# Essays on Noncausal and Noninvertible Time Series 

 byWeifeng Jin

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Economics

Universidad Carlos III de Madrid

Advisor:

Carlos Velasco

March 2023

This thesis is distributed under license "Creative Commons Atributtion - Non
Commercial - Non Derivatives".
(cc) $(G)$

To Víctor Royás, an angel who graced my life with his love.

## ACKNOWLEDGEMENTS

The pursuit of a Ph.D. was not a part of my life plan until I decided to come to Spain. The decision to embark on this journey was a regime switch in my life adventure, rather than a persistent shock. Despite the inevitable frustration, anxiety, and unexpected obstacles that came my way, this experience has taught me the importance of facing difficulties head-on, which will undoubtedly benefit me for the rest of my life beyond the realm of research. I extend my heartfelt gratitude to Universidad Carlos III de Madrid for providing me with this invaluable opportunity. Additionally, I applaud myself for persevering through the challenges and successfully completing this endeavor.

First and foremost, I am grateful for the continuous guidance and support of my supervisor, Carlos. He has been like an "encyclopedia" of econometrics, always available to help whenever I encountered obstacles in my research. His encouragement during times of demotivation and his tolerance for my personal emotions at work have been invaluable. Special thanks go to Juan Carlos and Miguel Delgado, whose insightful advice has greatly improved the quality of my papers. Additionally, I am fortunate to be surrounded by a faculty of exceptional econometricians at UC3M, especially professors Juanjo, Jesus, and Nazarii. They have created a friendly and supportive environment that has greatly contributed to my research. I am also grateful to Johannes, Alan, and Ignacio for their earnest help during my job search. I would like to express my appreciation to the administrative staff, particularly Angélica, Monica, and others, for their tireless efforts in managing the bureaucratic aspects of my Ph.D. life. Their contributions are vital, and I could not have succeeded without their support.

I also want to extend my thanks to my officemates, Álvaro and Miguel, for being good company on this journey, to Morteza, for being an excellent beer buddy, to Rui, Yuhao, Minghai, Meng, Junyi, Boyao, Ale, Lidia, Telmo, Henry, Alvaro, Flor, Agus, Fer, Puma, Mariale, Deli, Najiba, Isa, Ona, ..., every fantastic person I have met in the doctoral program. I am very happy to have met Maumau, Xiaofei, Xiaoyi, and Tria outside my work for hanging out together. Thanks to my parents, who have been always so supportive without asking me for anything in return.

This work is dedicated to my partner, Víctor, who passed away from a brain tumor last year. Víctor, thank you for always believing in me and for being my constant source of encouragement. The time we spent together was the most memorable period of my life, and you taught me the most important lesson in life about love. Thanks to Pepi, Jesús, Lourdes, Marimar, and Carlos for being my Spanish
family and taking care of me with love. I love you all so much.
Although time moves forward in a linear fashion, each of us will generate a unique and complex nonlinear dynamic path throughout our lives.

## PUBLISHED AND SUBMITTED CONTENT

- The first chapter, titled "Quantile Autoregression-based Non-causality Testing", was presented in the seminar CEBA Talk and for the convenience of dissemination afterward, I made it available online. The possible online source is: Source.


## OTHER RESEARCH MERITS

This thesis has been possible thanks to the financial support from the grant BES-2017-082695 from the Ministerio de Economía Industria y Competitividad.


#### Abstract

Over the last two decades, there has been growing interest among economists in nonfundamental univariate processes, generally represented by noncausal and noninvertible time series. These processes have become increasingly popular due to their ability to capture nonlinear dynamics such as volatility clustering, asymmetric cycles, and local explosiveness - all of which are commonly observed in Macroeconomics and Finance. In particular, the incorporation of both past and future components into noncausal and noninvertible processes makes them attractive options for modeling forward-looking behavior in economic activities. However, the classical techniques used for analyzing time series models are largely limited to causal and invertible counterparts. This dissertation seeks to contribute to the field by providing theoretical tools robust to noncausal and noninvertible time series in testing and estimation.

In the first chapter, "Quantile Autoregression-Based Non-causality Testing", we investigate the statistical properties of empirical conditional quantiles of non-causal processes. Specifically, we show that the quantile autoregression (QAR) estimates for non-causal processes do not remain constant across different quantiles in contrast to their causal counterparts. Furthermore, we demonstrate that non-causal autoregressive processes admit nonlinear representations for conditional quantiles given past observations. Exploiting these properties, we propose three novel testing strategies of non-causality for non-Gaussian processes within the QAR framework. The tests are constructed either by verifying the constancy of the slope coefficients or by applying a misspecification test of the linear QAR model over different quantiles of the process. Some numerical experiments are included to examine the finite sample performance of the testing strategies, where we compare different specification tests for dynamic quantiles with the Kolmogorov-Smirnov constancy test. The new methodology is applied to some time series from financial markets to investigate the presence of speculative bubbles. The extension of the approach based on the specification tests to AR processes driven by innovations with heteroskedasticity is studied through simulations. The performance of QAR estimates of non-causal processes at extreme quantiles is also explored.

In the second chapter, "Estimation of Time Series Models Using the Empirical Distribution of Residuals", we introduce a novel estimation technique for general linear time series models, potentially noninvertible and noncausal, by utilizing the empirical cumulative distribution function of residuals. The proposed method relies on the generalized spectral cumulative function to characterize the pairwise dependence of residuals at all lags. Model identification can be achieved by exploiting


the information in the joint distribution of residuals under the iid assumption. This method yields consistent estimates of the model parameters without imposing stringent conditions on the higher-order moments or any distributional assumptions on the innovations beyond non-Gaussianity. We investigate the asymptotic distribution of the estimates by employing a smoothed cumulative distribution function to approximate the indicator function, considering the non-differentiability of the original loss function. Efficiency improvements can be achieved by properly choosing the scaling parameter for residuals. Finite sample properties are explored through Monte Carlo simulations. An empirical application to illustrate this methodology is provided by fitting the daily trading volume of Microsoft stock by autoregressive models with noncausal representation. The flexibility of the cumulative distribution function permits the proposed method to be extended to more general dependence structures where innovations are only conditional mean or quantile independent.

In the third chapter, "Directional Predictability Tests", joint with Carlos Velasco, we propose new tests of predictability for non-Gaussian sequences that may display general nonlinear dependence in higher-order properties. We test the null of martingale difference against parametric alternatives which can introduce linear or nonlinear dependence as generated by ARMA and all-pass restricted ARMA models, respectively. We also develop tests to check for linear predictability under the white noise null hypothesis parameterized by an all-pass model driven by martingale difference innovations and tests of non-linear predictability on ARMA residuals. Our Lagrange Multiplier tests are developed from a loss function based on pairwise dependence measures that identify the predictability of levels. We provide asymptotic and finite sample analysis of the properties of the new tests and investigate the predictability of different series of financial returns.

## CONTENTS

1. CHAPTER I: QUANTILE AUTOREGRESSION-BASED NON-CAUSALITY TESTING ..... 1
1.1. Introduction ..... 1
1.2. Mixed Causal and Non-causal Autoregressions ..... 3
1.3. QAR-based Non-causality Tests ..... 8
1.3.1. Benchmark model: $\operatorname{MAR}(r, s)$ ..... 8
1.4. Monte Carlo Simulations ..... 18
1.5. Empirical Applications ..... 24
1.6. Extensions ..... 26
1.6.1. Some Extensions ..... 26
1.6.2. Perspective from Extreme Quantiles ..... 30
1.7. Conclusion ..... 30
1.8. Appendices ..... 32
1.8.1. Some Properties of Non-causal Autoregressive Processes ..... 32
1.8.2. Asymptotic Properties of QAR Estimates ..... 35
1.8.3. Proof to Theorem 1.3.1 ..... 36
1.8.4. Simulations of Extensions ..... 37
2. CHAPTER II: ESTIMATION OF TIME SERIES MODELS USING THE EMPIRICAL DISTRIBUTION OF RESIDUALS ..... 39
2.1. Introduction ..... 39
2.2. Pairwise dependence measures based on residuals ..... 42
2.3. Model estimation under serial independence ..... 46
2.3.1. Model identification under serial independence ..... 46
2.3.2. Estimates of model parameters under serial independence ..... 48
2.3.3. Asymptomatic distribution of estimates under serial independence ..... 50
2.3.4. Standard error calculation ..... 54
2.4. Simulations ..... 56
2.4.1. Indicator approach: $h=0$. ..... 57
2.4.2. Smooth $c d f$ approach: $h>0$ ..... 58
2.5. Empirical application ..... 59
2.5.1. Clustering volatility ..... 59
2.6. Discussion: Measures of dependence under a martingale difference as- sumption ..... 60
2.7. Appendix A: Proofs ..... 65
2.7.1. Proof of Lemma 2.2.1 ..... 65
2.7.2. Proof of Lemma 2.3.1 ..... 67
2.7.3. Truncation effect ..... 70
2.7.4. Proof of Theorem 2.3.2 ..... 71
2.7.5. Proof of Theorem 2.3.3 ..... 75
2.7.6. Proof of Theorem 2.3.4 ..... 77
2.7.7. Proof of Theorem 2.3.5 ..... 82
2.8. Appendix B: Numerical Calculus ..... 89
2.8.1. Calculations of Sample Objective Function ..... 89
3. CHAPTER III: DIRECTIONAL PREDICTABILITY TESTS ..... 91
3.1. Introduction ..... 91
3.2. Predictability, linear and nonlinear dependence ..... 93
3.2.1. Predictability introduced by all-pass filters on mds sequences ..... 94
3.2.2. Higher order dependence introduced by all-pass filters ..... 97
3.3. Directional Predictability Hypothesis Tests ..... 98
3.3.1. Tests of the $m d s$ hypothesis ..... 101
3.3.2. Tests of the AP model ..... 104
3.4. Directional Nonlinear Dependence Tests for Residuals ..... 106
3.5. Monte Carlo Simulations ..... 108
3.6. Empirical Application ..... 110
3.7. Conclusions ..... 111
3.8. Appendix A: Proofs of Section 2 ..... 113
3.9. Appendix B ..... 121
3.9.1. General ARMA(p,p) model ..... 121
3.9.2. Technical Assumptions ..... 123
3.9.3. Proofs of Section 3 ..... 124
3.9.4. Tests based on iid dependence measures ..... 127
BIBLIOGRAPHY ..... 130

## LIST OF FIGURES

1.2.1 Trajectories of MAR $(1,1)$ processes with different parameters $(\phi, \psi)$, $\mathrm{T}=500$ ..... 6
1.3.1 $\mathrm{QAR}(1)$ estimates of a pair of $\mathrm{AR}(1)$ processes ..... 11
1.4.1 Four cases of $\operatorname{QAR}(1)$ on $\operatorname{MAR}(1,0)$ and $\operatorname{MAR}(0,1), \mathrm{T}=500, \phi(\psi)=$ 0.6 : asymmetric distributions ..... 21
1.4.2 Three cases of $\operatorname{QAR}(1)$ on $\operatorname{MAR}(1,0)$ and $\operatorname{MAR}(0,1), \mathrm{T}=500, \phi(\psi)=$ 0.6: symmetric distributions ..... 21
1.5.1 Financial series trajectories ..... 26
1.5.2 Sample partial autocorrelation functions of six financial series ..... 27
1.8.1 conditional density of $Y_{t}$ given different $x$ ..... 35
2.1.1 Simulated processes from causal and noncausal AR(1) models ..... 42
2.6.1 Microsoft daily trading volume from $6 / 3 / 1993$ to $5 / 26 / 1999$ ..... 63
2.6.2 Diagnostics of residuals from both causal and non-causal models: a comparison in ACF of residuals in squared values ..... 64

## LIST OF TABLES

1.4.1 Empirical size and power of non-causality test using constancy ${ }^{\star}$ test in QAR in various cases ..... 20
1.4.2 Empirical size and power of non-causality test using constancy test with different trimmed quantile interval ..... 22
1.4.3 Empirical size and power of non-causality test using EV test in QAR in various cases ..... 23
1.4.4 Empirical size and power of non-causality test using EG test in QAR in various cases ..... 24
1.4.5 Comparison of QAR-based non-causality tests ..... 25
1.5.1 Non-causality tests for six financial series in Fries and Zakoïan (2019) ..... 28
1.8.1 Empirical size and power of non-causality tests for AR-ARCH models with known heteroskedasticity ..... 38
2.6.1 Comparison of estimates using empirical $c d f$ and standard normal $c d f$ under different distributions of innovations: study in $\operatorname{AR}(1)$ case ..... 62
2.6.2 Estimates of $\operatorname{AR}(2)$ generated by innovations following $\chi_{5}^{2}$. ..... 62
2.6.3 Performance of the estimates by approximating the indicator func- tion using smoothed $c d f$ with different convergence rates of $h: \mathrm{AR}(1)$ driven by $\chi_{5}^{2}$ innovations ..... 63
3.5.1 Rejection rates of $5 \%$ LM tests: $\operatorname{ARMA}(1,1)$ models, $\varepsilon_{t} \sim \chi_{4}^{2}$. ..... 109
3.5.2 Rejection rates of $5 \%$ LM tests: $\operatorname{ARMA}(1,1)$ models, $\varepsilon_{t} \sim t_{4}$ ..... 110
3.6.1 iid estimation and tests for quarterly returns. ..... 112
3.6.2 mds estimation and tests for quarterly returns ..... 113

# 1. CHAPTER I: QUANTILE AUTOREGRESSION-BASED NON-CAUSALITY TESTING 

### 1.1. Introduction

A stationary autoregressive moving-average (ARMA) process is defined as causal concerning the specified innovation sequence if all roots of the autoregressive polynomials are outside of the unit circle, so it can be represented by an infinite sum of past innovations. However, non-causal autoregressive (AR) processes, due to their ability to display various non-linear dynamics, have been drawing increasing attention in the econometrics literature during the last two decades. The same concept has been fully exploited as non-minimum phase stochastic systems with applications to natural sciences. Unlike the causal AR process in the classical time series context, mixed causal and non-causal AR processes do not impose presumptions on the location of the lag polynomial roots except for the exclusion of the unit root, which allows these stationary processes to be dependent on past and future innovations at the same time. Breidt et al. (2001) showed that non-causal AR processes can capture stylized facts like clustering volatility in financial data, which usually is associated with GARCH models. The same argument is made in the paper by Lanne et al. (2013), where they derived a closed-form expression for the correlation of squared values in levels in the $\operatorname{ARMA}(1,1)$ case. Fries and Zakoïan (2019); Gouriéroux and Zakoïan (2017) and Cavaliere et al. (2020) proposed to model speculative bubbles with non-causal AR or mixed causal and non-causal AR processes generated by heavy-tailed innovations because they can display local explosive behavior. Regarding forecasting, Lanne et al. (2012) and Hecq and Voisin (2021) argued that there is a gain in forecasting performance after introducing non-causality into the modeling procedure. Moreover, Lanne and Luoto (2013) pointed out that non-causal AR processes can be an alternative to non-invertible processes for forward-looking behavior, see Alessi et al. (2011) for a comprehensive survey of empirical applications of non-invertible (non-fundamental) processes to Macroeconomics and Finance.

The emerging applications of non-causal AR processes promote an interest in testing non-causality in practice, given that autocorrelation functions fail to discriminate non-causal processes from their causal counterparts. Needless to say, the non-causality check can be naturally achieved through testing classical linear hypotheses under robust estimation techniques applicable to possibly non-causal processes. These estimation methods have been developed by Breid et al. (1991) and Lii
and Rosenblatt $(1992,1996)$ through non-Gaussian maximum likelihood schemes or minimum distance estimation exploiting information contained in higher order moments/cumulants or characteristic/cumulative distribution functions of residuals, see Velasco and Lobato (2018), Velasco (2022), and Jin (2022). With this approach, the test procedure is confined to the assumptions required for the corresponding estimates, which can be somehow stringent. Apart from that, a factorization of the coefficients is necessary for disentangling the roots, which becomes rather complicated as the order of the AR process increases. Therefore, a testing strategy before the estimation, which can potentially work as a model selection, needs investigation. Besides, the testing can serve as a detection tool for the existence of speculative bubbles in empirical time series for the ability of non-causal processes with heavy tails to exhibit local explosive behavior.

However, except for the robust estimation techniques introduced before, little has been done on the theory of testing non-causality in AR processes. Nevertheless, all the estimation techniques above deliver the same message that additional information beyond second-order moments is required to identify non-causality. Following this message, we propose some testing strategies for the non-causality of time series within the quantile autoregressions framework (QAR hereafter) (Koenker and Xiao (2006)), which allows us to make use of the complete distribution to measure dependence. Similarly, Hecq and Sun (2021) applied QAR to the target process and entertained the sum of rescaled absolute residuals as an information criterion to select between purely causal and non-causal models. The approach of Hecq and Sun (2021) obtains the proposed test statistic by running QAR in direct and reserved time, respectively. This approach can bring up ambiguity in the model selection when the complexity of the causality structure escalates as the order of the AR model increases. Thereby, this strategy may yield misleading results when the time series is mixed causal and non-causal.

In this paper, we propose three testing procedures based on the well-developed inference for QAR estimates exploiting the non-linear characteristics of non-causal processes. Recall that the coefficients in the conditional quantile regression for the location model ${ }^{1}$ are quantile-invariant except for the intercept. When the process is causal, the independence of the current innovation with past observations contributes to the invariant property of coefficients in the QAR model across different quantiles. However, under non-causality the true conditional quantile of a nonGaussian AR process exhibits non-linearity in the past information. Induced from this, the best linear approximation to the true conditional quantile is expected to show varying coefficients across distinct quantiles. Using this feature, we introduce

[^0]our first strategy for the objective of testing non-causality by carrying out the constancy test over the entire quantile interval. Its easy-to-implement attribute makes it a perfect candidate for a preliminary check of non-causality. The other approach to detect non-causality is achieved through a specification test of the linear functional form for the conditional quantile. With this specification-based approach, no distributional knowledge of the innovations beforehand is required, nor does the correct specification of the conditional quantile need to be spelled out.

The testing strategies introduced in this paper fill a gap in the theory of testing non-causality. Apart from that, they share an appealing property of retaining power when the AR process is higher-order with a mixture of causal and non-causal representation. Like the significant advantage of quantile regression over conditional mean regression, our approach is robust to outliers, making it suitable for heavytailed processes commonly employed in finance. Our specification-based approach can also be tentatively extended to an AR process with conditional heteroskedasticity. The tests achieve considerable power in relatively small samples. In addition, some Monte Carlo simulation results suggest that the estimates in linear quantile models approach the true parameters for non-causal processes as the quantile estimand approaches to either extremum ( 0 or 1 ), which can be a potential viewpoint to investigate the nonlinear features of non-causal processes in future research.

The rest of the paper is organized as follows. Section 1.2 introduces mixed causal and non-causal Autoregressions and some of their statistical properties. Section 1.3 investigates the non-causality testing within the QAR framework. Section 1.4 discusses the finite sample performance of our proposed testing procedures through Monte Carlo Simulations. Section 1.5 illustrates the tests using financial data with the possible existence of speculative bubbles. Section 1.6 presents some extensions of the proposed tests. Section 1.7 closes the paper with conclusions

### 1.2. Mixed Causal and Non-causal Autoregressions

In the context of classical time series analysis, it is customary to restrict attention to the causal representation of autoregressive processes while modeling stationary univariate time series. The reason is mentioned in (Brockwell and Davis, 2009, p. 105). Every non-causal autoregressive process is a stationary solution to a future-independent autoregressive process with explosive roots. It provides the same second-order structure as its causal counterpart. However, to feature higher-order dynamics, a general framework of autoregressive processes named mixed causal and non-causal autoregressive processes $(\operatorname{MAR}(r, s) \text { hereafter })^{2}$ is proposed where tem-

[^1]poral dependence in both past and future is introduced in the processes, defined by
\[

$$
\begin{equation*}
\phi(L) \psi\left(L^{-1}\right) Y_{t}=u_{t} \tag{1.2.1}
\end{equation*}
$$

\]

where $\phi(L)=1-\phi_{1} L-\phi_{2} L^{2}-\cdots-\phi_{r} L^{r}$ and $\psi\left(L^{-1}\right)=1-\psi_{1} L^{-1}-\psi_{2} L^{-2}-\cdots-\psi_{s} L^{-s}$ are polynomials with backward and forward operators, respectively. $\left\{u_{t}\right\}_{t \in \mathbb{Z}}$ is a sequence of independent identically distributed (iid) innovations with zero mean. $p=r+s$ is the total order encompassing both causal and non-causal polynomials with $\phi_{r} \neq 0$ and $\psi_{s} \neq 0$. An equivalent expression of equation (1.2.1) in moving average can be given by

$$
\begin{equation*}
Y_{t}=\phi(L)^{-1} \psi\left(L^{-1}\right)^{-1} u_{t}=\sum_{j=-\infty}^{\infty} \rho_{j} u_{t-j} \tag{1.2.2}
\end{equation*}
$$

where $Y_{t}$ may depend on both future and past innovations. The stationarity of $Y_{t}$ is assured by the absolute summability of $\rho_{j}, \sum_{j=-\infty}^{\infty}\left|\rho_{j}\right|<\infty$ and $\mathbb{E}\left|u_{t}\right|^{1+\delta}<\infty$ for $\delta>0$. The former condition is guaranteed once the roots of both polynomials $\phi(z)$ and $\psi(z)$ are confined to locate outside the unit circle.

When $\psi(z)=1$ for all $z$, the process is reduced to a purely causal process $\operatorname{MAR}(r, 0)$ like in conventional studies. While if $\phi(z)=1$ for all $z$, the process becomes purely non-causal. Below, we attach several examples to illustrate the statistical properties of a general autoregressive process $\operatorname{MAR}(r, s)$.

Example 1.2.1 (Second-order Property). Define a purely non-causal MAR $(0,1)$ sequence $\left(1-\psi L^{-1}\right) Y_{t}=u_{t}$ starting from an iid sequence of innovations $u_{t}$ of zeromean and finite variance $\sigma^{2}$. The autocovariance function of $Y_{t}$ is provided by

$$
\begin{aligned}
\operatorname{Cov}\left(Y_{t+h}, Y_{t}\right) & =\operatorname{Cov}\left(\sum_{j=0}^{\infty} \psi^{j} u_{t+h+j}, \sum_{j=0}^{\infty} \psi^{j} u_{t+j}\right) \\
& =\sum_{j=0}^{\infty} \operatorname{Cov}\left(\psi^{j} u_{t+h+j}, \psi^{j+h} u_{t+h+j}\right) \\
& =\frac{\psi^{h}}{1-\psi^{2}} \sigma^{2} \text { for } h=0,1,2, \ldots
\end{aligned}
$$

It is worth noting, after some simple calculation, that $\left\{Y_{t}\right\}_{t \in \mathbb{Z}}$ shares the same autocovariance function as the $\operatorname{MAR}(1,0)$ sequence $\left\{\tilde{Y}_{t}\right\}_{t \in \mathbb{Z}}$ defined by $(1-\psi L) \tilde{Y}_{t}=u_{t}$, which is its causal counterpart. A general conclusion can be drawn for $\operatorname{MAR}(r, s)$ through the autocovariance generating function (ACGF). The ACGF of a stationary

[^2]autoregression is given by
\[

$$
\begin{aligned}
G(L) & =\frac{1}{\left|\phi(L) \psi\left(L^{-1}\right)\right|^{2}} \sigma^{2}=\frac{1}{\phi(L) \phi\left(L^{-1}\right) \psi\left(L^{-1}\right) \psi(L)} \sigma^{2} \\
& =\frac{1}{\phi(L) \psi(L) \phi\left(L^{-1}\right) \psi\left(L^{-1}\right)} \sigma^{2} \\
& =\frac{1}{|\phi(L) \psi(L)|^{2}} \sigma^{2}
\end{aligned}
$$
\]

from where it is clear that $\operatorname{MAR}(r, s)$ takes the identical form of $A C G F$ with $M A R\left(r^{\prime}, s^{\prime}\right)$ as long as the total order $r+s=r^{\prime}+s^{\prime}$ is satisfied and the roots to all polynomials match.

This second-order property explains the failure to distinguish non-causal processes from causal processes based on the ACF. Moreover, it implies that nonGaussianity of innovations is required for identifying non-causal processes since second-order properties are sufficient to characterize a Gaussian probabilistic structure but not to other distributions.

Example 1.2.2 (Local Explosiveness). Define a $\operatorname{MAR}(1,1)$ process by

$$
(1-\phi L)\left(1-\psi L^{-1}\right) Y_{t}=u_{t}, \text { where } u_{t} \sim \operatorname{Lognormal}(0,2)-\exp (2)
$$

A MAR $(r, s)$ process with heavy-tailed innovations can generate multiple phases of local explosiveness, which can be employed to model speculative bubbles. A detailed investigation of probabilistic properties of a MAR process driven by $\alpha$-stable non-Gaussian innovations is provided by Fries and Zakoïan (2019), where they show the properties of the marginal and conditional distributions of a stable $\operatorname{MAR}(r, s)$. But in a general situation, there is no closed-form solution to either the marginal or the conditional distribution of a non-causal AR process. Here we illustrate the potential applications of $\operatorname{MAR}(r, s)$ processes in modeling speculative bubbles with some simulated trajectories. In this example, we plot four distinct scenarios by varying the parameters of both causal and non-causal components of the $\operatorname{MAR}(1,1)$ process with log-normal distributed innovations ${ }^{3}$. Generally, with non-causality, the processes can mimic bubbles by repetitive phases of upward trends followed by a sharp drop; see the lower panel in Figure 1.2.1. The upper panel in the same figure indicates more complicated dynamics can be generated by incorporating more

[^3]causal/non-causal components in the data-generating process regarding the number and magnitude of bubbles.

Figure 1.2.1: Trajectories of $\operatorname{MAR}(1,1)$ processes with different parameters $(\phi, \psi)$, $\mathrm{T}=500$


Note: the upper panel exhibits two $\operatorname{MAR}(1,1)$ processes with different parameters on the causal/non-causal components; the lower panel exhibits a purely non-causal AR(1) process (left) and a purely causal $A R(1)$ process (right) when one of the parameters degenerates to zero.

