0.437	0.626
		$T=500$	0.664	0.889	0.528	0.514	0.571	0.872
	$H=5$	$T=100$	0.440	0.406	0.179	0.320	0.120	0.168
		$T=200$	0.680	0.723	0.194	0.298	0.428	0.594
		$T=500$	0.673	0.891	0.493	0.486	0.552	0.863

C.4 Empirical Power of Specification Test Based on Many Transformations



(a) $T = 500$ (b) $T = 5000$

Figure A.2: Empirical power of specification test based on $H = 4$ and $K = 7$ as a function of coefficient ϕ under the alternative with $t(5)$ error distribution.

Figure A.2 provides empirical power of the specification test with many transformations. We generate MAR(1,1) and fit MAR(0,1) to the process; therefore, ϕ is our distance to the null hypothesis. For high persistence of noncausality, the power converges to one with both $T = 500$ and $T = 5000$. In both sample sizes, power decreases when ψ is low and ϕ gets closer to the unit root. Moreover, Figure A.3 demonstrates the convergence of empirical power to one when we increase the sample size. The more persistence in noncausality, the faster we converge to one.

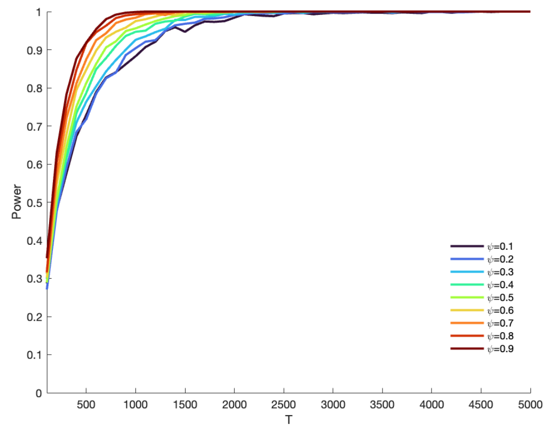
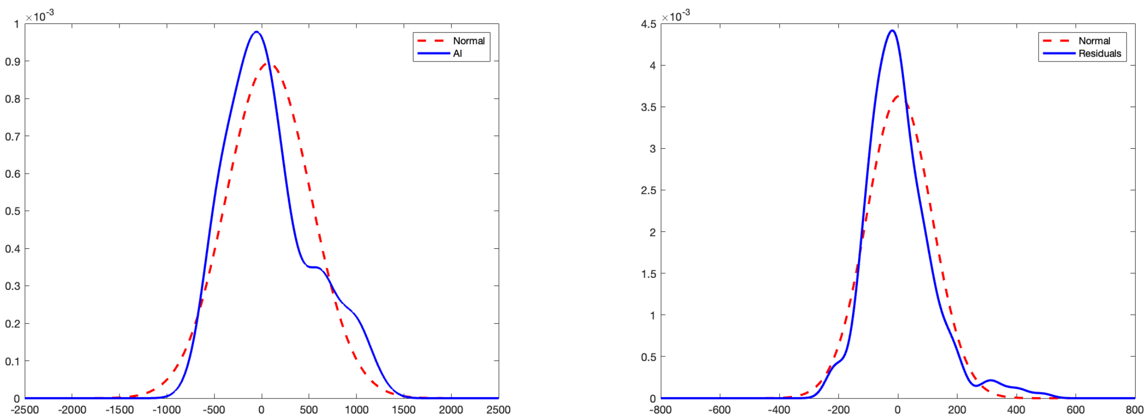


Figure A.3: Empirical power of specification test based on $H = 4$ and $K = 7$ with fixed $\phi = 0.6$ as a function of the number of observation T with $t(5)$ error distribution.