Example 1.2.3 (Conditional heteroskedasticity). Here we exemplify the applicability of $\operatorname{MAR}(r, s)$ in characterizing clustering volatility as an alternative to ARCH or other stochastic volatility models with a simple $\operatorname{MAR}(1,1)$ process. This argument is originally made by Breidt et al. (2001) in an empirical application to New Zealand/US exchange rate. Lanne et al. (2013) elaborate on it by deriving explicitly the expression of the autocorrelation of $Y_{t}^{2}$ in an all-pass model of order 1. However, no formal justification for the higher-order dependence structure has been made on general non-causal processes. Nevertheless, we present the following example that a non-causal process can exhibit higher-order dependence.

Continued with data generating process $\operatorname{MAR}(1,1)$, it can be reexpressed by an
$A R$ (2) with the $\frac{1-\psi L}{1-\psi L^{-1}}$ filter on the innovations,

$$
\begin{aligned}
& \quad(1-\phi L)\left(1-\psi L^{-1}\right) Y_{t}=u_{t} \quad u_{t} \sim \operatorname{IID}\left(0, \sigma^{2}\right) \\
& \Longleftrightarrow(1-\phi L)(1-\psi L) Y_{t}=\tilde{u}_{t} \\
& \text { with } \tilde{u}_{t}=\frac{1-\psi L}{1-\psi L^{-1}} u_{t}=\sum_{j=-\infty}^{\infty} \rho_{j} u_{t+j} \\
& \text { with } \rho_{j}= \begin{cases}0 & \text { if } j<-1 \\
-\psi & \text { if } j=-1 \\
\psi^{j}-\psi^{j+2} & \text { if } j \geq 0 .\end{cases}
\end{aligned}
$$

The $\frac{1-\psi L}{1-\psi L^{-1}}$ filter introduces higher-order dependence to an iid innovation sequence ${ }^{4}$ preserving its uncorrelatedness. By simple algebra, we can obtain the formula of the autocorrelation function of $\tilde{u}_{t}^{2}$,

$$
\operatorname{Corr}\left(\tilde{u}_{t}^{2}, \tilde{u}_{t+h}^{2}\right)=\frac{\kappa_{4}\left(u_{t}\right)\left(\sum_{j=-\infty}^{\infty} \rho_{j}^{2} \rho_{j+h}^{2}\right)+2 \sigma^{4}\left(\sum_{j=-\infty}^{\infty} \rho_{j} \rho_{j+h}\right)^{2}}{\kappa_{4}\left(u_{t}\right)\left(\sum_{j=-\infty}^{\infty} \rho_{j}^{4}\right)+2 \sigma^{4}\left(\sum_{j=-\infty}^{\infty} \rho_{j}^{2}\right)^{2}},
$$

which, in general, does not vanish ${ }^{5}$. This property demonstrates the capability of non-causal processes to exhibit volatility clustering, which is commonly observed in financial data. The higher-order dependence analysis of $\tilde{u}_{t}$ is relegated to Appendix 1.8.1, indicating the possibility of accommodating higher-order nonlinear dynamics with a general $\operatorname{MAR}(r, s)$ model.

As shown in the preceding examples, $\operatorname{MAR}(r, s)$ models can display various nonlinear characteristics with a linear process generation scheme. This deviation from "linearity" can be employed as a crucial feature to detect non-causality in linear time series. A fundamental result is formalized by Rosenblatt (2000) on the best one-step predictor in the mean square for a general AR process, where he demonstrates that the conditional expectation must be nonlinear in the past if there is non-causality involved in the non-Gaussian AR processes with finite variance. This statement provides us with the critical theoretical result where our tests for non-causality are grounded, which will be elaborated on in the next section. The extension to a general VARMA process framework has been developed by Chen et al. (2017) and applied to a test for non-invertibility. Afterward, Fries and Zakoïan (2019) consider the case when the $\operatorname{MAR}(1, s)$ process is driven by symmetric $\alpha$-stable innovations with infinite variance. They surprisingly find that the conditional expectation can be explicitly expressed by a linear function of the past information, in contrast to

[^4]a $\operatorname{MAR}(r, s)$ model with finite variance. However, no study has yet been done on the distributional characterization of a $\operatorname{MAR}(r, s)$ process due to no closed-form solution. Following a simulation-based approach, we perform a preliminary analysis of the conditional density function of a process given the past observations, see Appendix 1.8.1. In short, in the presence of non-causality, the shape of the conditional distribution of the process, say $f\left(Y_{t} \mid Y_{t-1}=y\right)$, is $y$-dependent. Still, the dependence pattern is hard to characterize since it varies across different distributions.

### 1.3. QAR-based Non-causality Tests

### 1.3.1. Benchmark model: $\operatorname{MAR}(r, s)$

In this section, we formalize the test procedures for non-causality. The null hypothesis of interest is $Y_{t}$ being a causal process, i.e.

$$
\begin{equation*}
\mathbb{H}_{0}: \psi_{1}=\cdots=\psi_{s}=0 \text { in (1.2.1) } \tag{1.3.1}
\end{equation*}
$$

against the alternative hypothesis denoted by $\mathbb{H}_{A}$, which is $\left\{Y_{t}\right\}$ is non-causal,

$$
\begin{equation*}
\mathbb{H}_{A}: \psi_{k} \neq 0 \text { for some } k \in\{1,2, \ldots, s\} \text { in (1.2.1). } \tag{1.3.2}
\end{equation*}
$$

A primitive strategy is to take it as a joint significance test for the coefficients of leads $\left\{Y_{t+j}\right\}_{j=1,2, \ldots, s}$ in (1.2.1). This approach would call for the identification of models and robust inference for the estimates under both hypotheses. In this paper, we propose a test for $\mathbb{H}_{0}$ employing the linearity property of $\left\{Y_{t}\right\}$ based on the QAR , which is easy to implement empirically and does not require consistent estimates for general autoregressive progresses.

Denoting the information set generated by the observations up to period $t$ by $I_{t}=$ $\sigma\left(Y_{t}, Y_{t-1}, \ldots\right)$ and the $\tau$-th quantile of $Y_{t}$ conditional on the past by $Q_{Y_{t}}\left(\tau \mid I_{t-1}\right)$, the following result on the linearity of $\left\{Y_{t}\right\}$ justifies our approach.

Assumption 1.1. Let $\left\{u_{t}\right\}_{t \in \mathcal{Z}}$ be a non-Gaussian iid sequence with $(k+1)$ th order moment finite and $(k+1)$ th cumulant nonzero for some $k \geq 2$.

Theorem 1.3.1. Under Assumption 1.1, a stationary $\operatorname{MAR}(r, s)$ process $\left\{Y_{t}\right\}$ has a non-degenerated non-causal component, i.e., $s \neq 0$ if and only if $Q_{Y_{t}}\left(\tau \mid I_{t-1}\right)$ is nonlinear in $\left\{Y_{t-j}\right\}_{j \geq 1}$ for at least one $\tau \in(0,1)$.

As discussed in Example 1.2.1, there is no meaning to discuss non-causality in a Gaussian structure as there is always an equivalent causal representation of a Gaussian AR process for its non-causal counterpart. The finiteness of the moment condition and nonzero cumulants are also required in Rosenblatt (2000). Theorem
1.3.1 is deduced from the nonlinearity of the best predictor of $Y_{t}$ in the mean square criterion for non-causal processes. The nonlinearity in the conditional mean implies dependence in the conditional distribution beyond the linear correlation. Note that the theorem does not point out the quantile(s) where this nonlinear relationship occurs, nor the manners in which this nonlinearity is expressed. Another remark is that the information set considered in the theorem can be replaced by $\sigma$-field generated by $Y_{t-1}$ up to $Y_{t-p}$ due to the Markovian property of $\operatorname{MAR}(r, s)$, which avoids the infinite-dimensional issue arising from $I_{t-1}$. The total number of the lags and leads of $Y_{t}$ included to explain the conditional quantile of $\operatorname{MAR}(r, s)$ processes, $p$, is determined by the partial autocorrelation function (PACF) of $Y_{t}$.

This theorem prompts us to adopt QAR as a medium to detect non-causality. QAR is a comprehensive analysis tool in the time series context, providing robust statistical analyses against outliers in the measurement of the response variable, which has proven to be rather prevailing in recent decades. Given a $\operatorname{MAR}(r, s)$ process of order $p$ defined by (1.2.1), if the non-causal component degenerates to 1 , i.e., $s=0, r=p$, the conditional quantile of $Y_{t}$ can be expressed by

$$
\begin{align*}
Q_{Y_{t}}\left(\tau \mid I_{t-1}\right) & =Q_{u_{t}}(\tau)+\phi_{1} Y_{t-1}+\phi_{2} Y_{t-2}+\cdots+\phi_{p} Y_{t-p} \\
& =\theta_{0}(\tau)+\theta_{1}(\tau) Y_{t-1}+\theta_{2}(\tau) Y_{t-2}+\cdots+\theta_{p}(\tau) Y_{t-p}  \tag{1.3.3}\\
& =\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}(\tau) \quad \forall \tau \in(0,1),
\end{align*}
$$

where $Q_{u_{t}}(\tau)$ denotes the $\tau$-quantile of innovations $u_{t}$, and $\phi_{j}$ 's are the coefficients of corresponding $Y_{t-j}$ in the polynomial expansion of (1.2.1). Therefore, after imposing pure causality, the conditional quantile of $Y_{t}$ can be fully characterized by a linear function of past observations, $\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}(\tau)$ with $\boldsymbol{X}_{t}=\left(1, Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}\right)^{\prime}$ and $\boldsymbol{\theta}(\tau)=$ $\left(\theta_{0}(\tau), \ldots, \theta_{p}(\tau)\right)^{\prime}=\left(Q_{u_{t}}(\tau), \phi_{1}, \ldots, \phi_{p}\right)^{\prime}$. The QAR estimates of coefficients $\boldsymbol{\theta}(\tau)$ in this linear quantile model can be obtained by minimizing the following problem for each $\tau$,

$$
\begin{equation*}
\hat{\boldsymbol{\theta}}(\tau)=\underset{\boldsymbol{\theta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{t=1}^{T} \rho_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}\right), \tag{1.3.4}
\end{equation*}
$$

with the check function $\rho_{\tau}(u)=u(\tau-\mathbb{I}(u<0))$. The asymptotic properties of linear QAR estimates were first established by Koenker and Xiao (2006). A brief review of QAR estimates (1.3.4) can be found in Appendix 1.8.2. However, if $Y_{t}$ has a non-degenerated non-causal component, the linear dynamic model (1.3.3) for its conditional quantile is misspecified for at least one $\tau \in(0,1)$, following Theorem 1.3.1.

Within the linear QAR framework, Hecq and Sun (2021) consider a statistic aggregating the information of the residuals over quantiles, which they employ as a model selection criterion between purely causal AR models and purely non-causal AR models. Given that the calculation of residuals is done either by running QAR
with direct or reversed time, this methodology may provide misleading conclusions regarding $\operatorname{MAR}(r, s)$ in presence of causality and non-causality at the same time. By contrast, our approaches address more the correctness of the linear specification of the conditional quantile through the QAR. For non-causal autoregressive processes, no closed form of the nonlinear conditional quantile of $Y_{t}$ is required. Consequently, our approach is robust to the general $\operatorname{MAR}(r, s)$ setting. Before we carry out the tests for non-causality, the following assumptions are imposed in the QAR framework.

Assumption 1.2. The distribution function of innovations $u_{t}, F(u)$, admits a continuous density function $f(u)$ away from zero on the domain $\mathcal{U}=\{u: 0<F(u)<$ $1\}$.

Assumption 1.3. Denote the family of conditional distribution $\left\{P\left(Y_{t}<y \mid \boldsymbol{X}_{t}=x\right), y \in\right.$ $\left.\mathbb{R}, x \in \mathbb{R}^{r+s}\right\}$ as $F_{x}(y)$ and its Lebesgue density as $f_{x}(y)$, that is uniformly bounded on the space of $y \times x \subseteq \mathbb{R} \times \mathbb{R}^{r+s}$ and uniformly continuous.

Corollary 1.3.1.1. Under Assumptions 1.1-1.3 and null hypothesis $\mathbb{H}_{0}$, the coefficients except for the intercept of the conditional quantile are constant across different quantiles in $(0,1)$.

Corollary 1.3.1.1 is a consequence of the independence between $u_{t}$ and $Y_{t-j}$ for $j=1,2, \ldots, p$ when the $\operatorname{MAR}(r, s)$ is purely causal. As shown in the equation (1.3.3), the slope coefficients $\left\{\theta_{j}(\tau)\right\}_{j=1, \ldots, p}$ are constant over $\tau \in(0,1)$ and uniquely determined by the expansion of autoregressive polynomials of $Y_{t}$. Instead, the performance of the QAR estimates is more complicated in the mixed causal and non-causal autoregressive processes since the linear model is under misspecification. Angrist et al. (2006) demonstrated in their paper that quantile regression is essentially an approximation to the true conditional quantile function in a weighted mean squared criterion, with weights associated with the densities of $Y_{t}$ ranging from the linear approximation (1.3.3) to the true conditional quantile $Q_{Y_{t}}\left(\tau \mid Y_{t-1}, \ldots, Y_{t-p}\right)$. Their statement presents a rough idea of how well the linear function fits the true conditional quantile.

Constancy Test Our first approach for detecting non-causality comes along with the constancy test of QAR coefficients over the entire quantile range. As stated in Corollary 1.3.1.1, if the process is causal, the constancy should hold for all $\theta_{j}(\tau)$ for all $j=1,2, \ldots, p$ and $\tau \in(0,1)$. Whereas for a non-causal process, the estimated coefficients of $Y_{t-j}$ in the quantile regression may vary across different quantiles much likely for the following intuitions: i) non-causal processes display highly nonlinear dynamics, one of which is asymmetric dynamics; ii) linear quantile model is misspecified; iii) the conditional distribution of $Y_{t}$ varies both in the location and shape
at different values of past observations. For instance, Figure 1.3.1 depicts the dynamic performance of $\mathrm{QAR}(1)$ estimates of the slope parameter across the quantiles for a non-Gaussian autoregressive process, including causal and non-causal cases in different colors. The QAR(1) estimates in the non-causal processes (solid blue lines) exhibit a trendy pattern over the quantile domain. However, since a general solution for the true conditional quantile function of $Y_{t}$ is infeasible, we do not attempt to provide a detailed characterization of the varying coefficient property in the noncausal situation. The test for non-causality based on the constancy test can only be used to check the necessary condition of AR processes being causal. Nevertheless, the accessibility and straightforwardness of this method make it a touchstone for non-causality testing in practice.

Figure 1.3.1: $\operatorname{QAR}(1)$ estimates of a pair of $\operatorname{AR}(1)$ processes


Note: The figure displays the $\mathrm{QAR}(1)$ estimates of a pair of $\operatorname{AR}(1)$ processes. One is causal, and the other is non-causal. The left part of the figure is applied to the processes generated by exponential innovations, and the right part is to the processes generated by chi-square innovations. The true parameter is $0.6(1 / 0.6)$ for the causal(non-causal) case.

To implement the constancy test of coefficients under the QAR framework, we consider the approach developed by Koenker and Xiao (2006), where the hypothesis is formulated in the manner analog to the classical linear hypothesis:

$$
\mathbb{H}_{0}^{1}: \boldsymbol{R} \boldsymbol{\theta}(\tau)=\boldsymbol{\phi} \text { with } \boldsymbol{R}=\left(\mathbf{0}_{p \times 1} \vdots \boldsymbol{I}_{p}\right) \text { for all } \tau \in(0,1)
$$

with the unknown $\tau$-invariant parameter vector $\boldsymbol{\phi}=\left(\phi_{1}, \phi_{2}, \ldots, \phi_{p}\right)^{\prime}$ which needs to be estimated ${ }^{6}$. Naturally, the naive test for this hypothesis is constructed on the quantile process

$$
V_{T}(\tau)=\sqrt{T}\left[\boldsymbol{R} \hat{\Sigma}_{1}^{-1} \hat{\Sigma}_{0} \hat{\Sigma}_{1}^{-1} \boldsymbol{R}^{\prime}\right]^{-1 / 2}(\boldsymbol{R} \hat{\boldsymbol{\theta}}(\tau)-\hat{\boldsymbol{\phi}})
$$

[^5]and the Kolmogorov-Smirnov (KS) type of test statistic is adopted for the interest of testing a compact set of quantiles
$$
K S V_{T}=\sup _{\tau \in \Upsilon \subset(0,1)} V_{T}(\tau), \text { where } \Upsilon \text { is a compact interval, }
$$
where $\hat{\boldsymbol{\theta}}(\tau)$ is the linear QAR estimate, and $\hat{\Sigma}_{1}, \hat{\Sigma}_{0}$ are the corresponding estimates of the asymptotic variance, see Appendix 1.8.4. $\hat{\boldsymbol{\phi}}$ is a $\sqrt{T}$ consistent estimator of $\phi$ and a simple choice is the QAR estimator $\hat{\boldsymbol{\theta}}\left(\tau^{*}\right)$ at any $\tau^{*} \in \Upsilon .{ }^{7}$ A closed interval $\left[\epsilon_{1}, 1-\epsilon_{2}\right]$ with trivial numbers $\epsilon_{1}, \epsilon_{2}$ is proposed for $\Upsilon$ to avoid missing much information from the entire quantile interval $(0,1)$. Under the hypothesis of constancy $\mathbb{H}_{0}^{1}$,
$$
V_{T}(\tau) \Longrightarrow \boldsymbol{B}_{p}(\tau)-\underbrace{f\left(F^{-1}(\tau)\right)\left[\boldsymbol{R} \Sigma_{0}^{-1} \boldsymbol{R}^{\prime}\right]^{-1 / 2} \boldsymbol{Z}}_{\text {Drift }}
$$
where $\boldsymbol{B}_{p}(\tau)$ is a $p$-dimensional standard Brownian bridge, and $\boldsymbol{Z}$, the limiting behavior of $\hat{\boldsymbol{\phi}}$, is a Gaussian component brought up by the estimation of the nuisance parameter $\phi$. As stated, the drift compels the final distribution of the original statistic $V_{T}(\tau)$ to be data-dependent. To annihilate this non-trivial effect, a martingale transformation $\mathcal{K}$ was introduced into $V_{T}(\tau)$ to retrieve the distribution-free merit of the KS test. Denote the derivative of the density function by $\dot{f}$ and define
\[

$$
\begin{aligned}
& g(x)=\left(1,\left(\dot{f}\left(F^{-1}(x)\right) / f\left(F^{-1}(x)\right)\right)\right)^{\prime} \text { and } \\
& C(z)=\int_{z}^{1} g(x) g(x)^{\prime} d x
\end{aligned}
$$
\]

and the martingale transformation $\mathcal{K}$ on the process $V_{T}(\tau)$ is constructed as follows

$$
\begin{aligned}
\tilde{V}_{T}(\tau) & =\mathcal{K} V_{T}(\tau) \\
& =V_{T}(\tau)-\int_{0}^{\tau}\left[g_{T}^{\prime}(x) C_{T}^{-1}(x) \int_{x}^{1} g(s) d V_{T}(s)\right] d x
\end{aligned}
$$

where $g_{T}(x)$ and $C_{T}(x)$ are uniformly consistent estimators of $g(x)$ and $C(x)$ in the considered domain, respectively. The proposed KS-type norm on the transformed process becomes

$$
K S \tilde{V}_{T}=\sup _{\tau \in \Upsilon}\left\|\tilde{V}_{T}(\tau)\right\|
$$

Corollary 1.3.1.2 (Constancy test for non-causality). Under Assumptions 1.1-1.3

[^6]and the causality hypothesis $\mathbb{H}_{0}$,
\[

$$
\begin{aligned}
& \tilde{V}_{T}(\tau) \Rightarrow \boldsymbol{W}_{p}(\tau) \\
& K S \tilde{V}_{T} \Rightarrow \sup _{\tau \in \Upsilon}\left\|\boldsymbol{W}_{p}(\tau)\right\|
\end{aligned}
$$
\]

where $\boldsymbol{W}_{p}(\tau)$ represents a p-dimensional standard Brownian motion.

The related discussion on the estimation of density and score functions is given in Koenker and Xiao (2002), providing suggestions on the choice of bandwidth in detail. In the command KhmaladzeTest implemented in R studio, Hall/Sheather bandwidth (Hall and Sheather (1988)) for sparsity estimation is set as default. The critical values are obtained through approximating $\boldsymbol{W}_{p}(\tau)$ by a Gaussian random walk, and the corresponding values at different significance levels can be found in tables in the Appendix of Koenker and Xiao (2002). One remark on the quantile interval in the corollary, typically a symmetric interval $[\epsilon, 1-\epsilon]$ trimmed by a small number $\epsilon$ close to 0 is considered for simplicity in practice. Monte Carlo experiments have evidenced that an appropriate trimming in the entire quantile interval alleviates the over-rejection of the null hypothesis ascribable to the instability of estimation at extremal quantiles without sacrificing the power.

Specification test-based approach Another direction to test non-causality is based on Theorem 1.3.1, where non-causality in the linear processes is translated into the misspecification of conditional quantiles of non-causal processes by linear dynamic quantile models (1.3.3). Equivalently, we aim to test

$$
\begin{equation*}
\mathbb{E}\left(\Psi_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}_{0}\right) \mid Y_{t-1}, \ldots, Y_{t-p}\right)=0 \text { a.s. for some } \boldsymbol{\theta}_{0} \in \mathcal{B} \text { and } \forall \tau \in \Upsilon \subset(0,1), \tag{1.3.5}
\end{equation*}
$$

where $\Psi_{\tau}(\cdot)=\mathbb{I}(\cdot \leq 0)-\tau$, and $\mathcal{B}$ is a family of uniformly bounded functions from $\Upsilon \subset(0,1)$ to $\Theta \subset \mathbb{R}^{p+1} .{ }^{8}$ Both $\Upsilon$ and $\Theta$ are compact sets. Escanciano and Velasco (2010) (hereafter EV) characterize this restriction by unconditional moments
$\mathbb{E}\left(\Psi_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}_{0}\right) \exp \left(i x^{\prime} \boldsymbol{X}_{t}\right)\right)=0, \quad \forall x \in \mathbb{R}^{p+1}$, for some $\boldsymbol{\theta}_{0} \in \mathcal{B}$ and $\forall \tau \in \Upsilon \subset(0,1)$,
where $i=\sqrt{-1}$. Following the strategy of the EV test, we consider the statistic based on the residual processes indexed by quantiles $\tau, \boldsymbol{\theta} \in \mathcal{B}$ and $x \in \mathbb{R}^{p+1}$

$$
\begin{equation*}
R_{T}^{E V}(x, \tau ; \hat{\boldsymbol{\theta}})=T^{-1 / 2} \sum_{t=1}^{T} \Psi_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \hat{\boldsymbol{\theta}}\right) \exp \left(i x^{\prime} \boldsymbol{X}_{t}\right) \tag{1.3.7}
\end{equation*}
$$

[^7]with true parameters $\boldsymbol{\theta}_{0}$ replaced by QAR estimates $\hat{\boldsymbol{\theta}}$ from a given sample $\left(Y_{t}, \boldsymbol{X}_{t}^{\prime}\right)_{t=1,2, \ldots, T}$. $\boldsymbol{X}_{t}$ is composed by a constant and the lags of $Y_{t}$ up to order $p . T^{-1 / 2} R_{T}^{E V}$ approaches zero when $T$ goes to infinity for any $x \in \mathbb{R}^{p+1}$ if $\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}_{0}(\tau)$ is the correct specification for $Q_{Y_{t}}\left(\tau \mid I_{t}\right)$ and $\hat{\boldsymbol{\theta}}(\tau)$ is a $\sqrt{T}$-consistent estimator of $\boldsymbol{\theta}_{0}(\tau)$ for $\tau \in \Upsilon$. Thus, the distance between the process $R_{T}^{E V}$ defined in (1.3.7) and zero naturally turns into a measure of the deviation of $\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}_{0}(\tau)$ from the true $Q_{Y_{t}}\left(\tau \mid I_{t}\right)$. The suggested Cramér-von Mises (CvM) norm on (1.3.7) is defined by
\[

$$
\begin{equation*}
\int_{\tau \in \Upsilon, x \in \mathbb{R}^{p+1}}\left|R_{T}^{E V}(x, \tau ; \hat{\boldsymbol{\theta}})\right|^{2} d \Phi(x) d W(\tau) \tag{1.3.8}
\end{equation*}
$$

\]

where $\Phi(x)$ and $W(\tau)$ are weighting functions defined on $\mathbb{R}^{p+1}$ and $\Upsilon$, respectively with positive derivative in the corresponding domain. This CvM norm permits us to consider model specifications for conditional quantiles at all $\tau$ 's of interest. Other possible options of norms aggregating the information over the quantiles and $x$, for instance, Kolmogorov-type, are also applicable here. The following proposition presents the asymptotic behavior of the EV test applied for testing non-causality.

Corollary 1.3.1.3 (EV Test for non-causality). Under $\mathbb{H}_{0}$ and Assumptions 1.1-1.3, let $\mathbb{E}\left(\boldsymbol{X}_{t} \boldsymbol{X}_{t}^{\prime}\right)$ be nonsingular in a neighborhood of $\boldsymbol{\theta}(\tau)=\boldsymbol{\theta}_{0}(\tau)$ for all $\tau \in \Upsilon$,

$$
R_{T}^{E V}(x, \tau ; \hat{\boldsymbol{\theta}}) \Longrightarrow \tilde{R}_{\infty}^{E V}(x, \tau)
$$

and
$\int_{\tau \in \Upsilon, x \in \mathbb{R}^{p+1}}\left|R_{T}^{E V}(x, \tau ; \hat{\boldsymbol{\theta}})\right|^{2} d \Phi(x) d W(\tau) \longrightarrow_{d} \int_{\tau \in \Upsilon, x \in \mathbb{R}^{p+1}}\left|\tilde{R}_{\infty}^{E V}(x, \tau)\right|^{2} d \Phi(x) d W(\tau)$
where $\tilde{R}_{\infty}^{E V}=R_{\infty}-\Delta R . R_{\infty}$ is a Gaussian process with mean zero and covariance function defined by

$$
\operatorname{Cov}\left(x_{1}, x_{2} ; \tau_{1}, \tau_{2}\right)=\left(\tau_{1} \wedge \tau_{2}-\tau_{1} \tau_{2}\right) \mathbb{E}\left(\exp \left(i\left(x_{1}-x_{2}\right)^{\prime} \boldsymbol{X}_{0}\right)\right)
$$

and the drift $\Delta R$ is introduced due to the asymptotic effect from estimation error of $\hat{\boldsymbol{\theta}}$,

$$
\Delta R(x, \tau)=G^{\prime}\left(x, \boldsymbol{\theta}_{0}(\tau)\right) Q(\tau)
$$

where $G\left(x, \boldsymbol{\theta}_{0}(\tau)\right)=\mathbb{E}\left[\boldsymbol{X}_{t} f\left(Q_{u_{t}}(\tau)\right) \exp \left(i x^{\prime} \boldsymbol{X}_{t}\right)\right]$ and $Q(\cdot)$ is $\Sigma_{0}^{-1 / 2} \boldsymbol{B}_{p+1} / f\left[F^{-1}(\cdot)\right]$. $\boldsymbol{B}_{p+1}$ is a $(p+1)$-dimensional standard Brownian bridge. $f$ and $F$ are the density and cumulative distribution functions of the innovation $u_{t}$, respectively, and $\Sigma_{0}=$ $\mathbb{E}\left(\boldsymbol{X}_{t} \boldsymbol{X}_{t}^{\prime}\right)$.

The result immediately follows from Escanciano and Velasco (2010), where they develop the result for a general class of quantile estimates covering various linear and nonlinear models with corresponding assumptions. Those conditions are satisfied
under Assumptions 1.1-1.3 in the context of $\operatorname{MAR}(r, s)$ processes under the null hypothesis. The limiting distribution of the test statistic is no longer distributionfree due to the estimation of nuisance parameters. Hence, a subsampling method is proposed to approximate the critical value. The operation for calculating the residual process (1.3.7) and the test statistic (1.3.8) is applied to a given subsample $\left(Y_{t}, \ldots, Y_{t+b}\right)$ of size $b$, denoted by $R_{b, t}^{E V}\left(x, \tau ; \hat{\boldsymbol{\theta}}_{b, t}\right)$ and $\Gamma\left(R_{b, t}^{E V}\right)$ respectively, and repeated for $t=1,2, \ldots, T-b+1$. The cdf of the limiting distribution of the proposed statistic is approximated by the empirical cdf across resamples, i.e.,

$$
\hat{P}\left[\Gamma\left(R_{b, t}^{E V}\right) \leq \omega\right]=\frac{1}{T-b+1} \sum_{t=1}^{T-b+1} I\left(\Gamma\left(R_{b, t}^{E V}\right) \leq \omega\right) .
$$

Therefore, the $1-\alpha$ th sample quantile, $c_{T, b}^{E V}(\alpha)$ defined as

$$
c_{T, b}^{E V}(\alpha)=\inf \left\{\omega: \hat{P}\left[\Gamma\left(R_{b, t}^{E V}\right) \leq \omega\right] \geq 1-\alpha\right\}
$$

intuitively serves as the critical value for this test at the $\alpha$-level of the significance. The validity of this subsampling approach has been verified by Escanciano and Velasco (2010), who suggest an appropriate choice for bandwidth $b=\left\lfloor k T^{2 / 5}\right\rfloor$ for the sake of optimal minimax accuracy ${ }^{9}$. Some numerical evidence has demonstrated that a diverse range of values of $k$, like 4,5 , and 6 provide reasonably good performance in finite samples. A centering strategy can be adopted for the resampling statistic to achieve better performance power-wise in finite samples. Alternatively, Escanciano and Goh (2014) (hereafter EG) translate the restriction (1.3.5) into
$\mathbb{E}\left[\Psi_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}_{0}\right) \mathbb{I}\left(\boldsymbol{X}_{t} \leq x\right)\right]=0 \quad \forall x \in \mathbb{R}^{p+1}$, for some $\boldsymbol{\theta}_{0} \in \mathcal{B}$ and for all $\tau \in \Upsilon \subset(0,1)$.

Naturally, a new test statistic can be constructed on the sample analog of moment conditions (1.3.9) with the replacement of $\boldsymbol{\theta}_{0}$ by their QAR estimates $\hat{\boldsymbol{\theta}}$,

$$
\begin{equation*}
T^{-1 / 2} \sum_{t=1}^{T} \Psi_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \hat{\boldsymbol{\theta}}\right) \mathbb{I}\left(\boldsymbol{X}_{t} \leq x\right) \quad \tau \in \Upsilon, x \in \mathbb{R}^{p+1} \tag{1.3.10}
\end{equation*}
$$

Unlike the approach in the EV test, where the asymptotic behavior of the test statistic is derived by incorporating the non-negligible effect from the estimates of nuisance parameters into the final limiting distribution, Escanciano and Goh (2014) introduce a variant of weighting functions $\mathbb{I}\left(\boldsymbol{X}_{t} \leq x\right)$ satisfying the orthogonality condition for the Taylor expansion of the process (1.3.10) around the true parameter

[^8]$\boldsymbol{\theta}_{0}$. With this consideration, the modified process becomes
$R_{T}^{E G}(x, \tau ; \hat{\boldsymbol{\theta}})=T^{-1 / 2} \sum_{t=1}^{T}\left\{\Psi_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \hat{\boldsymbol{\theta}}\right)\left(\mathbb{I}\left(\boldsymbol{X}_{t} \leq x\right)-\hat{D}_{T}^{\prime}(x, \hat{\boldsymbol{\theta}}(\tau))\left(T^{-1} \sum_{t=1}^{T} \hat{\delta}_{t, \tau} \hat{\delta}_{t, \tau}^{\prime}\right)^{-1} \hat{\delta}_{t, \tau}\right)\right\}$
with $\hat{D}_{T}(x, \hat{\boldsymbol{\theta}}(\tau))=T^{-1} \sum_{t=1}^{T} \hat{\delta}_{t, T} \mathbb{I}\left(\boldsymbol{X}_{t} \leq x\right)$ and
$$
\hat{\delta}_{t, \tau}=\hat{f}\left(\boldsymbol{X}_{t}^{\prime} \hat{\boldsymbol{\theta}}(\tau) \mid \boldsymbol{X}_{t}\right) \boldsymbol{X}_{t},
$$
where $\hat{f}\left(y \mid \boldsymbol{X}_{t}\right)$ is a consistent estimator of the conditional density function of $Y_{t}$ given the past information, $f\left(y \mid \boldsymbol{X}_{t}\right)$. One suggested kernel estimator is proposed by Escanciano and Goh (2012), defined by
\[

$$
\begin{equation*}
\hat{f}\left(\boldsymbol{X}_{t}^{\prime} \hat{\boldsymbol{\theta}}(\tau) \mid \boldsymbol{X}_{t}\right)=\frac{1}{M h_{M}} \sum_{j=1}^{M} K\left(\frac{\boldsymbol{X}_{t}^{\prime} \hat{\boldsymbol{\theta}}(\tau)-\boldsymbol{X}_{t}^{\prime} \hat{\boldsymbol{\theta}}\left(\tau_{j}\right)}{h_{M}}\right), \tag{1.3.12}
\end{equation*}
$$

\]

where $\left\{\tau_{j}\right\}_{j=1}^{M}$ is a sequence randomly selected from $\Upsilon$ following the uniform distribution with $M \longrightarrow \infty$ as well as $T \longrightarrow \infty . K(\cdot)$ is a kernel function, and $h_{M}$ denotes a smoothing parameter which may depend on the data and the quantiles considered for the estimation. Compared to other density estimator candidates, this estimator is computationally less cumbersome but still preserves the same convergence rate as Rosenblatt estimator when the following assumption is imposed for the kernel function and the corresponding smoothing parameter.

## Assumption 1.4. 1. For kernel function $K(s)$ :

(a) $K(s)$ is continuously differentiable;
(b) $\int_{-\infty}^{\infty} K(s)=1$;
(c) $K(s)$ is uniformly bounded;
(d) $K(s)$ is of second order, i.e. $\int_{-\infty}^{\infty} s K(s)=0, \int_{-\infty}^{\infty} s^{2} K(s) d s \in(0, \infty)$ and $\int_{-\infty}^{\infty} K^{2}(s) d s \in(0, \infty)$.
2. The convergence rate of smoothing parameter $h_{M}$ to 0 satisfies $P\left(a_{M} \leq h_{M} \leq b_{M}\right) \rightarrow$ 1, for some deterministic sequences of positive numbers $a_{M}$ and $b_{M}$ such that $b_{M} \longrightarrow 0$ and $a_{M}^{p+2} M / \log M \rightarrow \infty$ as $T \rightarrow \infty$.

These regularity conditions for kernel functions apply to commonly used options in practice, for instance, the Gaussian kernel.

Corollary 1.3.1.4 (EG Test for non-causality). Under $\mathbb{H}_{0}$ and Assumptions 1.11.4, let the matrix $\mathbb{E}\left(\delta_{t, \tau} \delta_{t, \tau}^{\prime}\right)$ be nonsingular in a neighborhood of $\boldsymbol{\theta}(\tau)=\boldsymbol{\theta}_{0}(\tau)$ for all $\tau \in \Upsilon$,

$$
R_{T}^{E G}(x, \tau ; \hat{\boldsymbol{\theta}}) \Longrightarrow R_{\infty}^{E G}(x, \tau),
$$

and
$\int_{\tau \in \Upsilon, x \in \mathbb{R}^{p+1}}\left|R_{T}^{E G}(x, \tau ; \hat{\boldsymbol{\theta}})\right|^{2} d \Phi(x) d W(\tau) \longrightarrow_{d} \int_{\tau \in \Upsilon, x \in \mathbb{R}^{p+1}}\left|R_{\infty}^{E G}(x, \tau)\right|^{2} d \Phi(x) d W(\tau)$,
where $R_{\infty}^{E G}$ is a Gaussian process with mean zero and covariance function characterized by

$$
\left(\tau_{1} \wedge \tau_{2}-\tau_{1} \tau_{2}\right) \mathbb{E}\left\{\Pi_{\tau_{1}} \mathbb{I}\left(\boldsymbol{X}_{t} \leq x_{1}\right) \Pi_{\tau_{2}} \mathbb{I}\left(\boldsymbol{X}_{t} \leq x_{2}\right)\right\}
$$

with the so-called orthogonal projection operator on the weighting function

$$
\Pi_{\tau} \mathbb{I}\left(\boldsymbol{X}_{t} \leq x\right) \equiv \mathbb{I}\left(\boldsymbol{X}_{t} \leq x\right)-D^{\prime}\left(x, \boldsymbol{\theta}_{0}(\tau)\right) \mathbb{E}^{-1}\left(\delta_{t, \tau} \delta_{t, \tau}\right) \delta_{t, \tau}
$$

and $\delta_{t, \tau}=f\left(\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}_{0}(\tau) \mid \boldsymbol{X}_{t}\right) \boldsymbol{X}_{t}, D\left(x, \boldsymbol{\theta}_{0}(\tau)\right)=\mathbb{E}\left(\delta_{t, \tau} \mathbb{I}\left(\boldsymbol{X}_{t} \leq x\right)\right)$.

As stated before, the main advantage of the EG test is the limiting distribution of the test statistic being invariant to the estimation effect of $\hat{\boldsymbol{\theta}}(\tau)$. Therefore, compared to the asymptotic distribution of the EV test, there is no "drift" term subtracted from a Gaussian process. However, this asymptotic distribution still depends on the data-generating process. Consequently, we cannot tabulate the critical values for the considered test statistic. This is coped with the aid of a multiplier bootstrap approach. The approximation based on a transformation on $R_{T}^{E G}(x, \tau ; \hat{\boldsymbol{\theta}})$ is obtained by multiplying by a sequence of $i i d$ random variables $\left\{W_{t}\right\}_{t=1}^{T}$ with zero mean and unit variance, independent on $\boldsymbol{X}_{t}$,

$$
\begin{align*}
& \tilde{R}_{T}^{E G}(x, \tau ; \hat{\boldsymbol{\theta}}) \\
= & T^{-1 / 2} \sum_{t=1}^{T}\left\{\Psi_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \hat{\boldsymbol{\theta}}\right)\left(\mathbb{I}\left(\boldsymbol{X}_{t} \leq x\right)-\hat{D}_{T}^{\prime}(x, \hat{\boldsymbol{\theta}}(\tau))\left(T^{-1} \sum_{t=1}^{T} \hat{\delta}_{t, \tau} \hat{\delta}_{t, \tau}^{\prime}\right)^{-1} \hat{\delta}_{t, \tau}\right)\right\} W_{t} . \tag{1.3.13}
\end{align*}
$$

One common choice of $\left\{W_{t}\right\}_{t=1}^{T}$ is

$$
\left\{\begin{array}{l}
P(W=1-\omega)=\omega / \sqrt{5}  \tag{1.3.14}\\
P(W=\omega)=1-\omega / \sqrt{5}, \text { with } \omega=(\sqrt{5}+1) / 2
\end{array}\right.
$$

This transformation has been proven by Escanciano and Goh (2014) to restore the limiting distribution of the original statistic. It allows us to use the empirical distribution of any continuous functional, including $\operatorname{CvM}$ form, $\Gamma\left(\tilde{R}_{T}^{E G}\right)$, i.e.,

$$
\hat{P}\left[\Gamma\left(\tilde{R}_{T}^{E G}\right) \leq \omega \mid\left\{W_{t}\right\}_{t=1}^{T}\right]=\frac{1}{N} \sum_{i=1}^{N} I\left(\Gamma\left(\tilde{R}_{T}^{E G}\right) \leq \omega\right)
$$

to consistently estimate the limiting distribution of the original statistic $\Gamma\left(R_{T}^{E G}\right)$,
where $N$ is the number of the replication of sequences $\left\{W_{t}\right\}_{t=1}^{T}$ for the calculation of the critical value in the multiplier bootstrap. Likewise, the $(1-\alpha)$-th empirical quantile of the transformed statistic,

$$
c_{T}^{E G}(\alpha)=\inf \left\{\omega: \hat{P}\left[\Gamma\left(\tilde{R}_{T}^{E G}\right) \leq \omega \mid\left\{W_{t}\right\}_{t=1}^{T}\right] \geq 1-\alpha\right\}
$$

will is a consistent estimate of the critical value at $\alpha$ significance level.
In the presence of non-causality, $\psi\left(L^{-1}\right)$ does not vanish from general $\operatorname{MAR}(r, s)$ processes. Then, by Corollary 1.3.1.1,

$$
\mathbb{E}\left(\Psi_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}_{1}(\cdot)\right) \exp \left(i \cdot \boldsymbol{X}_{t}\right)\right) \neq 0
$$

and

$$
\mathbb{E}\left(\Psi_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}_{1}(\cdot)\right) \mathbb{I}\left(\boldsymbol{X}_{t} \leq \cdot\right)\right) \neq 0
$$

in a set with a positive Lebesgue measure on $\mathbb{R}^{p+1} \times \Upsilon$, provided that $\Phi$ and $W$ are absolutely continuous on $\mathbb{R}^{p+1} \times \Upsilon$ with respect to the Lebesgue measure. Correspondingly, under the alternative,

$$
\Gamma\left(R_{T}^{E V}\right)=\int_{\tau \in \Upsilon, x \in \mathbb{R}^{p+1}}\left|R_{T}^{E V}(x, \tau ; \hat{\boldsymbol{\theta}})\right|^{2} d \Phi(x) d W(\tau) \rightarrow_{p} \infty
$$

and

$$
\Gamma\left(\tilde{R}_{T}^{E G}\right)=\int_{\tau \in \Upsilon, x \in \mathbb{R}^{p+1}}\left|\tilde{R}_{T}^{E G}(x, \tau ; \hat{\boldsymbol{\theta}})\right|^{2} d \Phi(x) d W(\tau) \rightarrow_{p} \infty
$$

so both specification-based tests are consistent.

### 1.4. Monte Carlo Simulations

In this section, we study the performance of the three proposed tests in finite samples and compare them with each other in terms of size and power.

Constancy Test In the first experiment, we focus on the approach based on the constancy test. In the simulation, we consider a pair of $\operatorname{MAR}(1,0)$ and $\operatorname{MAR}(0,1)$ models

$$
\left\{\begin{array}{l}
(1-\phi L) Y_{t}=u_{t}  \tag{1.4.1}\\
\left(1-\psi L^{-1}\right) Y_{t}^{*}=u_{t}
\end{array}\right.
$$

which are generated from 11 non-Gaussian distributed innovations, which cover a variety of distributions commonly used in the empirics, ranging from symmetric to asymmetric, bounded to unbounded support, with mixed types of tail behavior. The parameters $(\phi, \psi)$ with values $(0.3,0.6,0.9)$ enable us to investigate the sensitivity of the method responding to data-generating processes with different persistence.

The sample sizes are 200 and 500 with 500 replications. $\epsilon=0.05$ is the default choice for the quantile interval $\Upsilon=[\epsilon, 1-\epsilon] \subset(0,1)$. The empirical size and power of rejecting the constancy hypothesis to detect non-causality under the QAR framework are summarized in Table 1.4.1.

Regarding the size, the constancy test has an empirical size close to the nominal level in most cases but suffers from a severe over-rejection for heavy-tailed distributions. This corresponds to the estimation of conditional quantiles of these processes when $\tau$ is extremely close to 0 or 1 , which generally calls for a larger sample size to produce less biased and more stable estimates. The volatility of QAR estimates of extremal quantiles triggers the over-rejection of the constant coefficients under the causality hypothesis. Therefore, as seen in Figures 1.4.1c and 1.4.1d, where innovations follow truncated Cauchy distribution ${ }^{10}$ and log-normal distribution, respectively, the QAR estimates at quantiles close to 0 and 1 turn rather volatile compared to the estimates at other levels in $(0,1)$. To alleviate this issue, we propose to check different trimming strategies in the quantile interval for the test. There is an obvious trade-off in the selection of truncation of the quantile interval: over-trimmed, the power of the test will decrease due to the loss of valuable information; under-trimmed, the volatile estimate of extremal quantiles is not excluded, so the distortion in size will remain as before. Consequently, we conduct some experiments to analyze the sensitivity of the constancy test in response to truncated intervals; see Table 1.4.2. The results suggest a trimmed quantile interval [0.10, 0.90] would be appropriate for the sample size considered because the power remains relatively high while the size is close to the nominal level. Another possible reason highlighted by Koenker and Xiao (2002) is that the default smoothing parameter selection for estimating the density function of innovations, which comprises the test statistic in the KhmaladzeTest command in R studio, produces satisfactory performance for the class of distributions considered there, but is not designed for heavy-tailed distributed innovations. A more adaptive bandwidth choice of density function estimation of heavy-tailed distributions at extreme quantiles needs further investigation.

From the perspective of power, the test achieves a significant leap in power as the sample size increases from 200 to 500 in most scenarios. More particularly, the method provides favorable performance in the presence of asymmetry in the distributions of innovations with rejection rates of $40 \% \sim 90 \%$ in 200 -sized samples and $70 \% \sim 90 \%$ in 500 -sized samples, respectively. This finding coincides with the idea shared in Velasco and Lobato (2018), where the third-order moments contribute most to the identification of non-causal AR processes. This phenomenon is exceptionally well-illustrated in the first four cases (Exponential, Gamma, Beta, and F

[^9]distributions) when the distribution of innovations is skewed but has no heavy tails. However, in the cases where innovations do not display skewness or heavy-tailedness, the test can barely distinguish non-causal processes from their causal counterparts, see Figure 1.4.2. The point of the failure is illustrated in the same figure. The QAR estimates for non-causal $A R(1)$ with symmetric innovations appear to be indistinguishable from the ones for the causal counterparts, even under misspecification.

Table 1.4.1: Empirical size and power of non-causality test using constancy* test in QAR in various cases

| Distribution | test | $\mathrm{T}=200$ |  |  | $\mathrm{~T}=500$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $u_{t}$ |  | $\phi(\psi)=0.3^{\dagger}$ | $\phi(\psi)=0.6$ | $\phi(\psi)=0.9$ | $\phi(\psi)=0.3$ | $\phi(\psi)=0.6$ | $\phi(\psi)=0.9$ |
| $\operatorname{Eapp}(1)-1$ | size | $4.20 \%$ | $4.00 \%$ | $4.80 \%$ | $6.40 \%$ | $6.80 \%$ | $4.80 \%$ |
|  | power | $33.80 \%$ | $38.80 \%$ | $40.20 \%$ | $65.40 \%$ | $69.60 \%$ | $72.40 \%$ |
| Gamma(1,1)-1 | size | $4.40 \%$ | $3.00 \%$ | $3.80 \%$ | $5.00 \%$ | $5.80 \%$ | $4.40 \%$ |
|  | power | $34.40 \%$ | $37.00 \%$ | $42.20 \%$ | $64.40 \%$ | $68.40 \%$ | $70.20 \%$ |
| Beta(5,1)-5/6 | size | $6.80 \%$ | $6.60 \%$ | $5.40 \%$ | $6.80 \%$ | $9.40 \%$ | $6.40 \%$ |
|  | power | $15.40 \%$ | $27.20 \%$ | $39.00 \%$ | $24.80 \%$ | $65.00 \%$ | $77.20 \%$ |
| $\mathrm{~F}(5,5)-5 / 3$ | size | $5.00 \%$ | $6.60 \%$ | $3.40 \%$ | $8.40 \%$ | $6.00 \%$ | $4.20 \%$ |
|  | power | $80.00 \%$ | $81.80 \%$ | $70.00 \%$ | $97.20 \%$ | $98.20 \%$ | $96.20 \%$ |
| $\chi_{5}^{2}-5$ | size | $5.00 \%$ | $4.40 \%$ | $4.00 \%$ | $4.00 \%$ | $4.80 \%$ | $5.20 \%$ |
|  | power | $11.40 \%$ | $11.40 \%$ | $8.80 \%$ | $27.40 \%$ | $35.60 \%$ | $17.40 \%$ |
| skewed normal | size | $5.20 \%$ | $6.60 \%$ | $7.40 \%$ | $9.80 \%$ | $7.40 \%$ | $9.60 \%$ |
|  | power | $8.00 \%$ | $7.80 \%$ | $10.60 \%$ | $25.40 \%$ | $20.00 \%$ | $12.80 \%$ |
| truncated Cauchy | size | $26.60 \%$ | $34.80 \%$ | $43.00 \%$ | $20.80 \%$ | $19.20 \%$ | $33.20 \%$ |
|  | power | $70.00 \%$ | $96.40 \%$ | $92.40 \%$ | $80.00 \%$ | $99.80 \%$ | $91.80 \%$ |
| log normal | size | $21.20 \%$ | $19.60 \%$ | $23.60 \%$ | $23.20 \%$ | $26.40 \%$ | $26.60 \%$ |
|  | power | $98.20 \%$ | $99.00 \%$ | $99.60 \%$ | $100.00 \%$ | $100.00 \%$ | $100.00 \%$ |
| $t_{3}$ | size | $5.40 \%$ | $4.00 \%$ | $5.00 \%$ | $4.00 \%$ | $6.80 \%$ | $4.80 \%$ |
|  | power | $9.80 \%$ | $13.00 \%$ | $15.80 \%$ | $11.60 \%$ | $20.00 \%$ | $25.80 \%$ |
| uniform | size | $6.20 \%$ | $7.60 \%$ | $6.00 \%$ | $6.00 \%$ | $6.80 \%$ | $5.80 \%$ |
|  | power | $13.40 \%$ | $8.00 \%$ | $13.20 \%$ | $40.40 \%$ | $7.80 \%$ | $48.80 \%$ |
| Laplace | size | $3.40 \%$ | $4.00 \%$ | $4.60 \%$ | $2.60 \%$ | $3.80 \%$ | $4.60 \%$ |
|  | power | $8.40 \%$ | $5.60 \%$ | $17.00 \%$ | $10.80 \%$ | $8.60 \%$ | $29.00 \%$ |

$\star$ : constancy test over the quantile interval $[0.05,0.95]$.
$\dagger: \phi(\psi)=0.3$ means the coefficient in the lag polynomial of the $\operatorname{MAR}(1,0)$ process (purely causal) is 0.3 ; the coefficient in the lead polynomial of the $\operatorname{MAR}(0,1)$ process (purely non-causal) is 0.3 as well, for comparison.