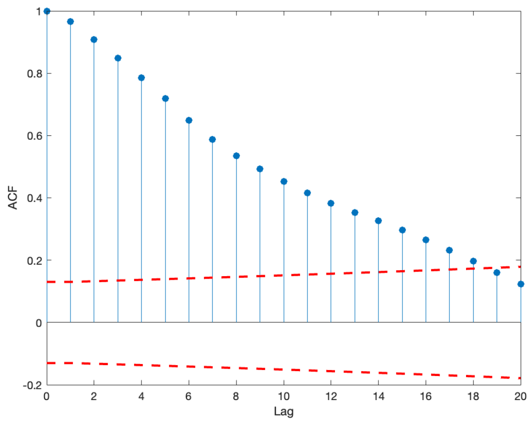
C.5 Empirical Application



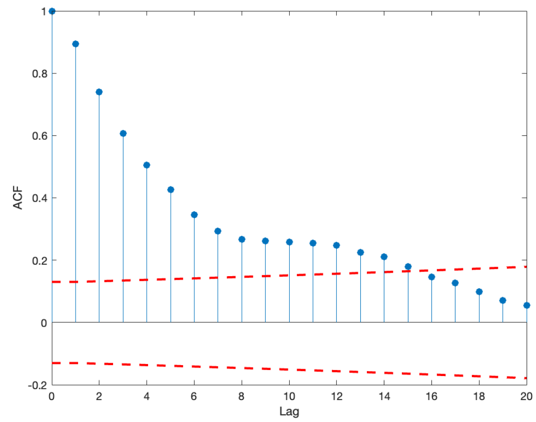
(a) Demeaned-Detrended aluminum price

(b) MAR(1,1) residuals

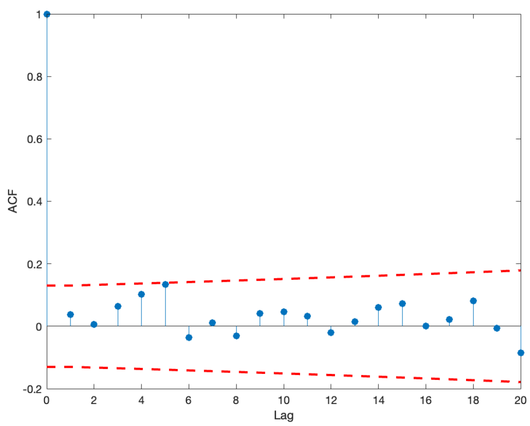
Figure A.4: Densities of demeaned aluminum price and MAR(1,1) residuals, compared with the normal density



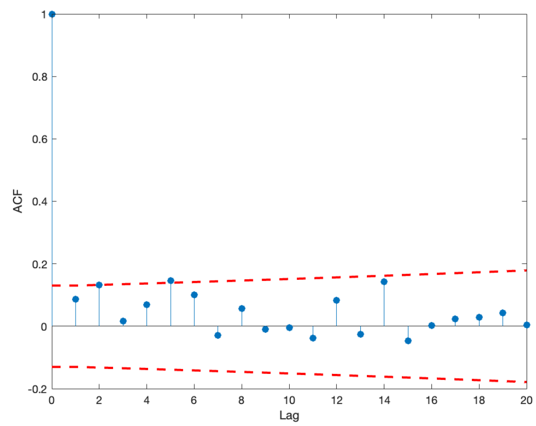
(a) ACF of the series



(b) ACF of the series square



(c) ACF of the residuals



(d) ACF of the square of the residuals

Figure A.5: ACF of aluminum prices and squared prices (panels a and b) and MAR(1,1) residuals and squared residuals (panels c and d)

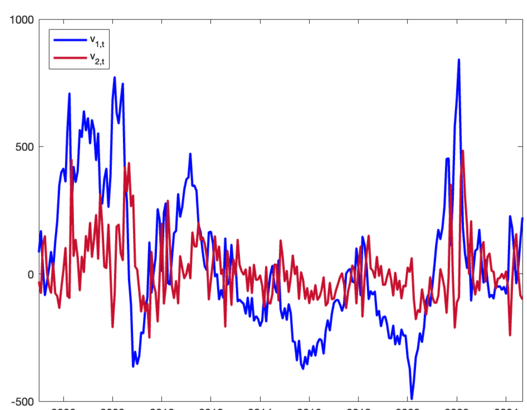


Figure A.6: MAR(1,1) causal-noncausal components