EV Test With the same setting as the one in the Constancy Test, we take the case with the coefficient $\phi(\psi)=0.6$ for $\operatorname{MAR}(1,0)(\operatorname{MAR}(0,1))$ as the representative to examine the performance of EV test in discriminating non-causal processes from their causal alternatives. The sample size varies from 100 to 200 and 500. The number of replications is 500. Regarding the bandwidth for the subsampling scheme to approximate the critical value, we choose $b=\left\lfloor 4 T^{2 / 5}\right\rfloor$. As for the weighting functions involved in the expression (1.3.8), a uniform distribution is applied to $W(\tau)$ defined on the evenly discretized quantile interval $\Upsilon=[0.01,0.99]$ and 2dimensional standard normal distribution to $\Psi(x)$ for the sake of simplicity in the

Figure 1.4.1: Four cases of $\operatorname{QAR}(1)$ on $\operatorname{MAR}(1,0)$ and $\operatorname{MAR}(0,1), \mathrm{T}=500, \phi(\psi)=0.6$ : asymmetric distributions


Figure 1.4.2: Three cases of $\operatorname{QAR}(1)$ on $\operatorname{MAR}(1,0)$ and $\operatorname{MAR}(0,1), T=500, \phi(\psi)=0.6$ : symmetric distributions


(c) Laplace distribution

Table 1.4.2: Empirical size and power of non-causality test using constancy test with different trimmed quantile interval

| Sample size | $[0.05,0.95]$ | $\phi(\psi)=0.3$ | $\phi(\psi)=0.6$ | $\phi(\psi)=0.9$ | $\phi(\psi)=-0.4$ | $\phi(\psi)=-0.6$ | $\phi(\psi)=-0.8$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathrm{~T}=100$ | size | $2.40 \%$ | $2.20 \%$ | $2.60 \%$ | $2.80 \%$ | $2.00 \%$ | $2.40 \%$ |
|  | power | $22.40 \%$ | $24.60 \%$ | $22.00 \%$ | $4.80 \%$ | $32.80 \%$ | $26.60 \%$ |
| $\mathrm{~T}=200$ | size | $3.60 \%$ | $3.60 \%$ | $4.00 \%$ | $5.00 \%$ | $2.80 \%$ | $4.40 \%$ |
|  | power | $40.80 \%$ | $38.60 \%$ | $41.80 \%$ | $8.40 \%$ | $48.00 \%$ | $44.00 \%$ |
| $\mathrm{~T}=500$ | size | $3.40 \%$ | $6.20 \%$ | $4.80 \%$ | $6.60 \%$ | $4.80 \%$ | $3.40 \%$ |
|  | power | $67.20 \%$ | $67.00 \%$ | $72.40 \%$ | $14.20 \%$ | $66.00 \%$ | $72.20 \%$ |
| Sample size | $[0.10,0.90]$ | $\phi(\psi)=0.3$ | $\phi(\psi)=0.6$ | $\phi(\psi)=0.9$ | $\phi(\psi)=-0.4$ | $\phi(\psi)=-0.6$ | $\phi(\psi)=-0.8$ |
| $\mathrm{~T}=100$ | size | $3.40 \%$ | $3.20 \%$ | $3.20 \%$ | $3.80 \%$ | $2.40 \%$ | $3.40 \%$ |
|  | power | $28.00 \%$ | $25.40 \%$ | $22.80 \%$ | $6.00 \%$ | $32.80 \%$ | $28.20 \%$ |
| $\mathrm{~T}=200$ | size | $3.20 \%$ | $2.80 \%$ | $5.00 \%$ | $3.00 \%$ | $4.20 \%$ | $4.00 \%$ |
|  | power | $37.40 \%$ | $43.40 \%$ | $43.20 \%$ | $5.80 \%$ | $49.40 \%$ | $48.80 \%$ |
| $\mathrm{~T}=500$ | size | $4.80 \%$ | $5.40 \%$ | $6.00 \%$ | $4.60 \%$ | $3.80 \%$ | $5.00 \%$ |
|  | power | $71.20 \%$ | $67.80 \%$ | $69.20 \%$ | $14.60 \%$ | $65.40 \%$ | $72.80 \%$ |
| Sample size | $[0.15,0.85]$ | $\phi(\psi)=0.3$ | $\phi(\psi)=0.6$ | $\phi(\psi)=0.9$ | $\phi(\psi)=-0.4$ | $\phi(\psi)=-0.6$ | $\phi(\psi)=-0.8$ |
| $\mathrm{~T}=100$ | size | $4.40 \%$ | $2.60 \%$ | $3.60 \%$ | $5.00 \%$ | $3.00 \%$ | $3.00 \%$ |
|  | power | $27.80 \%$ | $28.20 \%$ | $21.80 \%$ | $8.80 \%$ | $37.60 \%$ | $30.20 \%$ |
| $\mathrm{~T}=200$ | size | $5.60 \%$ | $3.80 \%$ | $3.60 \%$ | $6.60 \%$ | $3.40 \%$ | $4.40 \%$ |
|  | power | $39.00 \%$ | $40.00 \%$ | $42.40 \%$ | $9.60 \%$ | $51.60 \%$ | $46.60 \%$ |
| $\mathrm{~T}=500$ | size | $5.40 \%$ | $4.40 \%$ | $5.00 \%$ | $4.20 \%$ | $4.80 \%$ | $5.40 \%$ |
|  | power | $64.80 \%$ | $65.80 \%$ | $69.00 \%$ | $21.80 \%$ | $68.00 \%$ | $71.60 \%$ |

The innovations follow an exponential distribution
calculation. The results of the empirical size and power are displayed in Table 1.4.3. Concerning size, the results present some fluctuations around the nominal level. The test tends to under-reject the correct null hypothesis in the light-tailed scenarios while over-rejects in the heavy-tailed scenarios. This comes from the subjective choice in the subsampling size, whose performance can be sensitive to the quantile interval included in the test and the data-generating process. Nevertheless, the distortion in size is not so significant, and we believe this can be eased by choosing different subsampling schemes. As for power, the test achieves reasonably good performance generally. From the simulation results, we observe that the power of detecting non-causality is close to $100 \%$ in most cases when the sample size is 500 , which is as expected. It is worth noting that the convergence rate of power towards $100 \%$ varies across different distributions. The numerical evidence indicates that the further the innovations depart from Gaussianity, the faster the convergence rate is. In such log-normal and uniform distributions, a reasonably high power, like $86 \%$ or $92 \%$, has been obtained in relatively small samples. While others (chi-square) may need larger samples to reach the same level of power

EG Test The setup of DGP keeps unchanged, like in the EV test. The sample size varies from 50 to 100 and 200. The weighting functions $\Psi(x)$ and $W(\tau)$ in the CvM form of $R_{T}^{E G}(x, \tau ; \hat{\boldsymbol{\theta}})$ are chosen to be the empirical distribution of $\boldsymbol{X}_{t}$ and uniform distribution over the grid of quantiles from $\Upsilon=[0.01,0.99]$ considered in the esti-

Table 1.4.3: Empirical size and power of non-causality test using EV test in QAR in various cases

| Distribution | parameter: |  | $\phi(\psi)=0$ |  |
| :---: | :---: | :---: | :---: | :---: |
| $u_{t}$ |  | 100 | 200 | 500 |
| Gaussian | size | 2.20\% | 3.40\% | 3.40\% |
|  | power | 0.40\% | 0.60\% | 2.60\% |
| exponential | size power | 4.40\% | 1.80\% | 2.00\% |
|  |  | 29.20\% | 43.20\% | 97.20\% |
| Gamma | size | 1.80\% | 3.00\% | 2.80\% |
|  | power | 10.60\% | 41.60\% | 97.60\% |
| Beta | size | 3.20\% | 3.20\% | 3.40\% |
|  | power | 26.40\% | 67.80\% | 99.00\% |
| F | size | 3.60\% | 5.80\% | 3.40\% |
|  | power | 24.00\% | 67.60\% | 99.00\% |
| $\chi_{5}^{2}-5$ | size power | 4.80\% | 4.40\% | 3.60\% |
|  |  | 13.40\% | 22.20\% | 68.40\% |
| log normal | size | 5.40\% | 4.60\% | 6.80\% |
|  | power | 54.00\% | 92.40\% | 100.00\% |
| $t_{3}$ | size | 6.20\% | 7.00\% | 6.40\% |
|  | power | 13.80\% | 42.40\% | 93.40\% |
| Uniform | size | 5.00\% | 6.00\% | 3.40\% |
|  | power | 44.00\% | 86.80\% | 100.00\% |
| Laplace | size power | 4.40\% | 4.80\% | 4.40\% |
|  |  | 14.00\% | 47.20\% | 98.20\% |

mation. The critical value is obtained through the multiplier bootstrap introduced in the methodology section. The number of iid sequences of multipliers $\left\{W_{t}\right\}_{t=1}^{T}$ is 200. Implementing the multiplier bootstrap avoids computing the estimates for each subsample. The results are summarized in Table 1.4.4. Regarding the empirical size performance, this approach delivers stable rejection frequencies under the null hypothesis associated with their nominal level in all cases considered in the simulation exercise. This is anticipated in accordance with the argument in the previous section that there is no subjective choice involved in the approximation of the critical value. In terms of power, the EG test has an increasing trend as the sample is enlarged. Similar to the EV test, the EG approach outperforms in the presence of skewness and excess kurtosis (regardless of negative or positive), with a rejection probability over $70 \%$ in relatively small samples (200). For the cases where the performance is less satisfactory, such as $t_{3}$ and Laplace distributions, power still increases when the sample size becomes larger. An overall comparison of the three methods is exhibited in Table 1.4.5. Size-wise, the EG test has an appealing attribute of undistorted size in general scenarios compared to the other two approaches. On the contrary, the constancy test suffers from over-rejection in heavy-tailed cases, and the EV test delivers less accuracy than the EG test in some cases. Power-wise, the EG test is the most robust one as it produces relatively good results in all situations but is extraordinarily competent for asymmetric distributions. By contrast, the EV test achieves the highest power among the three in the symmetric cases. In comparison, the con-

Table 1.4.4: Empirical size and power of non-causality test using EG test in QAR in various cases

| Distribution | parameter: |  | $\phi(\psi)=0.6$ |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  |  | $\mathrm{~T}:$ | 50 | 100 |  |
| Gaussian | size | $6.00 \%$ | $5.00 \%$ | $4.00 \%$ |  |
|  | power | $4.80 \%$ | $5.60 \%$ | $5.80 \%$ |  |
| exponential | size | $5.20 \%$ | $4.60 \%$ | $6.00 \%$ |  |
|  | power | $24.80 \%$ | $49.20 \%$ | $76.40 \%$ |  |
| Gamma | size | $5.00 \%$ | $5.60 \%$ | $5.00 \%$ |  |
|  | power | $23.80 \%$ | $45.80 \%$ | $75.80 \%$ |  |
| Beta | size | $6.40 \%$ | $5.80 \%$ | $6.00 \%$ |  |
|  | power | $24.20 \%$ | $42.20 \%$ | $75.80 \%$ |  |
| F | size | $4.40 \%$ | $5.00 \%$ | $5.40 \%$ |  |
|  | power | $22.60 \%$ | $37.40 \%$ | $67.00 \%$ |  |
|  | size | $6.60 \%$ | $5.40 \%$ | $4.60 \%$ |  |
|  | power | $12.80 \%$ | $22.60 \%$ | $41.60 \%$ |  |
|  | size | $5.60 \%$ | $5.40 \%$ | $4.00 \%$ |  |
|  | power | $32.00 \%$ | $60.60 \%$ | $85.80 \%$ |  |
| $t_{3}$ | size | $4.40 \%$ | $6.60 \%$ | $7.00 \%$ |  |
|  | power | $6.80 \%$ | $14.80 \%$ | $36.20 \%$ |  |
|  | size | $4.80 \%$ | $6.60 \%$ | $6.60 \%$ |  |
|  | power | $10.20 \%$ | $25.40 \%$ | $64.80 \%$ |  |

EG test: critical value obtained from multiplier bootstrap.
stancy test can be a powerful tool in detecting non-causality in processes with heavy tails. However, given that the consistency of the constancy test for non-causality is not guaranteed and the size is distorted, it may give misleading conclusions in empirical analysis. Thus, the constancy test can only be considered a preliminary test to check whether the process is likely non-causal, followed by the implementation of the EV or EG tests, which serve as the formal tests for non-causality. In practice, a combination of the constancy test and EV (EG) test is suggested.

### 1.5. Empirical Applications

In this section, we apply our non-causality tests to six financial series studied in Fries and Zakoïan (2019): cotton price, soybean price, sugar price, coffee price, Hang Seng Index (HSI), and Shiller Price/Earning ratio (Shiller PE), where single or multiple spikes and asymmetric dynamics are exhibited. Fries and Zakoïan (2019) found numerical evidence in favor of $\operatorname{MAR}(r, s)$ models in fitting time series with local explosiveness phases compared to purely causal AR models. The frequency of data is quarterly for the Shiller PE series and monthly for the rest ${ }^{11}$. The trajectories of these series are displayed in Figure 1.5.1.

[^10]Table 1.4.5: Comparison of QAR-based non-causality tests

| Distribution <br> $u_{t}$ | test type: T : | constancy test |  | EV test |  | EG test |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 100 | 200 | 100 | 200 | 100 | 200 |
| Gaussian | size | 2.80\% | 5.40\% | 2.20\% | 3.40 \% | 5.00\% | 4.00\% |
|  | power | 4.20\% | $3.40 \%$ | 0.40\% | 0.60\% | 5.60\% | 5.80\% |
| exponential | size | 3.20\% | 4.00\% | 4.40\% | 1.80\% | 4.60\% | 6.00\% |
|  | power | 25.40\% | 38.80\% | 29.20\% | 43.20\% | 49.20\% | 76.40\% |
| Gamma | size | 3.80\% | 3.00\% | 1.80\% | 3.00\% | 5.60\% | 5.00\% |
|  | power | 26.60\% | 37.00\% | 10.60\% | 41.60\% | 45.80\% | 75.80\% |
| Beta | size | 4.20\% | 6.60\% | 3.20\% | 3.20\% | 5.80\% | 6.00\% |
|  | power | 18.60\% | 27.20\% | 26.40\% | 67.80\% | 42.20\% | 75.80\% |
| F | size | 6.80\% | 6.60\% | 3.60\% | 5.80\% | 5.00\% | 5.40\% |
|  | power | 62.40\% | 81.80\% | 24.00\% | 67.60\% | 34.70\% | 67.00\% |
| $\chi_{5}^{2}$ | size | 5.80\% | 4.40\% | 4.80\% | 4.40\% | 5.40\% | 4.60\% |
|  | power | 9.60\% | 11.40\% | 13.40\% | 22.20\% | 22.60\% | 41.60\% |
| log normal | size | 20.20\% | 19.60\% | 5.40\% | 4.60\% | 5.40\% | 4.00\% |
|  | power | 78.80\% | 99.60\% | 54.00\% | 92.40\% | 60.60\% | 85.80\% |
| $t_{3}$ | size | 3.40\% | 4.00\% | 6.20\% | 7.00\% | 6.60\% | 7.00\% |
|  | power | 10.20\% | 13.00\% | 13.80\% | 42.40\% | 14.80\% | 36.20\% |
| Uniform | size | 6.80\% | 7.60\% | 5.00\% | 6.00\% | 6.60\% | 6.60\% |
|  | power | 7.40\% | 8.00\% | 44.00\% | 86.80\% | 25.40\% | 64.80\% |
| Laplace | size | 5.80\% | 4.00\% | 4.40\% | 4.80\% | 5.00\% | 6.80\% |
|  | power | 5.20\% | 5.60\% | 14.00\% | 47.20\% | 14.40\% | 41.60\% |

Constancy test: with trimmed quantile interval [0.05, 0.95]

Before proceeding to our testing strategies, we must ensure that the series is stationary. The augmented Dickey-Fuller tests indicate no unit roots in the cotton, soybean, sugar, and coffee price. Neither the evidence of unit root is found in the series of HSI after a linear trend is subtracted. The stationarity of the Shiller PE series is secured after the first difference in the levels. The sample partial autocorrelation function for each series is computed, see Figure 1.5.2, to determine the order of the lag orders: cotton: 2; soybean: 2; sugar: 4; coffee: 3; HSI: 1; Shiller PE: 7 . Three non-causality testing procedures: the constancy test, the EV test, and the EG test, are applied to each series, respectively. The corresponding results are reported in Table 1.5.1. For the constancy test, we consider two trimmed quantile intervals: $[0.05,0.95]$ and $[0.10,0.90]$ to mitigate the instability effect on the power from the subjective choice in the quantiles. Both in the cases of cotton and sugar series, a significant fluctuation in the coefficients is observed. Yet no strong evidence against linear quantile specification is found based on the EV or EG test. The strong rejection in the constancy test for these series may result from the over-rejection issue in heavy-tailed scenarios, which makes the results from the EV or EG tests more reliable. This does not deviate much from the results obtained by Fries and Zakoïan (2019), where they found one non-causal root (0.94) in the cotton series and a root (0.92) in the sugar series. As seen in the numerical experiments, a non-causal

Figure 1.5.1: Financial series trajectories


AR model with coefficients near the unity makes it harder to distinguish from its causal counterpart, as well as generates less distinct dynamics than its causal counterpart. For three series (soybean, coffee, and Shiller PE), the tests show strong evidence favoring non-causal processes at $5 \%$ or even $1 \%$ level from all three testing strategies. For HSI, the test shows mild evidence of non-causality at the significance level of $10 \%$ based on the EG test. This result is compatible with the conclusion drawn in Fries and Zakoïan (2019), where mixed models with nontrivial non-causal components ${ }^{12}$ are selected.

### 1.6. Extensions

### 1.6.1. Some Extensions

So far, the preceding discussion has been confined to $\operatorname{MAR}(r, s)$ driven by iid innovations. Within this framework, the only possible source of non-linearity in $\operatorname{MAR}(r, s)$ is non-causality, which contributes to the consistency of the test in the aforementioned methods. However, stylized nonlinear dynamics like conditional heteroskedas-

[^11]Figure 1.5.2: Sample partial autocorrelation functions of six financial series


Table 1.5.1: Non-causality tests for six financial series in Fries and Zakoïan (2019)

| Financial series | total AR order $(r+s)$ | test type: | constancy test |  | EV test | EG test |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | [0.05, 0.95] ${ }^{\dagger}$ | [0.10, 0.90] |  |  |
| Cotton | 2 | statistic | $6.897^{* *}$ | 4.738** | 0.012 | 0.025 |
|  |  | critical value ${ }^{a}$ | (3.393) | (3.287) | (0.046) | (0.032) |
| Soybean | 2 | statistic | 4.703** | 4.445** | $0.209^{* *}$ | $0.027^{* *}$ |
|  |  | critical value | (3.393) | (3.287) | (0.158) | (0.021) |
| Sugar | 4 | statistic | $306.043^{* *}$ | 144.025** | 0.009 | 0.065 |
|  |  | critical value | (5.560) | (5.430) | (0.102) | (0.100) |
| coffee | 3 | statistic | 8.457** | 6.992** | 0.149** | $0.044^{* *}$ |
|  |  | critical value | (4.523) | (4.383) | (0.089) | (0.025) |
| Hang Seng Index | 1 | statistic | 0.017 | 0.012 | 0.168 | 0.078* |
|  |  | critical value | (2.140) | (2.102) | (0.176) | (0.083) |
| Shiller's P/E ratio | 7 | statistic | $15.105^{* *}$ | $9.027^{* *}$ | $0.197^{* *}$ | $0.135^{* *}$ |
|  |  | critical value | (8.578) | (8.368) | (0.189) | (0.066) |

$\dagger$ the trimmed quantile interval considered in the constancy test.
** stands for significance at level $5 \%$ and $*$ stands for significance at $10 \%$.
the critical value at $10 \%$ significance level for the EG test in the case of Hang Seng Index is 0.061.
${ }^{a}$ the default level of significance is $5 \%$
ticity or asymmetric dynamics are prevalently observed in the financial and macroeconomic data, which renders it more demanding to detect non-causality in a time series process. A more robust methodology applicable to AR models accommodating non-linear features needs investigation. The critical point is how to disentangle non-linearity induced by non-causality from the other alternatives. If these nonlinear features can be captured by a parametric model, one plausible solution is to incorporate these non-linear terms into the baseline model (1.2.1). Below we list some possible extensions where this strategy is employed.

Asymmetric Dynamics This can be solved by allowing varying coefficients in the $\operatorname{MAR}(r, s)$ model in the spirit of the random coefficient model, defined by

$$
\begin{equation*}
\tilde{\phi}(L) \tilde{\psi}\left(L^{-1}\right) Y_{t}=u_{t} \tag{1.6.1}
\end{equation*}
$$

where $\tilde{\phi}(L)=1-\phi_{1}\left(U_{t}\right) L-\cdots-\phi_{r}\left(U_{t}\right) L^{r}$ and $\tilde{\psi}=1-\psi_{1}\left(U_{t}\right) L^{-1}-\cdots-\psi_{s}\left(U_{t}\right) L^{-s}$, $U_{t}$ is an iid sequence of random variables following standard uniform distribution, and $u_{t}$ is an iid innovation sequence satisfying Assumptions 1.1. Denote

$$
\Omega_{c}=\left(\begin{array}{cccc}
\phi_{1}\left(U_{t}\right) & \ldots & \phi_{r-1}\left(U_{t}\right) & \phi_{r}\left(U_{t}\right) \\
& \boldsymbol{I}_{r-1} & & \mathbf{0}_{(r-1)}
\end{array}\right)
$$

and

$$
\Omega_{n c}=\left(\begin{array}{cccc}
\psi_{1}\left(U_{t}\right) & \ldots & \psi_{s-1}\left(U_{t}\right) & \psi_{s}\left(U_{t}\right) \\
& \boldsymbol{I}_{s-1} & & \mathbf{0}_{(s-1)}
\end{array}\right)
$$

Similar to the $p$-th order autoregressive process, which is designed to accommodate asymmetric dynamics in Koenker and Xiao (2006) for linear QAR model, we need
to assume $\mathbb{E}\left(\Omega_{c} \otimes \Omega_{c}\right)$ and $\mathbb{E}\left(\Omega_{n c} \otimes \Omega_{n c}\right)$ have eigenvalues with moduli less than one. This equation (1.6.1) is able to mimic asymmetric dynamics since $\phi_{j}, \psi_{j}$ 's are functions $[0,1] \rightarrow \mathbb{R}$. The definition of non-causality in this context will be adapted to that $\tilde{\psi}\left(L^{-1}\right)$ does not decline to constant. In the causal situation, this model (1.6.1) works like a random coefficient model with lags. The linearity of the conditional mean is restored, and the linear quantile dynamic model with varying coefficients over different quantiles remains the correct specification for conditional quantiles of $Y_{t}$. On the other hand, when the process is non-causal, it is conceivable that linearity will not hold anymore. Therefore, the methodology relying on the specification tests is applicable here.

Volatility Clustering Concerning volatility clustering, which is routinely modeled by the quadratic ARCH/GARCH model in squared residuals. Another popular choice is to replace the squared value with the absolute value suggested by Taylor (2008) and make the model a linear ARCH.

$$
\begin{align*}
& \phi(L) \psi\left(L^{-1}\right) Y_{t}=v_{t} \text { where } v_{t}=\sigma_{t} u_{t} \\
& \sigma_{t}=\gamma_{0}+\gamma_{1}\left|v_{t-1}\right|+\cdots+\gamma_{q}\left|v_{t-q}\right| \tag{1.6.2}
\end{align*}
$$

and $\phi(L)$ and $\psi\left(L^{-1}\right)$ are defined following (1.2.1), and $u_{t}$ is a sequence of iid innovations. The linear ARCH model is able to capture the correlation in the variance and meanwhile preserves a relatively simple linear specification compared to other alternatives like GARCH. Under the $\mathbb{H}_{0}$ where $\psi\left(L^{-1}\right)$ degenerates to 1 , the linearity of conditional quantile specification still holds after adding $\left\{\left|v_{t-j}\right|\right\}_{j=1}^{q}$ into the regression equation.

$$
\begin{align*}
Q_{Y_{t}}\left(\tau \mid Y_{t-1}, Y_{t-2}, \ldots\right) & =\phi_{1} Y_{t-1}+\cdots+\phi_{r} Y_{t-r}+Q_{u_{t}}(\tau)\left(\gamma_{0}+\gamma_{1}\left|v_{t-1}\right|+\cdots+\gamma_{q}\left|v_{t-q}\right|\right) \\
& =\underbrace{Q_{u_{t}}(\tau) \gamma_{0}}_{\tilde{\gamma}_{0}(\tau)}+\underbrace{Q_{u_{t}}(\tau) \gamma_{1}}_{\tilde{\gamma}_{1}(\tau)}\left|v_{t-1}\right|+\cdots+\underbrace{Q_{u_{t}}(\tau) \gamma_{q}}_{\tilde{\gamma}_{q}(\tau)}\left|v_{t-q}\right|+\phi_{1} Y_{t-1}+\cdots+\phi_{r} Y_{t-r} \\
& =\tilde{\gamma}_{0}(\tau)+\tilde{\gamma}_{1}(\tau)\left|v_{t-1}\right|+\cdots+\tilde{\gamma}_{q}(\tau)\left|v_{t-q}\right|+\phi_{1} Y_{t-1}+\cdots+\phi_{r} Y_{t-r}, \tag{1.6.3}
\end{align*}
$$

where $\left|v_{t-j}\right|$ can be recovered by $Y_{t-j}-\phi_{1} Y_{t-j-1}-\cdots-\phi_{r} Y_{t-j-r}$. Under noncausality, the explicit expression of the conditional quantile of $Y_{t}$ remains unclear. Nevertheless, it cannot be characterized by encompassing linear combinations of residuals in the model. Consequently, the linear dynamic quantile model would not be the correct specification conceivably.

Overall, these two possible extensions to cases with nonlinear dynamics are tentative since the statistical properties of conditional quantiles of $Y_{t}$ defined by (1.6.1) and (1.6.2) require further investigation, which opens a couple of lines for future
research. Some simulation trials in Appendix 1.8.4 have shown the validity of the proposed strategies.

### 1.6.2. Perspective from Extreme Quantiles

One intriguing observation from the simulation is that QAR estimates at extreme quantiles might be informative for identifying the true models even though linear quantile specification is incorrect for the conditional quantile of non-causal processes. As depicted in Figure 1.4.1, for $\operatorname{MAR}(0,1)$ processes with coefficient 0.6 driven from asymmetric innovations,

$$
\left(1-0.6 L^{-1}\right) Y_{t}=u_{t} \Leftrightarrow Y_{t}=(0.6)^{-1} Y_{t-1}-(0.6)^{-1} u_{t-1} .
$$

The estimated slope of $Y_{t-1}$ approaches $(0.6)^{-1}$ when the quantile gets close to 0 or 1. Somehow it indicates that the linear correlation at extreme quantiles between $Y_{t}$ and $Y_{t-1}$ can help to discriminate causality and non-causality. A similar idea has been adopted for model selection based on the extreme clustering of residuals by Fries and Zakoïan (2019), but the method is restricted to $\alpha$-stable distributions. Rich data is required for further analysis to get a less biased estimator for conditional quantiles close to 0 or 1 . This opens a possible avenue for future research in line with identifying causal and non-causal processes using tail information of processes.

### 1.7. Conclusion

This paper introduces three novel testing strategies for non-causality in linear time series within the quantile regression framework. The tests exploit the non-linearity of autoregressive processes with non-causality and achieve the objective of detecting non-causality based on the well-developed inference under the QAR framework. The constancy test shares the simplicity of implementation but lacks consistency since the behavior of linear quantile autoregression for non-causal processes is not clear yet. This issue from the constancy test is tackled by testing the specification of linear conditional quantile models to detect non-causality. Specification-based noncausality testing procedures like EV and EG tests yield stable size at a nominal level and fairly satisfactory power. On the one hand, the EV test outperforms the EG test in platykurtic and leptokurtic situations or symmetric distributions. On the other hand, the EG test is less computationally cumbersome and shows better performance when the process is skewed. However, no method is placed in a dominating situation. Thus, a combined testing procedure with the constancy test as a preliminary check complemented with either EV or EG test is suggested for practitioners.

Some possible extensions to accommodate different dependence in model innovations, which might bring obstacles in detecting non-causality, are proposed at the end of the paper. Some simulation results in QAR estimates at extreme quantiles indicate the possibility of identifying non-causal processes by employing information from the tails of processes.

### 1.8. Appendices

### 1.8.1. Some Properties of Non-causal Autoregressive Processes

Higher-order Dependence of All-pass Time Series Models Following the same setup in Example 1.2.3,

$$
\tilde{u}_{t}=\frac{1-\psi L}{1-\psi L^{-1}} u_{t}=\sum_{j=-\infty}^{\infty} \rho_{j} u_{t+j} .
$$

The skewness of $\tilde{u}_{t}$ is

$$
\mathbb{E}\left(\tilde{u}_{t}^{3}\right)=\sum_{j=-\infty}^{\infty} \rho_{j}^{3} \mathbb{E}\left(u_{t}^{3}\right)=\left(1-\frac{3 \psi^{2}(\psi+1)}{\psi^{2}+\psi+1}\right) \mathbb{E}\left(u_{t}^{3}\right),|\psi|<1,
$$

where $-1<\left(1-\frac{3 \psi^{2}(\psi+1)}{\psi^{2}+\psi+1}\right)<1$. From the above expression, it is easy to tell that the all-pass filter preserves the symmetry of the innovations if the original ones are not skewed but might alter the direction of skewness if $u_{t}$ is asymmetric by varying values of $\psi$. Apart from the correlation in the squared value of $\tilde{u}_{t}$ that has been explicitly shown in Example 1.2.3, here we derive the closed form solution for the dependence at order 3 . It suffices to show $\operatorname{Cov}\left(\tilde{u}_{t}^{3}, \tilde{u}_{t+h}^{3}\right)$ is nonzero for $h \neq 0$.

$$
\begin{aligned}
& \mathbb{E}\left(\tilde{u}_{t}^{3} \tilde{u}_{t+h}^{3}\right)=\mathbb{E}\left(\sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \rho_{j} \rho_{i} \rho_{m} \rho_{n} \rho_{l} \rho_{k} u_{t+j} u_{t+i} u_{t+m} u_{t+h+n} u_{t+h+k} u_{t+h+l}\right) \\
= & \left(\sum_{j=-\infty}^{\infty} \rho_{j}^{3} \rho_{j+h}^{3}\right)\left(\mathbb{E}\left(u_{t}^{6}\right)-15 \mathbb{E}\left(u_{t}^{4}\right) \mathbb{E}\left(u_{t}^{2}\right)-10 \mathbb{E}^{2}\left(u_{t}^{3}\right)-15 \mathbb{E}^{3}\left(u_{t}^{2}\right)\right) \\
& +3\left(\sum_{j=-\infty}^{\infty}\left(\rho_{j+h} \rho_{j}^{3}+\rho_{j+h}^{3} \rho_{j}\right)\right) \mathbb{E}\left(u_{t}^{4}\right) \mathbb{E}\left(u_{t}^{2}\right) \\
& +\left(\left(\sum_{j=-\infty}^{\infty} \rho_{j}^{3}\right)^{2}+9\left(\sum_{j=-\infty}^{\infty} \rho_{j+h}^{2} \rho_{j}\right)\left(\sum_{j=-\infty}^{\infty} \rho_{j+h} \rho_{j}^{2}\right)\right) \mathbb{E}^{2}\left(u_{t}^{3}\right)
\end{aligned}
$$

after using $\sum_{j=-\infty}^{\infty} \rho_{j} \rho_{j+h}=0$ (coincides with no correlation property of all-pass time series process) and $\sum_{j=-\infty}^{\infty} \rho_{j}^{2}=1$ (variance preserving property),

$$
\operatorname{Cov}\left(\tilde{u}_{t}^{3}, \tilde{u}_{t+h}^{3}\right)=\mathbb{E}\left(\tilde{u}_{t}^{3} \tilde{u}_{t+h}^{3}\right)-\mathbb{E}\left(\tilde{u}_{t}^{3}\right) \mathbb{E}\left(\tilde{u}_{t+h}^{3}\right)
$$

generally is not zero. For instance, $h=1$, after simplification

$$
\operatorname{Cov}\left(\tilde{u}_{t}^{3}, \tilde{u}_{t+1}^{3}\right)=\alpha_{1} \mathbb{E}\left(u_{t}^{6}\right)+\left(\alpha_{2} \mathbb{E}\left(u_{t}^{4}\right)+\alpha_{4} \mathbb{E}^{2}\left(u_{t}^{2}\right)\right) \mathbb{E}\left(u_{t}^{2}\right)+\alpha_{3} \mathbb{E}^{2}\left(u_{t}^{3}\right)
$$

with

$$
\left\{\begin{array}{l}
\alpha_{1}=\frac{-3 \psi^{5}\left(1-\psi^{2}\right)^{3}}{\psi^{4}+\psi^{2}+1} \\
\alpha_{2}=-3 \psi^{3}\left(1-\psi^{2}\right)^{3}+\frac{45 \psi^{5}\left(1-\psi^{2}\right)^{3}}{\psi^{4}+\psi^{2}+1} \\
\alpha_{3}=\frac{30 \psi^{5}\left(1-\psi^{2}\right)^{3}}{\psi^{4}+\psi^{2}+1}-\frac{9\left(1-\psi^{2}\right)^{3}(2 \psi+1) \psi^{2}}{\left(\psi^{2}+\psi+1\right)^{2}} \\
\alpha_{4}=\frac{45 \psi^{5}\left(1-\psi^{2}\right)^{3}}{\psi^{4}+\psi^{2}+1}
\end{array}\right.
$$

where zeros are not attained at the same value of $\psi \in(-1,1) \backslash\{0\}$.

Conditional Density Function of Non-causal Autoregressions In this section, we try to study the properties of the conditional density function of the response variable in the presence of non-causality in the autoregressive process through simulations. Consider a pair of $\operatorname{MAR}(1,0)$ and $\operatorname{MAR}(0,1)$ processes generated from iid innovations following the same distribution with the density function $f_{u}(\cdot)$,

$$
\left\{\begin{array}{l}
Y_{t}=0.6 Y_{t-1}+u_{t}  \tag{1.8.1}\\
\tilde{Y}_{t}=0.6^{-1} \tilde{Y}_{t-1}+u_{t}=0.6 \tilde{Y}_{t+1}-0.6 u_{t+1}
\end{array}\right.
$$

One is purely causal, and the other is purely non-causal with a coefficient of 0.6. We start by analyzing $f\left(Y_{t} \leq y \mid Y_{t-1}=x\right)$ in the causal case.

$$
\begin{align*}
f\left(Y_{t}=y \mid Y_{t-1}=x\right) & =\frac{d P\left(Y_{t} \leq y \mid Y_{t-1}=x\right)}{d y} \\
& =\frac{d P\left(0.6 Y_{t-1}+u_{t} \leq y \mid Y_{t-1}=x\right)}{d y}  \tag{1.8.2}\\
& =\frac{d P\left(u_{t} \leq y-0.6 x \mid Y_{t-1}=x\right)}{d y} \\
& =f_{u}(y-0.6 x)=f_{u}\left(y-0.6 Y_{t-1}\right)
\end{align*}
$$

It is not difficult to conclude from equation 1.8.2 that the conditional density of $Y_{t}$ given $Y_{t-1}$ is shifting horizontally as the value of $Y_{t-1}$ varies. Despite the change in the location of the density function, the rest remains the same across different values of $Y_{t-1}$.

Similarly, we derive the conditional density function for the non-causal case,

$$
\begin{align*}
f\left(\tilde{Y}_{t}=y \mid \tilde{Y}_{t-1}=x\right) & =\frac{f\left(\tilde{Y}_{t-1}=x \mid \tilde{Y}_{t}=y\right) f\left(\tilde{Y}_{t}=y\right)}{f\left(\tilde{Y}_{t-1}=x\right)} \text { by Bayes rule } \\
& =\frac{f\left(0.6 \tilde{Y}_{t}-0.6 u_{t}=x \mid \tilde{Y}_{t}=y\right) f\left(\tilde{Y}_{t}=y\right)}{f\left(Y_{t-1}=x\right)} \text { by definition of } \tilde{Y}_{t-1} \\
& =\frac{f_{u}\left(y-0.6^{-1} x\right) f\left(\tilde{Y}_{t}=y\right)}{f\left(Y_{t-1}=x\right)} \text { by the independence of } u_{t} \text { and } \tilde{Y}_{t} \\
& =\frac{f_{u}\left(y-0.6^{-1} x\right) f\left(\tilde{Y}_{t}=y\right)}{\int_{-\infty}^{\infty} f\left(\tilde{Y}_{t-1}=x \mid \tilde{Y}_{t}=s\right) f\left(\tilde{Y}_{t}=s\right) d s} \text { law of total probability } \\
& =\frac{f_{u}\left(y-0.6^{-1} x\right) f\left(\tilde{Y}_{t}=y\right)}{\int_{-\infty}^{\infty} f_{u}\left(s-0.6^{-1} x\right) f\left(\tilde{Y}_{t}=s\right) d s} \tag{1.8.3}
\end{align*}
$$

There is no general closed-form solution to this expression. Note that no $x$ plays a role in $f\left(Y_{t}=y\right)$, and the denominator is $x$-dependent but highly nonlinear due to the integration. If we assign additive property ${ }^{13}$ to the marginal distribution of $Y_{t}$. This nonlinearity can be shown more clearly. Say $u_{t}$ follows an exponential distribution with rate $\lambda$, then the equation (1.8.3) has the explicit form

$$
\begin{aligned}
& f\left(Y_{t}=y \mid Y_{t-1}=x\right) \\
= & \frac{\left(\int_{-\infty}^{\infty} e^{-i s y} \prod_{j=0}^{\infty} \frac{\lambda}{\lambda-i s(0.6)^{j}} d s\right) \lambda e^{-(x-0.6 y)}}{\left(\int_{-\infty}^{\infty} e^{-i s x} \prod_{j=0}^{\infty} \frac{\lambda}{\lambda-i s(0.6)^{j}} d s\right)} \mathbb{I}(x-0.6 y \geq 0) .
\end{aligned}
$$

This suggests that the shape (functional form) of the density function would differ, corresponding to the choice of $x$. The following simulation experiment demonstrates this argument. In this simulation, we generate two $\operatorname{AR}(1)$ processes (1.8.1) by exponentially distributed innovations with rate 1 . The sample size is 500. Estimated conditional density functions $f\left(y \mid Y_{t-1}=x\right)$ are plotted in Figure 1.8.1, given five choices of $x: 10 \%, 30 \%, 50 \%, 70 \%$, and $90 \%$ percentiles of $Y_{t-1}$ sample. The density function is estimated by $a k j$ command in R Studio, which is a univariate adaptive kernel estimation used by Portnoy and Koenker (1989). The left panel displays the result for the causal case. As shown in (1.8.2), density functions in different colors (values of $x$ ) share the same shape but the location. Whereas in the right panel, where the estimated density is plotted, five estimated conditional density functions present distinct modes, skewness, and kurtosis.

[^12]Figure 1.8.1: conditional density of $Y_{t}$ given different $x$


### 1.8.2. Asymptotic Properties of QAR Estimates

QAR is first proposed by Koenker and Xiao (2006) to study the conditional quantile functions of the following $p$ th-order AR process,

$$
\begin{equation*}
Y_{t}=\theta_{0}\left(U_{t}\right)+\theta_{1}\left(U_{t}\right) Y_{t-1}+\cdots+\theta_{p}\left(U_{t}\right) Y_{t-p} \tag{1.8.4}
\end{equation*}
$$

where $U_{t}$ is an iid sequence distributed as a standard uniform. This expression can be regarded as an $\operatorname{AR}(p)$ process allowing coefficients of lags to be random but somewhat dependent on each other. By the property of monotone transformation, our target, the conditional quantile at each $\tau \in(0,1)$, can be written as

$$
\begin{equation*}
Q_{Y_{t}}\left(\tau \mid Y_{t-1}, \ldots, Y_{t-p}\right)=\theta_{0}(\tau)+\theta_{1}(\tau) Y_{t-1}+\cdots+\theta_{p}(\tau) Y_{t-p}, \quad \tau \in(0,1) \tag{1.8.5}
\end{equation*}
$$

The estimates of $\boldsymbol{\theta}(\tau)=\left(\theta_{0}(\tau), \theta_{1}(\tau), \ldots, \theta_{p}(\tau)\right)^{\prime}$ are obtained by minimizing the following objective function,

$$
\begin{equation*}
\hat{\boldsymbol{\theta}}(\tau)=\underset{\boldsymbol{\theta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{t=1}^{T} \rho_{\tau}\left(Y_{t}-\boldsymbol{X}_{t}^{\prime} \boldsymbol{\theta}\right), \tag{1.8.6}
\end{equation*}
$$

where $\boldsymbol{X}_{t}^{\prime}=\left(1, Y_{t-1}, \ldots, Y_{t-p}\right)$ and check function $\rho_{\tau}(u)=u(\tau-\mathbb{I}(u<0))$. A vectorized form of (1.8.4) is introduced to facilitate the asymptotic analysis of the estimates $\hat{\boldsymbol{\theta}}(\tau)$,

$$
\boldsymbol{Y}_{t}=\boldsymbol{A}_{t} \boldsymbol{Y}_{t-1}+\boldsymbol{V}_{t}
$$

with

$$
\boldsymbol{A}_{t}=\left(\begin{array}{cccc}
\theta_{1}\left(U_{t}\right) & \theta_{2}\left(U_{t}\right) & \ldots & \theta_{p}\left(U_{t}\right) \\
& \boldsymbol{I}_{p-1} & & \mathbf{0}_{(p-1) \times 1}
\end{array}\right) \text { and } \boldsymbol{V}_{t}=\binom{\epsilon_{t}}{\mathbf{0}_{(p-1) \times 1}}
$$

where $\epsilon_{t}=\theta_{0}\left(U_{t}\right)-\mathbb{E}\left(\theta_{0}\left(U_{t}\right)\right)$ and $\boldsymbol{Y}_{t}=\left(Y_{t}, Y_{t-1}, \ldots, Y_{t-p+1}\right)^{\prime}$. The study of asymptotic properties is based on the following conditions

1. $\left\{\epsilon_{t}\right\}$ are iid innovations with mean 0 and finite variance $\sigma^{2}<\infty$. The distribution function of $\epsilon_{t}, \mathrm{~F}$, admits a continuous density $f(\epsilon)$ away from zero on $E=\{\epsilon: 0<F(\epsilon)<1\}$.
2. The eigenvalues of $\mathbb{E}\left(\boldsymbol{A}_{t} \otimes \boldsymbol{A}_{t}\right)$ have moduli within unity.
3. The conditional distribution function $P\left(Y_{t}<\cdot \mid Y_{t-1}, Y_{t-2}, \ldots\right)$ denoted by $F_{t-1}(\cdot)$ has a density function $f_{t-1}(\cdot)$ uniformly integrable on $E$.

Under these three assumptions,

$$
\Sigma^{-1 / 2} \sqrt{T}(\hat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}(\tau)) \rightarrow_{d} \boldsymbol{B}_{p+1}(\tau)
$$

where $\boldsymbol{B}_{k}(\tau)$ is a $k$-dimensional Brownian bridge. By definition it can be written as $\mathcal{N}\left(\mathbf{0}, \tau(1-\tau) \boldsymbol{I}_{k}\right)$ for any given $\tau$. $\Sigma$ is a matrix characterized by density and distribution function of $\epsilon_{t}$. Let $\Sigma_{0}=\mathbb{E}\left(\boldsymbol{X}_{t} \boldsymbol{X}_{t}^{\prime}\right)$ and $\Sigma_{1}=\mathbb{E}\left(f_{t-1}\left(F_{t-1}^{-1}(\tau)\right) \boldsymbol{X}_{t} \boldsymbol{X}_{t}^{\prime}\right)$. Then $\Sigma$ is defined as $\Sigma_{1}^{-1} \Sigma_{0} \Sigma_{1}^{-1}$.

In the special case where the data generating process is a conventional causal AR model with fixed coefficients, the conditional density would be independent of $\boldsymbol{X}_{t}$. We will have $\Sigma_{1}=f\left(F^{-1}(\tau)\right) \mathbb{E}\left(\boldsymbol{X}_{t} \boldsymbol{X}_{t}^{\prime}\right)$. Further we can simplify the $\Sigma$ to $f^{-2}\left(F^{-1}(\tau)\right) \mathbb{E}^{-1}\left(\boldsymbol{X}_{t} \boldsymbol{X}_{t}^{\prime}\right)$.

### 1.8.3. Proof to Theorem 1.3.1

Non-causality $\Rightarrow$ Nonlinear conditional quantile We prove this statement by contradiction. Consider a stationary $\operatorname{MAR}(r, s)$ with innovations satisfying Assumption 1.1 and $s>0$. Assume all conditional quantiles of $Y_{t}$ are linear in $\left(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}\right)$, where $p=r+s$. That is,
$Q_{Y_{t}}\left(\tau \mid Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}\right)=\theta_{0}(\tau)+\theta_{1}(\tau) Y_{t-1}+\cdots+\theta_{p}(\tau) Y_{t-p}$ for any given $\tau \in(0,1)$.
By aggregating $Q_{Y_{t}}\left(\tau \mid Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}\right)$ over the entire quantile range, the linearity is maintained for the aggregation. Therefore, we have
$\int_{0}^{1} Q_{Y_{t}}\left(\tau \mid Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}\right) d \tau=\int_{0}^{1} \theta_{0}(\tau) d \tau+\int_{0}^{1} \theta_{1}(\tau) d \tau Y_{t-1}+\cdots+\int_{0}^{1} \theta_{p}(\tau) d \tau Y_{t-p}$
Equivalently, we can yield

$$
\mathbb{E}\left(Y_{t} \mid Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}\right)=\theta_{0}+\theta_{1} Y_{t-1}+\cdots+\theta_{p} Y_{t-p}
$$

which contradicts the statement in Corollary 5.2 .3 by Rosenblatt (2000) on the nonlinearity of conditional expectation in past information with the presence of
non-causality. Thus, the presumed statement is not valid. That is to say, there exists a conditional quantile of $Y_{t}$ which is nonlinear in the past information for at least one $\tau \in(0,1)$ if $s>0$.

Nonlinear conditional quantile $\Rightarrow$ Non-causality This can be demonstrated equivalently by its contrapositive statement: an AR process $Y_{t}$ being causal implies its conditional quantile $Q_{Y_{t}}\left(\tau \mid I_{t-1}\right)$ is linear for all $\tau \in(0,1)$.

If a $\operatorname{MAR}(r, s)$ is purely causal, i.e., $\mathrm{s}=0$ and $\mathrm{r}=\mathrm{p}$. Then we have

$$
Y_{t}=\phi_{1} Y_{t-1}+\phi_{2} Y_{t-2}+\cdots+\phi_{r} Y_{t-r}+u_{t},
$$

where $u_{t}$ is independent of past observations. The conditional quantile of the response variable can be directly expressed as a linear combination of $\left\{Y_{t-j}\right\}_{j=1, \ldots, r}$,

$$
\begin{aligned}
Q_{Y_{t}}\left(\tau \mid Y_{t-1}, \ldots, Y_{t-r}\right) & =Q_{u_{t}}\left(\tau \mid Y_{t-1}, \ldots, Y_{t-r}\right)+\phi_{1} Y_{t-1}+\phi_{2} Y_{t-2}+\cdots+\phi_{r} Y_{t-r} \text { for } \forall \tau \in(0,1) \\
& =\underbrace{Q_{u_{t}}(\tau)}_{\theta_{0}(\tau)}+\underbrace{\phi_{1}}_{\theta_{1}(\tau)} Y_{t-1}+\underbrace{\phi_{2}}_{\theta_{2}(\tau)} Y_{t-2}+\cdots+\underbrace{\phi_{r}}_{\theta_{r}(\tau)} Y_{t-r} .
\end{aligned}
$$

### 1.8.4. Simulations of Extensions

In this section, we present some simulation results of non-causality testing strategies extended to the autoregressive processes with heteroskedasticity. In this stage, we assume the form of heteroskedasticity is known.

Consider a pair of $\operatorname{MAR}(1,0)-\operatorname{ARCH}(1)$ and $\operatorname{MAR}(0,1)-\mathrm{ARCH}(1)$ processes defined by

$$
\left\{\begin{array}{l}
Y_{t}=0.7 Y_{t-1}+v_{t}  \tag{1.8.8}\\
Y_{t}^{*}=0.7^{-1} Y_{t-1}^{*}+v_{t}=0.7 Y_{t+1}^{*}-0.7 v_{t+1} \\
v_{t}=\sigma_{t} u_{t} \\
\sigma_{t}=0.2+0.8\left|v_{t-1}\right| \text { where } u_{t} \sim \operatorname{IID}(0,1)
\end{array}\right.
$$

As explained in the extension section, we want to detect non-causality by checking whether the coefficients (except the intercept) in the following linear dynamic quantile model are $\tau$-invariant

$$
\begin{equation*}
Q_{Y_{t}}\left(\tau \mid Y_{t-1}, v_{t-1}\right)=\theta_{0}(\tau)+\theta_{1}(\tau) Y_{t-1}+\theta_{2}(\tau)\left|v_{t-1}\right|, \quad \tau \in \Upsilon \subset(0,1) \tag{1.8.9}
\end{equation*}
$$

provided that $v_{t-1}$ is recovered with $100 \%$ accuracy.
Another approach is to check whether the linear model (1.8.8) is the correct specification for the conditional quantile of $Y_{t}$ and $Y_{t}^{*}$. The innovation $u_{t}$ varies from exponential to $t$ student and Laplace distributions. The sample size in this trial is

100, 200, 500, and 1000. The trimmed quantile interval for the constancy test is [0.05, 0.95] and $\Upsilon=[0.01,0.99]$ for the specification-based test (here in this experiment, we only apply the EV test to see its performance). The empirical size and power of non-causality tests for cases with heteroskedasticity are displayed in Table 1.8.1. It is clear that both methods have fairly good performance, even in relatively small samples. Regarding the empirical size, both approaches have a slight distortion compared to the nominal level. In the case of the EV test, a different bandwidth can be applied to adjust the empirical size to the expected level. Concerning the empirical power, the specification-based approach dominates the constancy test in most cases. Except in the case of asymmetric distribution, the constancy test outperforms the EV test in small samples ( $\mathrm{T}=100$ and 200). It is conceivable that when we replace $v_{t}$ by its estimate $\hat{v}_{t}$, the asymptotic effect of the estimation needs to be taken into account when we construct test statistics. This is beyond the scope of this paper.

Table 1.8.1: Empirical size and power of non-causality tests for AR-ARCH models with known heteroskedasticity

| Distribution$u_{t}$ | test type | $\mathrm{T}=100$ |  | $\mathrm{T}=200$ |  | $\mathrm{T}=500$ |  | $\mathrm{T}=1000$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | size | power | size | power | size | power | size | power |
| Exponential | constancy test | 3.60\% | $39.00 \%$ | 4.40\% | 54.60\% | 6.00\% | 75.80\% | 7.00\% | 90.80\% |
|  | EV test | 5.20\% | 16.80\% | 4.40\% | 34.80\% | 4.60\% | 82.80\% | 5.00\% | 97.80\% |
| t student | constancy test | 5.20\% | 24.60\% | 6.60\% | 39.40\% | 6.00\% | 63.60\% | 7.80\% | 78.00\% |
|  | EV test | 8.20\% | 51.00\% | 7.20\% | 83.20\% | 8.20\% | 99.80\% | 7.20\% | 100.00\% |
| Laplace | constancy test | 4.40\% | 14.40\% | 3.80\% | 25.00\% | 7.60\% | 38.80\% | 4.40\% | 55.00\% |
|  | EV test | 7.00\% | 45.40\% | 6.40\% | 82.60\% | 8.60\% | 99.20\% | 7.20\% | 10.00\% |

The bandwidth for approximating the critical value using subsampling for the EV test is $b=\left\lfloor 7 T^{2 / 5}\right\rfloor$.

# 2. CHAPTER II: ESTIMATION OF TIME SERIES MODELS USING THE EMPIRICAL DISTRIBUTION OF RESIDUALS 

### 2.1. Introduction

Time series models with nonfundamental solutions have drawn considerable attention in the econometrics literature during the last two decades. They are represented by noncausal and noninvertible time series models. In Macroeconomics, nonfundamentalness arising from the moving average part, namely, noninvertibility, has been interpreted as economic agents' information sets being larger than econometricians' limited observations (Hansen and Sargent, 1991, p. 77-116). Lippi and Reichlin $(1993,1994)$ pointed out the importance of exploration of noninvertible moving average representations for the analysis of impulse-response functions with empirical applications in GNP-unemployment and interest rate-inflation. Leeper et al. (2013) explained noninvertibility as a recurrent consequence of an agent's foresight from the perspective of the information flows with an analytical case in tax news. More empirical examples of noninvertible processes applied to modeling forward-looking behavior can be found in Alessi et al. (2011). Noncausal processes have been widely exploited in Engineering from the perspective of random stochastic systems; as discussed by Tekalp et al. (1986) and Gaeta et al. (1997). The interest in noncausal processes in Economics and Finance stems from their ability to replicate nonlinear dynamics like locally explosive behavior and asymmetric cycles in time series. Noncausal autoregressive (AR) processes with heavy-tailed innovations, for example, can simulate the trajectory of phases of repetitive upward trends followed by an instantaneous drop, which is opposite to the pattern followed by a causal process, see Figure 2.1.1, where simulated $\operatorname{AR}(1)$ processes with roots equal to 0.9 and $(0.9)^{-1}$ are depicted. This feature is useful in modeling speculative bubbles in stock markets, as demonstrated by Gouriéroux and Zakoïan (2017) and Hecq and Voisin (2021). Noncausal processes can also display clustering volatility like autoregressive conditional heteroskedasticity (ARCH) behavior, which is commonly observed in financial data (Breidt et al. (2001)). Lof and Nyberg (2017) incorporate noncausality into linear AR models to improve the forecast accuracy of commodity prices. Hecq et al. (2020) propose a mixed causal-noncausal AR model that includes exogenous regressors and demonstrate the benefits of such models in ex-post forecasting. Noncausal AR processes can also be used to explain forward-looking behavior, as an alternative to noninvertible moving-average processes (MA), as shown by Lanne and Luoto (2013).

However, conventional time series analysis is generally restricted to causal and invertible representations without valid arguments for excluding nonfundamental solutions. The standard estimation techniques based on second-order moments, like OLS, fail to distinguish causal (invertible) from noncausal (noninvertible) processes since all weakly stationary processes admit a casual and invertible representation with identical autocorrelation structures. The information in the variance-covariance matrix of residuals is not sufficient to characterize the serial independence of nonGaussian innovations under the iid assumption. There are linear transformations on iid data that generate white noise sequences with the same second-order moment structure but not serially independent, like an all-pass filter ${ }^{14}$. The logic behind pseudo-Gaussian maximum likelihood (ML) estimation is that second-order moments suffice to identify the Gaussian probabilistic structure but are not adequate for other non-Gaussian models whose dynamic properties can not be identified using only autocorrelations. Hence Gaussian ML estimation does not apply to noncausal and noninvertible processes driven by non-Gaussian innovations, and alternative estimation techniques of general time series models resorting to non-Gaussianity are required.

Breid et al. (1991) and Lii and Rosenblatt (1992) introduce approximate ML procedures for noncausal/noninvertible processes given complete knowledge of the distribution of innovations. ML estimation achieves efficiency but imposes restrictive assumptions since, in most empirical cases, the distribution of innovations is not known. To circumvent this stringent distributional condition, Huang and Pawitan (2000) and Lanne and Saikkonen (2011) adopt the Laplacian and t-student density functions to approximate the likelihood function, respectively, which yield consistent estimates of model parameters for ARMA models without imposing causality and invertibility. In a similar fashion, Fries and Zakoïan (2019) derive the loss function from $\alpha$-stable distributions for mixed causal-noncausal AR processes which can accommodate heavy tails. In parallel, some progress has been made in the estimation techniques of all-pass models using the non-Gaussian likelihood functions or rankbased residuals dispersion function, see Breidt et al. (2001) and Andrews et al. (2006, 2007). These methods can be employed to estimate general noncausal and noninvertible time series models through a two-step procedure: first, obtain the residuals from a causal and invertible ARMA model applied to the original data by Gaussian ML; second, fit the residuals by a purely noncausal/noninvertible all-pass model. In this approach, the validity of the second step depends on whether the residuals from the first step are white noise but dependent sequences. In addition, the asymptotic analysis of the estimates remains open since the estimation is carried out on the residuals rather than the raw data directly. In contrast to the approaches resorting

[^13]to density functions, Gospodinov and Ng (2015) and Velasco and Lobato (2018) develop estimation methods based on the higher-order cumulants of non-Gaussian innovations in the time domain and frequency domain, respectively. Still, these two methodologies require at least a finite sixth-order moment of innovations to achieve the global identification of ARMA model parameters. To relax the assumptions on higher-order moments, Velasco (2022) propose to estimate general time series models using a serial dependence measure of residuals through the characteristic functions under iid/mds assumptions.

In this paper, we employ a general dependence measure based on the distance between the joint distribution function and the product of marginal distribution functions of random variables. This intuitive measure was initially proposed by Hoeffding (1948) and has been extended to $m$-dimensional random vectors by Blum et al. (1961). Later on, Skaug and Tjøstheim (1993) and Delgado (1996) consider tests of first-order and $p$ serial dependence in the time series context using this measure. Hong (1998) proposes a consistent test against all pairwise dependence via the empirical distribution function by taking all lags into account. Throughout this paper, following Hong (2000), we adopt a generalized spectral distribution function based on the Fourier transform of the aforementioned measure to capture the serial dependence of a given random sequence at all lags. In the estimation method, we consider a minimum distance estimate based on the loss function equal to the $L_{2}$ distance between the proposed dependence measure in the unrestricted case and the restricted one applied to the empirical cumulative distribution function of residuals. Our one-step estimation technique has some appealing attributes compared to other alternatives. First, it achieves the identification of the model parameters without imposing causality and invertibility. Second, only some regular smoothness conditions on the distribution of the innovations are required instead of stringent conditions on moments or parametric distributional knowledge. Unlike other procedures using spectral densities, it does not involve any subjective choices of lag windows or trimming parameters. Moreover, compared to the approach based on characteristic functions, the cumulative distribution function is more robust to outliers and less computationally cumbersome without complex quantities. Owing to the flexibility of the cumulative distribution function, the method can be tentatively extended to time series models with different dependence structures for innovations, for example, quantile independence or conditional mean independence.

The rest of the paper is organized as follows. Section 2.2 introduces the measure of pairwise dependence based on the cumulative distribution function. Section 2.3 investigates the identification of the model, consistency, and asymptotic properties of the proposed estimator under the serial independence condition on innovations. Section 2.4 presents results from some Monte Carlo experiments and discusses the finite sample performance of estimates. Section 2.5 illustrates the merits of this

Figure 2.1.1: Simulated processes from causal and noncausal $\operatorname{AR}(1)$ models

method through an empirical application. Finally, Section 2.6 concludes and discusses some possible extensions of our estimates in the procedure.

### 2.2. Pairwise dependence measures based on residuals

Consider a time series generated by

$$
\begin{equation*}
Y_{t}=\sum_{j=-\infty}^{\infty} \varphi_{j} u_{t-j}, \tag{2.2.1}
\end{equation*}
$$

where $\left\{u_{t}\right\}_{t \in \mathbb{Z}}$ is a sequence of independent identically distributed (iid) innovations with zero mean. Double-sided summation in the representation of infinite moving average (2.2.1) allows the process to be noncausal or/and noninvertible. The stationarity of $Y_{t}$ is guaranteed under the conditions like $\varphi_{j}$ being absolutely summable and $\mathbb{E}\left|u_{t}\right|<\infty$. The operator $\varphi(\theta, L)=\sum_{j=-\infty}^{\infty} \varphi_{j}(\theta) L^{j}$ with coefficients $\varphi_{j}(\theta)$ and lag operator $L$ defines the generation of $Y_{t}$ in terms of parameter $\theta \in \Theta \subset \mathbb{R}^{d}$ so that $\varphi_{j}\left(\theta_{0}\right)=\varphi_{j}$ for all $j$.

A common example is the autoregressive moving average process of order $(p, q)$, abbreviated as $\operatorname{ARMA}(p, q)$,

$$
\begin{equation*}
\alpha(L) Y_{t}=\beta(L) u_{t} \tag{2.2.2}
\end{equation*}
$$

where $\alpha(L)=1-\sum_{j=1}^{p} \alpha_{j} L^{j}$ is an autoregressive polynomial of order $p$ and $\beta(L)=$ $1+\sum_{j=1}^{q} \beta_{j} L^{j}$ is a moving average polynomial of order $q$. We allow the roots of both polynomials to lie inside and outside the unit circle, while $\alpha(z)$ and $\beta(z)$ have no common zeroes. The parameters of interest in (2.2.2) are $\theta=\left(\alpha_{1}, \ldots, \alpha_{p}, \beta_{1}, \ldots, \beta_{q}\right)^{\prime} \subset$ $\left\{\theta \in \mathbb{R}^{p+q}: \alpha(z) \beta(z) \neq 0\right.$ for all $z \in \mathbb{C}$ such that $\left.|z|=1, \alpha_{p} \neq 0, \beta_{q} \neq 0\right\}$. The
restriction defined on the parameter space guarantees the existence of the Laurent expansion of $\alpha^{-1}(L) \beta(L)$, from which the coefficients $\varphi_{j}$ in the infinite moving average representation are determined.

The residuals evaluated at any value $\theta$ under the correct specification are computed by

$$
\begin{equation*}
u_{t}(\theta)=\varphi^{-1}(\theta, L) Y_{t}=\varphi^{-1}(\theta, L) \varphi\left(\theta_{0}, L\right) u_{t}=\phi(\theta ; L) u_{t} \tag{2.2.3}
\end{equation*}
$$

where $\varphi(\theta, L)=\sum_{j=-\infty}^{\infty} \varphi_{j}(\theta) L^{j}, \varphi^{-1}(\theta, L)=\sum_{j=-\infty}^{\infty} \varphi_{j}^{(-1)}(\theta) L^{j}$ and the linear filter on the true innovations $\phi(\theta ; L)=\varphi^{-1}(\theta, L) \varphi\left(\theta_{0}, L\right)$. Evaluation of $\phi(\theta ; L)$ at the true value of the parameter allows to retrieve the sequence of innovations, i.e., $u_{t}\left(\theta_{0}\right)=u_{t}$ when $\theta=\theta_{0}$.

Before proceeding to the estimation method based on the dependence measure of residuals, we introduce some notation. Let $\Theta$ be a compact set containing the true parameter $\theta_{0} . I(A)$ is the indicator function of the event $A$, and $\mathbb{P}(A)$ is the probability measure of the event $A . C$ is a generic positive constant that may vary in different situations.

To capture the general serial dependence of the residual sequence $\left\{u_{t}(\theta)\right\}$ without imposing moment conditions at higher orders, we consider the distance between the joint cumulative distribution function of any pair of residuals $\left(u_{t}(\theta), u_{t-j}(\theta)\right)$ and the product of marginal distribution functions at any given $(x, y) \in \mathbb{R}^{2}$,

$$
\begin{align*}
\sigma_{\theta, j}^{*}(x, y): & =\mathbb{P}\left(u_{t}(\theta) \leq x, u_{t-j}(\theta) \leq y\right)-\mathbb{P}\left(u_{t}(\theta) \leq x\right) \mathbb{P}\left(u_{t-j}(\theta) \leq y\right) \\
& =\mathbb{E}\left(I\left(u_{t}(\theta) \leq x, u_{t-j}(\theta) \leq y\right)\right)-\mathbb{E}\left(I\left(u_{t}(\theta) \leq x\right)\right) \mathbb{E}\left(I\left(u_{t-j}(\theta) \leq y\right)\right) \\
& =\mathbb{C o v}\left(I\left(u_{t}(\theta) \leq x\right), I\left(u_{t-j}(\theta) \leq y\right)\right), \quad j=0, \pm 1, \ldots, \tag{2.2.4}
\end{align*}
$$

which can also be interpreted as a generalization of the standard covariance between $u_{t}(\theta)$ and $u_{t-j}(\theta)$ by applying an indicator transformation on the given pair of random variables. It is worth noting that $\sigma_{\theta,-j}^{*}(x, y)=\sigma_{\theta, j}^{*}(y, x)$ for all $j$, so that, without losing generality, we can define our measure of generic pairwise dependence by

$$
\begin{equation*}
\sigma_{\theta, j}(x, y)=\sigma_{\theta,|j|}^{*}(x, y) \text { for } j=0, \pm 1, \pm 2, \ldots \quad \forall(x, y) \in \mathbb{R}^{2} \tag{2.2.5}
\end{equation*}
$$

This measure allows us to capture all pairwise dependence between $u_{t}(\theta)$ and $u_{t-j}(\theta)$ by varying $x$ and $y$ over the entire real coordinate space, without specifying any functional forms or the order of moments in the "type of dependence". If $\sigma_{\theta, j}(x, y)=0$ for all $(x, y) \in \mathbb{R}^{2}$, then $u_{t}(\theta)$ and $u_{t-j}(\theta)$ are independent. A closely related concept can be constructed by replacing the residual sequences with the corresponding cumulative distribution functions of residuals, namely, copula covariance,

$$
\sigma_{\theta, j}^{c}\left(x_{1}, x_{2}\right)=\operatorname{Cov}\left(I\left\{F\left(u_{t}(\theta)\right) \leq x_{1}\right\}, I\left\{F\left(u_{t-j}(\theta)\right) \leq x_{2}\right\}\right)
$$

where $F\left(u_{t}(\theta)\right)$ is the $c d f$ of $u_{t}(\theta)$ and $\left(x_{1}, x_{2}\right) \in[0,1]^{2}$, see Lee and Rao (2011). This measure also exploits the same distributional information in any pair of residuals as $\sigma_{\theta, j}\left(x_{1}, x_{2}\right)$. In addition, the copula covariance has an appealing property of being invariant to monotonic transformations of $u_{t}(\theta)$, compared to the general dependence measure based on the characteristic function (Velasco (2022)) or our proposed one. However, the non-trivial effect of estimating $F\left(u_{t}(\theta)\right)$ excessively complicates the asymptotic analysis in the model estimation, which is out of the scope of this paper. More commonly, the copula covariance has been employed to characterize nonlinear sequential dependence in the levels of economic variables that cannot be fully captured by correlations of higher-order moments in Hagemann (2011), Kley et al. (2016) and Baruník and Kley (2019).

If the dependence decays fast enough as $j$ increases in the sense that,

$$
\sup _{(x, y) \in \mathbb{R}^{2}} \sum_{j=-\infty}^{\infty}\left|\sigma_{\theta, j}(x, y)\right|<\infty
$$

we can define the generalized spectral density based on the measure (2.2.5) at any frequency $\omega$ in $[-\pi, \pi]$ with $i=\sqrt{-1}$, by

$$
h_{\theta}(x, y ; \omega):=\frac{1}{2 \pi} \sum_{j=-\infty}^{\infty} \sigma_{\theta, j}(x, y) e^{-i j \omega}, \quad \omega \in[-\pi, \pi],
$$

and the associated generalized spectral distribution function by

$$
\begin{equation*}
H_{\theta}(x, y ; \lambda)=2 \int_{0}^{\lambda \pi} h_{\theta}(x, y ; \omega) d \omega=\sigma_{\theta, 0}(x, y) \lambda+2 \sum_{j=1}^{\infty} \sigma_{\theta, j}(x, y) \frac{\sin j \pi \lambda}{j \pi} \quad \lambda \in[0,1] \tag{2.2.6}
\end{equation*}
$$

The spectral approach facilitates the description of all pairwise dependence in a random sequence by incorporating the general pairwise dependence at all lags into one statistic. The same approach has been considered by Hong (2000) for testing the hypothesis of serial independence against all possible pairwise dependence alternatives and by Du and Escanciano (2015) for a distribution-free test for serial independence of residuals.

Under the independent assumption on $u_{t}(\theta)$,

$$
\begin{equation*}
h_{\theta}(x, y ; \omega)=h_{\theta}^{i i d}(x, y ; \omega):=\frac{1}{2 \pi} \sigma_{\theta, 0}(x, y) \quad \forall(x, y) \in \mathbb{R}^{2} \tag{2.2.7}
\end{equation*}
$$

at any frequency $\omega$ since $\sigma_{\theta, j}(x, y)=0$ for all $j \neq 0$ and any given pair $(x, y) \in \mathbb{R}^{2}$. Moreover, the associated generalized spectral distribution function, under the state specification, becomes

$$
\begin{equation*}
H_{\theta}(x, y ; \lambda)=H_{\theta}^{i i d}(x, y ; \lambda):=\sigma_{\theta, 0}(x, y) \lambda, \quad \lambda \in[0,1] \tag{2.2.8}
\end{equation*}
$$

All pairwise dependence degenerates to zero except for the self-dependence for $u_{t}(\theta)$. The following assumption is introduced to formalize the investigation of the dependence structure of the residual sequence $u_{t}(\theta)$.

Assumption 2.1. For compact $\Theta$ and $\mu_{0}>1$,

$$
\sup _{\theta \in \Theta}\left|\varphi_{j}(\theta)\right|+\sup _{\theta \in \Theta}\left|\varphi_{j}^{(-1)}(\theta)\right| \leq C|j|^{-\mu_{0}}, \quad j= \pm 1, \pm 2, \ldots
$$

In Assumption 2.1, we impose a uniform upper bound on the coefficients at each order $j$ of the polynomial $\varphi(\theta, L)$ and its inverse $\varphi^{-1}(\theta, L)$ by a sequence of constants converging to zero when $j$ goes to infinity. The condition contributes to the absolute summability of the coefficients in the linear filter $\phi(\theta ; L)$ defined in (2.2.3), ensuring the stationarity of residual sequences indexed by $\theta \in \Theta$ together with $\mathbb{E}\left|u_{t}\right|<\infty$. This allows us to analyze some time-invariant statistical properties of $u_{t}(\theta)$. The classical mixing conditions have been investigated for one-sided moving average sequences. Some classes of two-sided moving averages that do not fulfill the mixing conditions are listed in Sidorov (2010). Nevertheless, with the following smoothness condition on the innovations, we can show that $u_{t}(\theta)$ and $u_{t-j}(\theta)$ behave more similarly to pairwise independent variables the further they depart from each other, which characterizes the weak dependence of residual sequences.

Assumption 2.2. The innovation $u_{t}$ admits a density $f(u)$ with first-order derivative $f^{(1)}(u)$ that is Lebesgue integrable and has a bounded moment of order $a^{\text {th }}$, i.e. $\int_{\mathbb{R}}\left|f^{(1)}(u)\right| d u<\infty$ and $\int_{\mathbb{R}}\left|u^{a} f^{(1)}(u)\right| d u<\infty$.

Lemma 2.2.1. Assume $u_{t}$ is iid, mean zero, $\mathbb{E}\left|u_{t}\right|<\infty$, then, under Assumption 2.1-2.2 with $a=1$, we have

$$
\left|\sigma_{\theta, j}(x, y)\right| \leq C j^{1-\mu_{0}}
$$

uniformly in $(x, y) \in \mathbb{R}^{2}$ and $\theta \in \Theta$, for $\mu_{0}>1$ and $C<\infty$

Lemma 2.2.1 provides a decaying rate of the covariance at any percentile in the distribution of residuals similar to the mixing condition. With $a=1$, Assumption 2.2 implies the finiteness of $\mathbb{E}\left|\frac{f^{(1)}\left(u_{t}\right)}{f\left(u_{t}\right)}\right|$ and $\mathbb{E}\left|\frac{u_{t} f^{(1)}\left(u_{t}\right)}{f\left(u_{t}\right)}\right|$, which restricts the tail behavior of $f(u)$. From that perspective, the condition on the derivative of the density functions can be regarded as a slightly stronger version of uniform boundedness of $f(u)$ and $\mathbb{E}\left|u_{t}\right|<\infty$ to control the "thickness" of the tail of the distributions. Most distributions employed in practice are compatible with this condition.

### 2.3. Model estimation under serial independence

In this section, we study the identification of general linear time series models and investigate the asymptotic properties of the proposed estimate based on the generalized spectral distribution function introduced in Section 2.2.

### 2.3.1. Model identification under serial independence

To achieve global identification of $\theta \in \Theta$ based on the residual sequences $u_{t}(\theta)$ for a general linear time series model without imposing causality and invertibility, we propose a criterion using a quadratic distance between the generalized spectral distribution function $H_{\theta}(x, y ; \lambda)$ of residuals $u_{t}(\theta)$ and the counterpart under iidness.

The linear filter $\phi(\theta ; L)$ in (2.2.3) plays a crucial role in the dependence of sequence of $u_{t}(\theta)$. If $u_{t}$ follows a non-Gaussian distribution, $u_{t}(\theta)$ will be serially dependent as long as $\phi(\theta ; L) \neq 1$. The conventional methods based on second-order moments fail to discriminate noncausal and noninvertible processes from causal and invertible counterparts, as this linear filter can generate uncorrelated but not independent sequences like all-pass models, see, e.g., Breidt et al. (2001). In the Gaussian probabilistic structure, being uncorrelated implies independence. Therefore, to achieve identification, we need to impose the following assumption to rule out this possibility.

Assumption 2.3. 1. Given a compact $\Theta \subset \mathbb{R}^{d}$, for any $\theta \neq \theta_{0}, \phi(\theta ; z) \neq a_{0} z^{j_{0}}$ for any $j_{0}$ and any nonzero constant $a_{0}$ in a subset of positive measure of $\mathbb{C}$ such that $|z|=1$.
2. If $|\phi(\theta ; z)|^{2}=1$ a.e. for $z \in \mathbb{C}$ such that $|z|=1$ for some $\theta \neq \theta_{0}$, then $u_{t}$ is non-Gaussian.

Assumption 2.3.1 guarantees that the true iid innovation sequence can be only recovered at $\theta_{0}$. Assumption 2.3.2 controls for the special case when the model and the parametric space permit non-unique serially uncorrelated solutions. If such a linear filter generating an uncorrelated sequence exists with $\phi(\theta ; z) \neq 1$ for some $\theta \neq \theta_{0}$, then the non-Gaussianity is needed for the innovation. Given Assumption 2.3, when $\theta \neq \theta_{0}, u_{t}(\theta)$ is not pairwise independent. This implies $\sigma_{\theta, j}(x, y) \neq 0$ for some $j \neq 0$ and $(x, y) \in \mathbb{R}^{2}$ since there must exist at least one dependent pair of $I\left(u_{t}(\theta) \leq x\right)$ and $I\left(u_{t-j}(\theta) \leq y\right)$ by definition. To exploit all information contained in the distribution of $u_{t}(\theta)$, the $L_{2}$ distance defined on the generalized spectral distribution function is aggregated over $(x, y)$ and frequency $\lambda$ in Cramér-von Mises
criterion,

$$
\begin{align*}
\mathcal{Q}_{0}(\theta) & :=L^{2}\left(H_{\theta}(x, y ; \lambda), H_{\theta}^{i i d}(x, y ; \lambda)\right) \\
& =\int_{\mathbb{R}^{2}} \int_{0}^{1}\left|2 \sum_{j=1}^{\infty} \sigma_{\theta, j}(x, y) \frac{\sin j \pi \lambda}{j \pi}\right|^{2} d \lambda d W(x, y)  \tag{2.3.1}\\
& =2 \sum_{j=1}^{\infty} \frac{1}{(j \pi)^{2}} \int_{\mathbb{R}^{2}} \sigma_{\theta, j}^{2}(x, y) d W(x, y),
\end{align*}
$$

where the last equality comes immediately from Parseval's identity and for any weighting function $W$ which satisfies the following condition:

Assumption 2.4. $W(x, y)=W(x, \infty) W(\infty, y)$ where $W$ is a probability distribution defined on $\mathbb{R}^{2}$, absolutely continuous and strictly increasing.

The unboundedness of the support of $W$ is required for the full characterization of pairwise independence of $\left(u_{t}(\theta), u_{t-j}(\theta)\right)$ at any $j$. The continuous weighting function rules out the special case in which $\sigma_{\theta, j}(x, y) \neq 0$ but $\mathcal{Q}_{0}(\theta)=0$, see Hoeffding (1948). The factorization of the weighting functions simplifies the subsequent analysis of estimates based on this population function without sacrificing any power of detecting pairwise dependence. In fact, the weight function $W$ is not necessarily limited to probability functions but can be extended to any continuous function with $W(\infty, \infty)-W(-\infty,-\infty)<C<\infty$, but Assumption 2.4 simplifies the subjective choice of the weighting function in practice without losing much of generality.

It is worth noting that the population distance function (2.3.1) can be interpreted as an infinite weighted sum of pairwise dependence measure $\int_{\mathbb{R}^{2}} \sigma_{\theta, j}^{2}(x, y) d W(x, y)$, which is a generalization of the test statistic proposed by Skaug and Tjøstheim (1993) by replacing the joint distribution of $u_{t}(\theta)$ and $u_{t-j}(\theta)$ by any function satisfying Assumption 2.4 as a weighting function. The summand is naturally downweighted for higher-order lags by the factor $(j \pi)^{-2}$, which avoids subjective choices in the weighting functions, by contrast to the approach based on the spectral density function.

Remark. Boldin et al. (1997) introduce a sign-based estimation method of causal $A R$ models where the information is exploited from the generalized autocovariance

$$
\mathbb{E}\left(\operatorname{sgn}\left(u_{t}(\theta)\right) \operatorname{sgn}\left(u_{t-j}(\theta)\right)\right)=\mathbb{E}\left(\left(2 I\left(u_{t}(\theta)>0\right)-1\right)\left(2 I\left(u_{t-j}(\theta)>0\right)-1\right)\right)
$$

at any lag $j \geq 1$ with the condition that $F(0)=1 / 2$. The proposed approach in this paper can be viewed as an extension of the sign-based estimator in the sense it takes into account all the percentiles of the distribution rather than only the median, by
replacing $(0,0)$ by any $(x, y) \in \mathbb{R}^{2}$

$$
\int_{\mathbb{R}^{2}} \mathbb{E}\left(\operatorname{sgn}\left(u_{t}(\theta), x\right) \operatorname{sgn}\left(u_{t-j}(\theta), y\right)\right) d W(x, y)
$$

where

$$
\begin{aligned}
\operatorname{sgn}\left(u_{t}(\theta), x\right) & =2 I\left(u_{t}(\theta)>x\right)-2\left(1-\mathbb{E}\left(I\left(u_{t}(\theta) \leq x\right)\right)\right) \\
& =-2 I\left(u_{t}(\theta) \leq x\right)+\mathbb{E}\left(I\left(u_{t}(\theta) \leq x\right)\right)
\end{aligned}
$$

with $\mathbb{E}\left(I\left(u_{t}(\theta) \leq x\right)\right)$ replaced by its sample counterpart and the weighting factors arising from the score of the population generalized covariance. Further analysis compared to the proposed method in this paper can be carried out on the basis of this adaptation.

Under Assumptions 2.1, 2.3 and 2.4, $u_{t}$ being iid with zero mean and $\mathbb{E}\left|u_{t}\right|<\infty$,

$$
\mathcal{Q}_{0}(\theta)>0 \text { when } \theta \neq \theta_{0}
$$

due to some non-degenerated term $\sigma_{\theta, j}^{2}(x, y)$. By the Weierstrass theorem, the continuous non-negative function $\mathcal{Q}_{0}(\theta)$ admits its minimum at 0 in the compact set $\Theta$. Assumption 2.3 guarantees that the minimum can be uniquely attained when $\theta=\theta_{0}$, and thus, the identification of the parameter $\theta_{0}$ in $\Theta$ is achieved.

### 2.3.2. Estimates of model parameters under serial independence

The population loss function $\mathcal{Q}_{0}(\theta)$ is approximated by its sample counterpart in a given sample. In this subsection, we introduce the estimates of model parameters derived from the sample loss function and discuss the consistency of the proposed estimates. In practice, the computed residuals are approximated by a truncated moving average representation due to the limited observations in a given sample of length $T$,

$$
\hat{u}_{t}(\theta)=\varphi^{-1}(\theta, L) Y_{t} I\{1 \leq t \leq T\}
$$

where the information loss,

$$
\delta_{T}(\theta)=: u_{t}(\theta)-\hat{u}_{t}(\theta)=\left(\sum_{j=t}^{\infty} \varphi_{j}^{(-1)}(\theta)+\sum_{j=-\infty}^{t-T-1} \varphi_{j}^{(-1)}(\theta)\right) Y_{t-j}
$$

can be shown to be asymptotically negligible as Assumption 2.1 guarantees that coefficients $\varphi_{j}^{(-1)}(\theta)$ decay at a sufficient rate when $j \rightarrow \infty$ uniformly in $\theta$, see Appendix 2.7.3. Based on the sequence of residuals $\hat{u}_{t}(\theta)$ for any $\theta$, the sample loss
function is constructed as

$$
\begin{equation*}
\hat{\mathcal{Q}}_{T}(\theta)=2 \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{(j \pi)^{2}} \int_{\mathbb{R}^{2}} \hat{\sigma}_{\theta, j}^{2}(x, y) d W(x, y) \tag{2.3.2}
\end{equation*}
$$

where

$$
\begin{equation*}
\hat{\sigma}_{\theta, j}(x, y)=\hat{F}_{\theta, j}(x, y)-\hat{F}_{\theta, j}(x, \infty) \hat{F}_{\theta, j}(\infty, y) \tag{2.3.3}
\end{equation*}
$$

with

$$
\begin{equation*}
\hat{F}_{\theta, j}(x, y)=\frac{1}{T-j} \sum_{t=j+1}^{T} I\left(\hat{u}_{t}(\theta) \leq x\right) I\left(\hat{u}_{t-j}(\theta) \leq y\right) \tag{2.3.4}
\end{equation*}
$$

as a sample analogue of $F_{\theta, j}(x, y)=\mathbb{E}\left(I\left(u_{t}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right)\right)$ and $\left(1-\frac{j}{T}\right)$ is a finite sample correction. The proposed estimator of $\theta_{0}$ is defined as the minimum of $\hat{\mathcal{Q}}_{T}(\theta)$,

$$
\hat{\theta}_{T}=\underset{\theta \in \Theta}{\operatorname{argmin}} \hat{\mathcal{Q}}_{T}(\theta) .
$$

The objective function remains continuous in $\theta$ thanks to the integration, albeit the indicator functions $I\left(\hat{u}_{t}(\theta) \leq x\right)$ introduces non-smoothness into the sample covariance, see Appendix 2.8.1. The consistency of the estimates follows from the identifiable uniqueness of the true model parameter $\theta_{0} \in \Theta$ for $\mathcal{Q}_{0}(\theta)$ and the uniform convergence of the sample loss function to the population one. The latter condition is shown by Theorem 2.1 in Newey (1991) with the assistance of the following Lemma 2.3.1.

Lemma 2.3.1. Assume $u_{t}$ is iid with zero mean, $\mathbb{E}\left(u_{t}^{2}\right)<\infty$, then, under Assumption 2.1-2.2 with $a=2$, for $\theta \in \Theta, \mu_{0}>1$ and $C<\infty, j=1,2, \ldots$,

$$
\mathbb{E}\left|\hat{F}_{\theta, j}(x, y)-F_{\theta, j}(x, y)\right|^{2} \leq C\left(1 \wedge \frac{j}{T-j}\right)+C \frac{j^{2-\mu_{0}}}{T-j}+C \frac{\log T}{(T-j)^{\mu_{0}-1}}
$$

uniformly in $(x, y)$.

Lemma 2.3.1 provides the convergence of the empirical cumulative distribution function to its theoretical counterpart, which further implies the mean square convergence of the sample covariance to the general covariance. Together with the previous assumptions, then, consistency is established in the following theorem.

Theorem 2.3.2. Assume $\left\{u_{t}\right\}$ is iid with zero mean and $\mathbb{E}\left(u_{t}^{2}\right)<\infty$, under Assumption 2.1-2.4 with $a=2, \theta_{0} \in \Theta, \mu_{0}>3, \mu_{1}>1$, as $T \rightarrow \infty$,

$$
\hat{\theta}_{T} \longrightarrow{ }_{p} \theta_{0}
$$

### 2.3.3. Asymptomatic distribution of estimates under serial independence

The non-differentiability of the objective function hinders the classical approach of deriving a CLT for the estimates using scores and Hessian matrices ${ }^{15}$. Neither is the convex analysis approach adopted in the quantile regression able to mitigate the issue since our proposed loss function may attain multiple local minima. To investigate the asymptotic distribution of the proposed estimates, we first approximate the indicator function with a smoothed cumulative distribution function $\Lambda(u)$ indexed by a smoothing parameter $h$ such that

$$
\Lambda\left(\frac{z}{h}\right) \rightarrow I(z>0) \text { for }|z|>0 \text { when } h \rightarrow 0
$$

The following assumption formalizes the conditions of the smoothed $c d f$
Assumption 2.5. The smoothed cdf $\Lambda(u)$ admits uniformly bounded positive Lebesgue pdf $\lambda(u)$ with differentiable first-order and second-order derivatives, $\dot{\lambda}(u)$ and $\ddot{\lambda}(u)$, respectively, uniformly bounded by some constants $C$.

The positiveness of the $p d f \lambda(u)$ ensures no loss of information in the transformation procedure. The uniform boundedness of higher-order derivatives of the density function $\lambda$, together with their smoothness are not required for the model identification but for the convergence of the score and Hessian matrix for the subsequent asymptotic analysis. The new smoothed loss function is obtained by replacing the indicator function with the smoothed $c d f$ in the original formula,

$$
\tilde{\mathcal{Q}}_{T}(\theta ; h)=2 \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{(j \pi)^{2}} \int_{\mathbb{R}^{2}} \tilde{\sigma}_{\theta, j}^{2}(x, y ; h) d W(x, y),
$$

where

$$
\begin{aligned}
\tilde{\sigma}_{\theta, j}(x, y ; h) & =\tilde{F}_{\theta, j}(x, y ; h)-\tilde{F}_{\theta, j}(x, \infty ; h) \tilde{F}_{\theta, j}(\infty, y ; h) \\
\tilde{F}_{\theta, j}(x, y ; h) & =\frac{1}{T-j} \sum_{t=j+1}^{T} \Lambda\left(\frac{x-\hat{u}_{t}(\theta)}{h}\right) \Lambda\left(\frac{y-\hat{u}_{t-j}(\theta)}{h}\right) .
\end{aligned}
$$

The corresponding estimator from the smoothed version is defined by

$$
\tilde{\theta}_{T}^{h}=\underset{\theta \in \Theta}{\operatorname{argmin}} \tilde{\mathcal{Q}}_{T}(\theta ; h) .
$$

We start analyzing the simple case when $h$ is fixed and positive. The identification of the parameter in the model is fulfilled for any given $h>0$ as $\Lambda(\dot{\bar{h}})$ is a transformation

[^14]from the class that is totally revealing, implying the dependence structure of the residual sequences remains unchanged after this smooth transformation. There are many feasible choices, like exponential or trigonometric functions, see Stinchcombe and White (1998). The consistency of this class of estimates $\tilde{\theta}_{T}^{h}$ is formalized by the following theorem after the introduction of Assumption 2.6.

Assumption 2.6. The filter $\phi(\theta ; z)$ is differentiable with the first order derivative $\phi^{(1)}(\theta ; z):=\frac{\partial}{\partial \theta} \phi(\theta ; z)=\sum_{j=-\infty}^{\infty} \phi_{j}^{(1)}(\theta) z^{j}$ such that there exists a $\mu_{1}>1$,

$$
\sup _{\theta \in \Theta}\left\|\phi_{j}^{(1)}(\theta)\right\| \leq C|j|^{-\mu_{1}}, \quad j= \pm 1, \pm 2, \ldots
$$

Theorem 2.3.3. Let $\left\{u_{t}\right\}$ be iid with zero mean, $\mathbb{E}\left|u_{t}\right|<\infty, \theta_{0} \in \Theta$, under Assumptions 2.1, 2.3, 2.4, 2.5 and 2.6, with $\mu_{0}>3, \mu_{1}>1$, as $T \rightarrow \infty$,

$$
\tilde{\theta}_{T}^{h} \longrightarrow_{p} \theta_{0} \text { for any fixed } h>0
$$

The proof of the consistency for $\tilde{\theta}_{T}^{h}$ given any fixed positive $h$ is similar to the consistency theorem in Velasco (2022) by replacing the characteristic function with our proposed smoothed $c d f \Lambda$. Assumption 2.6 imposes further restrictions on the smoothness of the linear filter $\phi_{j}(\theta)$ to achieve uniform boundedness of the derivative of $\tilde{\sigma}_{\theta, j}(x, y ; h)$, leading to the uniform consistency of $\left.\mathcal{Q}_{T} \tilde{\theta} ; h\right)$ to its population counterpart. Thus, unlike the original approach based on the indicator function, Assumption 2.2 is not needed for consistency of the estimates after transformation. In addition, the choice of non-zero smoothing parameter $h$ does not affect consistency once it is fixed.

Prior to analyzing the asymptotic distribution of $\tilde{\theta}_{T}^{h}$, we define the following variables to simplify notation,

$$
\begin{aligned}
e_{t}^{h} & :=\int_{\mathbb{R}}\left(\Lambda\left(\frac{x-u_{t}}{h}\right)-\varphi^{h}(x)\right) \lambda^{h}(x) d W(x) \\
\nu_{t}^{h} & :=\int_{\mathbb{R}}\left(\Lambda\left(\frac{x-u_{t}}{h}\right)-\varphi^{h}(x)\right) \mu^{h}(x) d W(x) \\
E_{t-1}^{h} & :=\sum_{j=1}^{\infty} \frac{1}{j^{2}} \phi_{-j}^{(1)}\left(\theta_{0}\right) e_{t-j}^{h} \\
V_{t-1}^{h} & :=\sum_{j=1}^{\infty} \frac{1}{j^{2}} \phi_{j}^{(1)}\left(\theta_{0}\right) \nu_{t-j}^{h},
\end{aligned}
$$

where
$\varphi^{h}(x):=\mathbb{E}\left(\Lambda\left(\frac{x-u_{t}}{h}\right)\right), \quad \lambda^{h}(x):=\frac{1}{h} \mathbb{E}\left(\lambda\left(\frac{x-u_{t}}{h}\right)\right), \quad \mu^{h}(x):=\mathbb{E}\left(u_{t} \Lambda\left(\frac{x-u_{t}}{h}\right)\right)$.
It is easy to notice that $\left\{e_{t}^{h}\right\},\left\{\nu_{t}^{h}\right\}$ are $i i d$ sequences with mean zero as they are
measurable functions of true iid innovations $u_{t}$. The corresponding variance and covariance are denoted by $\left\{\sigma_{e ; h}^{2}, \sigma_{\nu ; h}^{2}\right\}$ and $\sigma_{e \nu ; h}^{2}$, respectively. From the independence of $u_{t}$ with $\left\{u_{t-j}\right\}_{j \geq 1},\left\{e_{t}^{h} V_{t-1}^{h}\right\},\left\{\nu_{t}^{h} E_{t-1}^{h}\right\}$ are martingale difference sequences conditional on the $\sigma$-field generated by $\left\{u_{t-j}\right\}_{j \geq 1}$. Let

$$
\Sigma_{0, a}:=\sum_{j=1}^{\infty} j^{-2 a} \phi_{j}^{(1)}\left(\theta_{0}\right) \phi_{j}^{(1)}\left(\theta_{0}\right)^{\prime}, \quad \Sigma_{0, a}^{*}:=\sum_{j=1}^{\infty} j^{-2 a} \phi_{-j}^{(1)}\left(\theta_{0}\right) \phi_{-j}^{(1)}\left(\theta_{0}\right)^{\prime}
$$

and $\Sigma_{0, a}^{\dagger}:=\sum_{j=1}^{\infty} j^{-2 a} \phi_{j}^{(1)}\left(\theta_{0}\right) \phi_{-j}^{(1)}\left(\theta_{0}\right)^{\prime}$ for $a=1,2$, and

$$
\begin{aligned}
& H_{1, h}:=\sigma_{e ; h}^{2} \sigma_{\nu ; h}^{2}\left(\Sigma_{0,2}+\Sigma_{0,2}^{*}\right)+\sigma_{e \nu ; h}^{2}\left(\Sigma_{0,2}^{\dagger}+\Sigma_{0,2}^{\dagger^{\prime}}\right) \\
& H_{0, h}:=\rho_{1}^{h} \rho_{2}^{h}\left(\Sigma_{0,1}+\Sigma_{0,1}^{*}\right)+\left(\rho_{12}^{h}\right)^{2}\left(\Sigma_{0,1}^{\dagger}+\Sigma_{0,1}^{\dagger^{\prime}}\right),
\end{aligned}
$$

where
$\rho_{1}^{h}:=\int_{\mathbb{R}}\left(\mu^{h}(x)\right)^{2} d W(x), \quad \rho_{2}^{h}:=\int_{\mathbb{R}}\left(\lambda^{h}(x)\right)^{2} d W(x), \quad \rho_{12}^{h}:=\int_{\mathbb{R}} \mu^{h}(x) \lambda^{h}(x) d W(x)$.
Finally, two further assumptions are imposed for asymptotic normality.
Assumption 2.7. The filter $\phi(\theta ; z)=\sum_{j=-\infty}^{\infty} \phi_{j}(\theta) z^{j}$ with three derivatives $\phi^{(a)}(\theta ; z)$ satisfies following condition:

$$
\sup _{\theta \in \Theta}\left\|\phi_{j}^{(a)}(\theta)\right\|<C|j|^{-\eta_{a}} \text { with } \eta_{a}>1
$$

for $a=1,2,3$ and $C<\infty$.
Assumption 2.8. $H_{0, h}$ is positive definite.

Stronger conditions on the smoothness of the linear filter $\phi(\theta ; z)$ and the $c d f$ $\Lambda$ are imposed for the analysis of the score and the Hessian matrix of $\tilde{\mathcal{Q}}_{T}(\theta ; h)$. The local identification condition Assumption 2.8.2 ensures the components of the covariance-variance matrix are invertible.

Theorem 2.3.4. Let $\left\{u_{t}\right\}$ be iid with zero mean, $\mathbb{E}\left|u_{t}\right|^{3}<\infty$ and $\mu_{0}>3, \mu_{1}>$ 1, $\theta_{0} \in \Theta$, Under Assumptions 2.1, 2.3, 2.4, 2.5 and 2.7-2.8, as $T \rightarrow \infty$,

$$
T^{1 / 2}\left(\tilde{\theta}_{T}^{h}-\theta_{0}\right) \longrightarrow_{d} \mathcal{N}\left(0, H_{0, h}^{-1} H_{1, h} H_{0, h}^{-1}\right)
$$

This intermediate result on the asymptotic distribution of $\tilde{\theta}_{T}^{h}$ by fixing $h$ provides a rough description of the limiting behavior of the original estimate based on the indicator transformation of the residuals. To numerically approximate the asymptotic distribution of $\hat{\theta}_{T}$, we set $h$ appropriately close to zero with $h \rightarrow 0$ as $T \rightarrow \infty$.

Then, as $h \rightarrow 0$,

$$
\varphi^{h}(x) \rightarrow F(x), \quad \lambda^{h}(x) \rightarrow f(x)
$$

and $\mu^{h}(x) \rightarrow \mu(x) \equiv \mathbb{E}\left(u_{t} I\left(u_{t} \leq x\right)\right)$. Correspondingly, the asymptotic variance becomes

$$
\begin{aligned}
& H_{1}:=\sigma_{e}^{2} \sigma_{\nu}^{2}\left(\Sigma_{0,2}+\Sigma_{0,2}^{*}\right)+\sigma_{e \nu}^{2}\left(\Sigma_{0,2}^{\dagger}+\Sigma_{0,2}^{\dagger^{\prime}}\right) \\
& H_{0}:=\rho_{1} \rho_{2}\left(\Sigma_{0,1}+\Sigma_{0,1}^{*}\right)+\left(\rho_{12}\right)^{2}\left(\Sigma_{0,1}^{\dagger}+\Sigma_{0,1}^{\dagger^{\prime}}\right)
\end{aligned}
$$

with $\left\{\sigma_{e}, \sigma_{\nu}, \sigma_{e \nu}, \rho_{1}, \rho_{2}, \rho_{12}\right\}$ described below, as the limits of $\left\{\sigma_{e ; h}, \sigma_{\nu ; h}, \sigma_{e \nu ; h}, \rho_{1}^{h}, \rho_{2}^{h}, \rho_{12}^{h}\right\}$ when $h \rightarrow 0$,

$$
\begin{aligned}
\sigma_{e}^{2} & :=\int_{\mathbb{R}} \int_{\mathbb{R}} F(x \wedge y) f(x) f(y) d W(x) d W(y)-\left(\int_{\mathbb{R}} F(x) f(x) d W(x)\right)^{2} \\
\sigma_{\nu} & :=\int_{\mathbb{R}} \int_{\mathbb{R}} F(x \wedge y) \mu(x) \mu(y) d W(x) d W(y)-\left(\int_{\mathbb{R}} F(x) \mu(x) d W(x)\right)^{2} \\
\sigma_{e \nu} & :=\int_{\mathbb{R}} \int_{\mathbb{R}} F(x \wedge y) f(x) \mu(y) d W(x) d W(y)-\int_{\mathbb{R}} F(x) f(x) d W(x)\left(\int_{\mathbb{R}} F(y) \mu(y) d W(y)\right),
\end{aligned}
$$

and

$$
\rho_{1}:=\int_{\mathbb{R}} \mu^{2}(x) d W(x), \quad \rho_{2}:=\int_{\mathbb{R}} f^{2}(x) d W(x), \quad \rho_{12}:=\int_{\mathbb{R}} \mu(x) f(x) d W(x)
$$

The loss function $\mathcal{Q}_{0}(\theta ; h)$ still identifies $\theta_{0}$ under the same structure as $h$ goes to zero. This immediately follows from

$$
\mathcal{Q}_{0}(\theta ; h)=\mathcal{Q}_{0}(\theta)+O\left(h^{2}\right) \text { uniformly in } \theta \in \Theta
$$

Some extra care needs to be taken on the smoothing parameter $h$ to preserve the classical rate $T^{1 / 2}$ in the application of CLT when $h \rightarrow 0$ with $T \rightarrow \infty$. In effect, the rate of convergence of $\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta, j}(x, y ; h)$ needs to be controlled unchanged as $h \rightarrow 0$. Based on our analysis of the mean squared error of the covariance derivative of estimates,

$$
\mathbb{E}\left\|\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y ; h)-\frac{\partial}{\partial \theta} \sigma_{\theta_{0}, j}(x, y)\right\|^{2}=O\left(h^{4}+h^{-1}(T-j)^{-1}\right)
$$

where $\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y ; h):=\left.\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta, j}(x, y ; h)\right|_{\theta=\theta_{0}}$ and $\frac{\partial}{\partial \theta} \sigma_{\theta_{0}, j}(x, y):=\left.\frac{\partial}{\partial \theta} \sigma_{\theta, j}(x, y)\right|_{\theta=\theta_{0}}$, we can conclude that for the approximation effect to be asymptotically negligible, the value of $h$ must approach zero, but at a rate not faster than $T^{-1}$. Another important consideration is the non-centering error of the residual covariance as $h \rightarrow 0$,

$$
\varphi^{h}(x) \varphi^{h}(y)=F(x) F(y)+O\left(h^{2}\right)
$$

uniformly in $(x, y)$, which imposes one further restriction. The bias must be $o\left(T^{-1 / 2}\right)$
to be negligible in the asymptotic distribution of the estimates $\tilde{\theta}_{T}^{h}$ when $h$ converges to zero. Hence, $h$ cannot go to zero too slowly, implying that it must approach zero at a rate of $o\left(T^{-1 / 4}\right)$. Some of these rates can be possibly improved with higherorder kernels, but since $h$ has no direct effect on the first-order asymptotics, we do not pursue this.

The following conditions in parallel with Assumption 2.8 are imposed for the analysis of the limiting behavior of $\tilde{\theta}_{T}^{h}$ as $h \rightarrow 0$.

Assumption 2.9. 1. The innovation $u_{t}$ admits uniformly bounded probability density function $f(u)$ with differentiable derivatives $f^{(a)}(u)$ of order a uniformly bounded by some constants $C<\infty$ for $a=1,2$.

## 2. $H_{0}$ is positive definite.

Theorem 2.3.5. Under Assumptions 2.1, 2.3, 2.4 and Assumptions 2.7-2.9, ut iid with zero mean, $\mathbb{E}\left|u_{t}\right|^{3}<\infty, \theta_{0} \in \Theta, \mu_{0}>3, \mu_{1}>1$, as $T \rightarrow \infty, h^{-1} T^{-1}+h^{4} T \rightarrow 0$

$$
T^{1 / 2}\left(\tilde{\theta}_{T}^{h}-\theta_{0}\right) \longrightarrow_{p} \mathcal{N}\left(0, H_{0}^{-1} H_{1} H_{0}^{-1}\right)
$$

Theorem 2.3.5 presents the asymptotic distribution of the estimates based on the smoothed $c d f$ when $\Lambda$ approaches the indicator function as $T \rightarrow \infty$. We expect that this also represents the asymptotic behavior of $\hat{\theta}_{T}$ obtained from the sample loss function based on the empirical $c d f$, as numerical experiments show that the difference between both estimates in magnitude is trivial as $h \rightarrow 0$, see the evidence in Simulation 2.4.2.

The asymptotic variance in the causal and invertible models can be simplified to

$$
\kappa \Sigma_{0,1}^{-1} \Sigma_{0,2} \Sigma_{0,1}^{-1}
$$

with $\kappa=\sigma_{e}^{2} \sigma_{\nu}^{2}\left(\rho_{1} \rho_{2}\right)^{-2}$ since some components like $\Sigma_{0, a}^{*}$ and $\Sigma_{0, a}^{\dagger}$ are degenerated to zero for $a=1,2$. There are potential gains in the efficiency of the estimates by replacing the spectral $c d f$ with the spectral $p d f$ in the loss function, in which the natural down weights $j^{-2}$ on the higher lags are not imposed. In such case, the asymptotic variance of estimates becomes $\kappa \Sigma_{0,0}^{-1}$, and it can be shown that $\Sigma_{0,1}^{-1} \Sigma_{0,2} \Sigma_{0,1}^{-1}-\Sigma_{0,0}^{-1}$ is positive semidefinite.

### 2.3.4. Standard error calculation

As shown in Theorems 2.3.4 and 2.3.5, the asymptotic variance of the proposed estimates crucially depends on the data-generating processes (DGP), resulting in a lack of closed-form expressions of $H_{0}$ and $H_{1}$ in general. Cavaliere et al. (2020) have investigated several bootstrapping schemes for the inference study of noncausal processes
with heavy-tailed innovations. In this section, we outline a method for constructing the approximated standard errors of the estimator $\hat{\theta}_{T}$ based on natural estimates of the unknown quantities depending on the DGP. In comparison to bootstrapping methods, our approach involves estimating the components in the asymptotic variance directly, which reduces computational expenses by avoiding repetitive calculations. Using the definitions of $e_{t}$ and $\nu_{t}$, we can substitute them by their sample counterparts:

$$
\begin{aligned}
& \hat{e}_{t}:=\int_{\mathbb{R}}\left(I\left(\hat{u}_{t}\left(\hat{\theta}_{T}\right) \leq x\right)-\hat{F}_{T}(x)\right) \hat{f}_{T}(x) d W(x) \\
& \hat{\nu}_{t}:=\int_{\mathbb{R}}\left(I\left(\hat{u}_{t}\left(\hat{\theta}_{T}\right) \leq x\right)-\hat{F}_{T}(x)\right) \hat{\mu}_{T}(x) d W(x)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{F}_{T}(x) & =\frac{1}{T} \sum_{t=1}^{T} I\left(\hat{u}_{t}\left(\hat{\theta}_{T}\right) \leq x\right) \\
\hat{f}_{T}(x) & =\frac{\hat{F}_{T}(x+a)-\hat{F}_{T}(x-a)}{2 a} \text { for a properly chosen value of } a \\
\text { or } & =\frac{1}{T h} \sum_{t=1}^{T} K\left(\frac{x-\hat{u}_{t}\left(\hat{\theta}_{T}\right)}{h}\right) \text { for a sufficiently small } h \text { and kernel function } K \\
\hat{\mu}_{T}(x) & =\frac{1}{T} \sum_{t=1}^{T} \hat{u}_{t}\left(\hat{\theta}_{T}\right) I\left(\hat{u}_{t}\left(\hat{\theta}_{T}\right) \leq x\right)
\end{aligned}
$$

The estimation of the density function can be obtained using a naive estimator or any consistent kernel density estimator. To calculate the integration in $\hat{e}_{t}$ and $\hat{\nu}_{t}$, a numerical integration algorithm will be employed with weighting functions $W(x)$. Then, $\left\{\hat{\sigma}_{e}^{2}, \hat{\sigma}_{\nu}^{2}, \hat{\sigma}_{e \nu}^{2}\right\}$ are the elements of the sample variance-covariance matrix of $\left(\hat{e}_{t}, \hat{\nu}_{t}\right)$. The estimates of $\left\{\rho_{1}, \rho_{2}, \rho_{12}\right\}$ also require numerical integration of $\left\{\left(\hat{\mu}_{T}(x)\right)^{2},\left(\hat{f}_{T}(x)\right)^{2}, \hat{\mu}_{T}(x) \hat{f}_{T}(x)\right\}$ with respect to $x$ over a chosen $W(x)$. The derivatives of the linear filter can be obtained from the model once the order is determined and $\left\{\Sigma_{0, a}, \Sigma_{0, a}^{*}, \Sigma_{0, a}^{\dagger}\right\}_{a=1,2}$ are estimated by plugging $\hat{\theta}_{T}$ in the corresponding expressions. Alternatively, to avoid numerical integration, one can choose the empirical $c d f$ as a weighting function to replace integration by averages at data points, in such case, the estimators can be computed as follows:

$$
\begin{aligned}
& \hat{e}_{t}:=\frac{1}{T} \sum_{s=1}^{T}\left(I\left(\hat{u}_{t}\left(\hat{\theta}_{T}\right) \leq \hat{u}_{s}\left(\hat{\theta}_{T}\right)\right)-\hat{F}_{T}\left(\hat{u}_{s}\left(\hat{\theta}_{T}\right)\right)\right) \hat{f}_{T}\left(\hat{u}_{s}\left(\hat{\theta}_{T}\right)\right) \\
& \hat{\nu}_{t}:=\frac{1}{T} \sum_{s=1}^{T}\left(I\left(\hat{u}_{t}\left(\hat{\theta}_{T}\right) \leq \hat{u}_{s}\left(\hat{\theta}_{T}\right)\right)-\hat{F}_{T}\left(\hat{u}_{s}\left(\hat{\theta}_{T}\right)\right)\right) \hat{\mu}_{T}\left(\hat{u}_{s}\left(\hat{\theta}_{T}\right)\right) .
\end{aligned}
$$

The rest components can be easily adapted in accordance with the empirical $c d f$.
Remark. When computing the standard error of the estimate $\tilde{\theta}_{T}^{h}$ for any fixed $h$, the
procedures are similar, with some modifications. In particular, the indicator function in the formula is replaced by $\Lambda(\cdot)$ index by $h$ and the corresponding density function by $\lambda(\cdot)$, as the smoothed distribution function is determined at the outset, and other kernel functional forms would not be a better fit compared to its real functional form.

### 2.4. Simulations

In this section, we conduct Monte Carlo simulations to investigate the finite sample properties of the proposed estimates. Our methods are robust to the choice of weighting functions $W$ as long as it satisfies Assumption 2.4. However, in practice, the finite sample performance can vary over different options. Since the indicator function in the loss function is not scale-invariant, the weights imposed on $\hat{\sigma}_{\theta, j}(x, y)$ may change before and after the rescaling of residuals $u_{t}(\theta)$. For example, if $W$ is selected to be the standard normal distribution, residuals $u_{t}(\theta)$ that fall outside the interval $(-3,3)$ will be assigned very small weights due to the $3-\sigma$ rule of thumb. This could lead to decreased efficiency of the estimates as they do not fully utilize the information in the extremes (distribution tails) of the residuals, which may be more informative than the median for detecting noncausality in some cases, as discussed in Jin (2023). To address this issue, we propose two solutions.

The first approach involves standardizing the residuals by dividing the original sequence by its standard deviation

$$
u_{t}^{*}(\theta):=\frac{u_{t}(\theta)}{\sqrt{\frac{1}{T} \sum_{t=1}^{T}\left(u_{t}(\theta)-\overline{u(\theta)}\right)^{2}}},
$$

where $\overline{u(\theta)}$ is the sample mean of the residual sequence. The standard normal distribution is chosen to be $W .{ }^{16}$ The detailed calculation of the loss function in finite samples can be found in Appendix 2.8.1. It can be shown that this standardization does not change the asymptotic properties of the proposed estimates but improves finite sample performance. The second approach relies on fitting the weighting function using the empirical distribution function of residuals, which plays a role as an automated rescaling scheme for each sequence of residuals correspondingly. In such case, the loss function can be simplified to

$$
2 \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{(j \pi)^{2}} \sum_{s=1}^{T} \sum_{t=1}^{T} \hat{\sigma}_{\theta, j}^{2}\left(u_{s}(\theta), u_{t}(\theta)\right)
$$

Both approaches avoid numerical integration and subjective choice of rescaling pa-

[^15]rameters.

### 2.4.1. Indicator approach: $h=0$

In the first experiment, we examine the finite sample performance of the estimates based on the indicator function, i.e., $\hat{\theta}_{T}$, under different processes and sample sizes. The first example is designed to evaluate the robustness of the methodology in $\mathrm{AR}(1)$ processes with different types of innovations. We consider iid innovations distributed by uniform distribution $U_{[-5,5]}$, t-distribution $t_{3}$ and centered chi-square distribution $\chi_{5}$ for the generation of $\mathrm{AR}(1)$ processes, including basic non-Gaussian distributions with symmetry and asymmetry $\left(\kappa_{3}=0,0\right.$ and $\left.\sqrt{\frac{8}{5}}\right)$, bounded support, and heavytailed property $\left(\kappa_{4}=-\frac{6}{5}, \infty\right.$ and $\left.\frac{12}{5}\right) \cdot{ }^{17}$ The sample sizes $T$ are 100 and 200 and the parameters of the $\mathrm{AR}(1)$ models are chosen to be $0.4\left(0.4^{-1}\right)$ and $0.9\left(0.9^{-1}\right)$ in the causal (noncausal) cases. The results, based on 100 replications, are reported in Table 2.6.1 in terms of the percentage of correct root identification (PCI), the bias of the estimates (Bias), and the mean squared errors (MSE) given the sample size, the distribution of innovations and the true parameters of the $\operatorname{AR}(1)$ models using both weighting schemes.

The results of our proposed method show that the method performs better when the innovation exhibits either heavy tails or asymmetry. This is because we are able to gain more information from high skewness and kurtosis, which are commonly employed to measure non-Gaussianity. Our findings are in line with the estimation technique proposed by Velasco and Lobato (2018), which also leverages information from higher-order moments. As the sample size increases, the percentage of correct root identification improves and the mean squared error (MSE) of the estimates decreases. Both standardization and weighting by the empirical cumulative distribution function ( $c d f$ ) demonstrate good performance. However, when the innovation follows $t_{3}$ or $\chi_{5}$ distribution, the standard normal $c d f$ approach outperforms the empirical $c d f$ approach. Conversely, when the innovation follows a uniform distribution, the empirical $c d f$ approach performs better. Based on these results, we suggest using the standard normal $c d f$ approach for the empirical examples, unless there is strong evidence suggesting a uniform distribution or negative kurtosis with symmetry for the innovation. Additionally, we also report the results of estimation when the true parameter is close to unity. As expected, it can be challenging to determine whether the root is located inside or outside the unit circle, as they are similarly magnified.

In the second example, our aim is to estimate an $\mathrm{AR}(2)$ process generated by innovations with a centered $\chi_{5}$ distribution. In this situation, the objective is to in-

[^16]vestigate the identification of causality and noncausality of AR processes with higher orders, as $\mathrm{AR}(p)$ processes with $p \geq 2$ can have different combinations of nonfundamental solutions, such as the mixed causal-noncausal processes discussed by Hecq et al. (2016). We select $\theta_{0,1}=0.4$ and $\theta_{0,2}=0.8$ as two parameters of the polynomial $\left(1-\theta_{0,1} L\right)\left(1-\theta_{0,2} L\right)$ in the causal case, and generate three other types of processes. The sample size is varied from 100 to 200 , with 100 replications per simulation. The results are presented in Table 2.6.2, which displays four types of processes in terms of causality: two are mixed causal-noncausal $\mathrm{AR}(2)$ processes, one purely causal $\mathrm{AR}(2)$ process, and one purely noncausal $\mathrm{AR}(2)$ process. In this experiment, our discussion is limited to the identification of the number of noncausal roots detected in the processes ${ }^{18}$. The first row of Table 2.6.2 reports the percentage of correct identification of the number of noncausal roots in the process, while the second row presents the percentage of identification that detects the existence of noncausality. The results indicate that the proposed method performs better in processes with noncausality than in causal cases. In the purely casual case, the performance is relatively poor when the sample size is small $(T=100)$, but the proportion of correct identification increases significantly when the sample size increases to 200 .

### 2.4.2. Smooth $c d f$ approach: $h>0$

In Section 2.3.3, we present the strategy to approximate the asymptotic distribution of the estimates based on the indicator function. The approximation is achieved by substituting the indicator function with its smooth counterpart, allowing $h \rightarrow 0$ when $T \rightarrow \infty$. We also provide a general analysis of the convergence rate of $h$. In this subsection, we carry out some experiments to evaluate the numerical approximation performance of the smooth $c d f$ estimates using different values of $h$. As per Theorem 2.3.5, $h$ is restricted between $T^{-1}$ and $T^{-1 / 4}$, so we select 4 rates of $h$ associated with $T: T^{-0.99}, T^{-3 / 4}, T^{-1 / 2}, T^{-0.26}$. We compare the corresponding estimates with the one based on the indicator function $(h=0)$. We generate the $\mathrm{AR}(1)$ process from an iid innovation sequence with $\chi_{5}^{2}$ distribution and the parameter 0.4 in the causal case and 2.5 in the noncausal case. The sample size is 100 for each replication. We set the smooth $c d f \Lambda$ to the logistic distribution and adopt empirical $c d f$ of residuals as weighting functions to avoid numerical integration in calculating the

[^17]loss function ${ }^{19}$. The numerical results from 100 replications are presented in Table 2.6.3. The first three items, namely, PCI, Bias, and Root MSE (RMSE), provide a general evaluation of the performance of the estimates with different values of $h$. Overall, the estimates' performance does not exhibit substantial differences across different choices of $h$. With regard to PCI, the estimates $\tilde{\theta}_{T}^{h}$ tend to achieve a higher correct root identification rate in the causal case when $h$ converges to 0 at a slower rate. This behavior may be attributed to more concentration near zero when $h$ is larger in terms of $T$, which is the median of the logistic distribution, when $h$ converges more slowly. As for the bias and the RMSE, both measures show only subtle fluctuations. The last item, relative RMSE (RRMSE), aims to assess the approximation performance of $\tilde{\theta}_{T}^{h}$ to $\hat{\theta}_{T}$ by measuring their similarity. The RRMSE is computed by the RMSE normalized by the RMSE of the estimates based on the indicator functions. The results in Table 2.6.3 indicate a better approximation performance of $\tilde{\theta}_{T}^{h}$ to $\hat{\theta}_{T}$ when the sample size increases. Moreover, it is worth noting that the approximation performance of $\tilde{\theta}_{T}^{h}$ to $\hat{\theta}_{T}$ is not proportional to $h$, and it does not vary too much with different choices of $h$.

### 2.5. Empirical application

### 2.5.1. Clustering volatility

In this section, we apply our proposed method to analyze a series of 753 daily trading volumes of Microsoft (MSFT) stock from $6 / 3 / 1996$ to $5 / 26 / 1999$. Breidt et al. (2001) have argued that a noncausal $\operatorname{AR}(1)$ model better fits the data than a causal $\mathrm{AR}(1)$ model, based on diagnostic tests of residuals computed from both models. To prepare the data for analysis, we remove the heteroskedasticity and drift by taking the logarithm and demeaning the sequence, as shown in the upper panel of Figure 2.6.1. The augmented Dickey-Fuller (ADF) test indicates that the resulting sequence is stationary, with no unit root present. Additionally, the partial autocorrelation of the sequence suggests that an AR model with order 1 or 3 might be appropriate, as shown in the lower panel of Figure 2.6.1. We fit the data by an AR model of order 1 and our proposed method yields a noncausal model as a stationary solution,

$$
\hat{u}_{t}=Y_{t}-\underset{(0.6059)}{1.7953 Y_{t-1} .}
$$

[^18]In other words, the forward-looking behavior of investors has a stronger impact on stock investment than the backward-looking behavior. This is evident in the clustering phenomenon observed in the stock market, which can be attributed to investors' uncertainty about the future market environment. Intuitively, the perception of a more volatile market environment in the future would lead to more variant investment strategies of investors of different types. Instead of fitting the data with the AR-ARCH model, with the noncausal $\operatorname{AR}(1)$ model, which can mimic clustering volatility dynamics, as demonstrated by Breidt et al. (2001), we can avoid estimating more parameters. We plot the autocorrelation function (ACF) of the squared value of residuals $\left\{\tilde{u}_{t}^{2}\right\}$ and $\left\{\hat{u}_{t}^{2}\right\}$ from the causal model using conventional Gaussian MLE ( $\tilde{u}_{t}=Y_{t}-0.5854 Y_{t-1}$ ) and noncausal model using our approach, respectively. The upper part of the Figure 2.6.2 displays a clear correlation in the squared residuals from the causal $\mathrm{AR}(1)$ model at first lag, indicating ARCH would be an appropriate alternative to characterize the residual dependence. The lower part of Figure 2.6.2 shows residuals generated from the noncausal model do not show a significant correlation on the squared value of the residuals. Apart from volatility clustering behavior, noncausal autoregressive processes also can capture the dependence between different percentiles, which is not possible with conventional GARCH models. These nonlinear dynamics observed in Microsoft stock data may be linked to the "informational heterogeneity" of investors who receive idiosyncratic messages from the shocks in the financial market, resulting in belief dynamics of different orders ${ }^{20}$ (Kasa et al. (2006)).

### 2.6. Discussion: Measures of dependence under a martingale difference assumption

The assumption on the innovations used to generate data process can be possibly relaxed to martingale difference sequence (mds), which broadens the class of time series models by including linear models with ARCH-type innovations.
Denote the $\sigma$-field generated by the past sequence of $u_{t}$ by $\sigma_{t-1}=\sigma\left(u_{t-1}, u_{t-2}, \ldots\right)$. By the definition of mds, we obtain

$$
\mathbb{E}\left(u_{t} \mid \sigma_{t-1}\right)=0,
$$

which implies

$$
\mathbb{E}\left(u_{t} I\left(u_{t-j}(\theta) \leq x\right)\right)=0 \text { for } j \geq 1 \quad \forall x \in \mathbb{R}
$$

[^19]The indicator function can be replaced by the exponential function $\exp (\cdot)$ or any other parametric family considered in Escanciano (2006). Here we choose the indicator function to be consistent with the notation introduced throughout this paper and its simplicity in computation.
We define the following measure of dependence

$$
\gamma_{\theta, j}(x)=\mathbb{E}\left(u_{t}(\theta) I\left(u_{t-j}(\theta) \leq x\right)\right) \text { for } j \geq 1, \forall x \in \mathbb{R}
$$

and $\gamma_{\theta, j}(x)=\gamma_{\theta,|j|}(x)$ for $j<0$. The corresponding spectral density and distribution functions based on this measure are

$$
\begin{aligned}
d_{\theta}(x ; \omega) & =\frac{1}{2 \pi} \sum_{j=-\infty}^{\infty} \gamma_{j}(x) e^{-i j \omega}, \quad \omega \in[-\pi, \pi] \\
D_{\theta}(x ; \lambda) & =\gamma_{\theta, 0}(x) \lambda+2 \sum_{j=1}^{\infty} \gamma_{\theta, j}(x) \frac{\sin j \pi \lambda}{j \pi}, \quad \lambda \in[0,1] .
\end{aligned}
$$

Following the same approach in the iid case, the population loss function is constructed by a $L_{2}$ distance of the generalized $c d f$ in the unrestricted case and the conjectured one in the restricted case

$$
\begin{aligned}
\mathcal{Q}_{0}^{m d s}(\theta): & =L^{2}\left(D_{\theta}(x ; \lambda), D_{\theta}^{m d s}(x ; \lambda)\right) \\
& =2 \int_{\mathbb{R}^{2}} \sum_{j=1}^{\infty} \gamma_{\theta, j}^{2}(x) \frac{1}{(j \pi)^{2}} d W(x)
\end{aligned}
$$

The study on the identification and estimation of the model is left to further research.

Table 2.6.1: Comparison of estimates using empirical $c d f$ and standard normal $c d f$ under different distributions of innovations: study in $\operatorname{AR}(1)$ case

| $u_{t}$ | T | $\begin{aligned} & \overline{\mathrm{W}}: \\ & \theta_{0}: \end{aligned}$ |  | empirical $c d f$ |  |  |  | standard normal $c d f$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | 0.4 | $0.4^{-1}$ | 0.9 | $0.9^{-1}$ | 0.4 | $0.4^{-1}$ | 0.9 | $0.9^{-1}$ |
| $U_{[-5,5]}$ | 100 | PCI |  | 67.00\% | 84.00\% | 53.00\% | $54.00 \%$ | 62.00\% | 77.00\% | 45.00\% | 41.00\% |
|  |  | Bias |  | -0.0148 | 0.3112 | -0.0254 | 0.0421 | -0.0211 | 0.5532 | -0.0405 | 0.0598 |
|  |  | MSE |  | 0.0092 | 0.5532 | 0.0067 | 0.0146 | 0.0122 | 0.09017 | 0.0076 | 0.0177 |
|  | 200 | PCI |  | 87.00\% | 86.00\% | 64.00\% | 63.00\% | 82.00\% | 78.00\% | 50.00\% | 54.00\% |
|  |  | Bias |  | -0.0101 | 0.0507 | -0.0130 | 0.0155 | -0.0169 | 0.0964 | -0.0203 | 0.0206 |
|  |  | MSE |  | 0.0046 | 0.2240 | 0.0031 | 0.0047 | 0.0054 | 0.3443 | 0.0037 | 0.0049 |
| $\overline{t_{3}}$ | 100 | PCI |  | 60.00\% | 69.00\% | 69.00\% | 57.00\% | 81.00\% | 86.00\% | 75.00\% | 72.00\% |
|  |  | Bias |  | -0.0089 | 0.1421 | -0.0079 | 0.0170 | -0.0145 | 0.0792 | -0.0162 | 0.0257 |
|  |  | MSE |  | 0.0131 | 0.4052 | 0.0036 | 0.0079 | 0.0102 | 0.2954 | 0.0035 | 0.0075 |
|  | 200 | PCI |  | 76.00\% | 75.00\% | 63.00\% | 59.00\% | 90.00\% | 84.00\% | 71.00\% | 76.00\% |
|  |  | Bias |  | -0.0105 | 0.1740 | 0.0121 | 0.0053 | -0.0087 | 0.1285 | 0.0023 | 0.0162 |
|  |  | MSE |  | 0.0037 | 0.2849 | 0.0026 | 0.0043 | 0.0034 | 0.2090 | 0.0024 | 0.0040 |
| $\chi_{5}$ | 100 | PCI |  | 94.00\% | 94.00\% | 71.00\% | 69.00\% | 94.00\% | 93.00\% | 78.00\% | 73.00\% |
|  |  | Bias |  | -0.0035 | 0.1921 | -0.0372 | 0.0550 | -0.0062 | 0.2085 | -0.0377 | 0.0699 |
|  |  | MSE |  | 0.0078 | 0.6751 | 0.0062 | 0.0149 | 0.0071 | 0.6660 | 0.0063 | 0.0181 |
|  | 200 | PCI |  | 99.00\% | 98.00\% | 80.00\% | 76.00\% | 99.00\% | 99.00\% | 81.00\% | 77.00\% |
|  |  | Bias |  | -0.0101 | 0.1082 | 0.0060 | 0.0134 | -0.0085 | 0.1674 | 0.0034 | 0.0025 |
|  |  | MSE |  | 0.0045 | 0.2281 | 0.0032 | 0.0041 | 0.0045 | 0.3838 | 0.0031 | 0.0043 |

PCI: percentage of correct root identification using our method
Bias: computed by $\hat{\mathbb{E}}_{T}\left(\hat{\theta}_{T}\right)-\theta_{0}$
MSE (mean squared error): computed by sum of $\widehat{\operatorname{Var}_{T}\left(\hat{\theta}_{T}\right)}$ and Bias ${ }^{2}$ using the correctly identified replications

Table 2.6.2: Estimates of $\operatorname{AR}(2)$ generated by innovations following $\chi_{5}^{2}$

| $\chi^{2}(5)-5$ |  |  |  |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\mathrm{~T}=100$ | $(0.4,0.8)$ | $\left(0.4^{-1}, 0.8^{-1}\right)$ | $\left(0.4^{-1}, 0.8\right)$ | $\left(0.4,0.8^{-1}\right)$ | $(0.4,0.8)$ | $\left(0.4^{-1}, 0.8^{-1}\right)$ | $\left(0.4^{-1}, 0.8\right)$ |
|  |  | $\left(0.4,0.8^{-1}\right)$ |  |  |  |  |  |  |
| PCI |  | $81.00 \%$ | $84.00 \%$ | $85.00 \%$ | $80.00 \%$ | $94.00 \%$ | $95.00 \%$ | $90.00 \%$ |
| PN |  | $95.00 \%$ | $96.00 \%$ | $85.00 \%$ | $20.00 \%$ | $100.00 \%$ | $98.00 \%$ | $99.00 \%$ |

PCI: percentage of correct root identification including the number of roots lying inside unit circle
PN: percentage of detecting the existence of noncausality in the process. i.e., There is at least one root lying inside unit circle.

Table 2.6.3: Performance of the estimates by approximating the indicator function using smoothed $c d f$ with different convergence rates of $h$ : $\operatorname{AR}(1)$ driven by $\chi_{5}^{2}$ innovations

|  | causal AR(1) |  |  |  |  | noncausal AR(1) |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\overline{T=100}$ | $h=0$ | $h=T^{-0.99}$ | $h=T^{-3 / 4}$ | $h=T^{-1 / 2}$ | $h=T^{-0.26}$ | $h=0$ | $h=T^{-0.99}$ | $h=T^{-3 / 4}$ | $h=T^{-1 / 2}$ | $h=T^{-0.26}$ |
| PCI | 88.00\% | 90.00\% | 92.00\% | 93.00\% | 96.00\% | 90.00\% | 90.00\% | 90.00\% | 88.00\% | 90.00\% |
| Bias | -0.0097 | -0.0067 | -0.0096 | -0.0112 | -0.0162 | 0.0994 | 0.0938 | 0.0796 | 0.0684 | 0.0883 |
| RMSE | 0.0837 | 0.0777 | 0.0763 | 0.0739 | 0.0750 | 0.7214 | 0.6424 | 0.6415 | 0.6345 | 0.6427 |
| RRMSE | 1.0000 | 0.9279 | 0.9111 | 0.8823 | 0.8957 | 1.0000 | 0.8905 | 0.8892 | 0.8796 | 0.8909 |
| $\overline{T=200}$ | $h=0$ | $h=T^{-0.99}$ | $h=T^{-3 / 4}$ | $h=T^{-1 / 2}$ | $h=T^{-0.26}$ | $h=0$ | $h=T^{-0.99}$ | $h=T^{-3 / 4}$ | $h=T^{-1 / 2}$ | $h=T^{-0.26}$ |
| $\overline{\text { PCI }}$ | 99.00\% | 99.00\% | 99.00\% | 99.00\% | 100.00\% | 100.00\% | 100.00\% | 100.00\% | 100.00\% | 100.00\% |
| Bias | -0.0094 | -0.0091 | -0.0096 | -0.0096 | -0.0091 | 0.0519 | 0.0521 | 0.0497 | 0.0512 | 0.0518 |
| RMSE | 0.0596 | 0.0585 | 0.0595 | 0.0597 | 0.0593 | 0.4476 | 0.4479 | 0.4458 | 0.4403 | 0.4407 |
| RRMSE | 1.0000 | 0.9828 | 0.9993 | 1.0026 | 0.9952 | 1.0000 | 1.0006 | 0.9960 | 0.9835 | 0.9846 |

PCI: percentage of correct root identification using our method
Bias: computed by $\hat{\mathbb{E}}_{T}\left(\hat{\theta}_{T}\right)-\theta_{0}$
RMSE (root mean squared error): computed by squared value of the sum of $\mathbb{V a r}_{T}\left(\hat{\theta}_{T}\right)$ and $\operatorname{Bias}^{2}$ using the correctly identified replications
RRMSE (relative RMSE): root mean squared error of the estimates normalized by the RMSE of the estimates based on the indicator functions.

Figure 2.6.1: Microsoft daily trading volume from 6/3/1993 to 5/26/1999


Figure 2.6.2: Diagnostics of residuals from both causal and non-causal models: a comparison in ACF of residuals in squared values


Sample autocorrelation function of squared residuals from non-causal AR(1)


### 2.7. Appendix A: Proofs

### 2.7.1. Proof of Lemma 2.2.1

Define

$$
F_{\theta}(x):=\mathbb{E}\left(\hat{F}_{\theta, j}(x, \infty)\right)=\mathbb{E}\left(I\left(u_{t}(\theta) \leq x\right)\right)=\mathbb{E}\left(F\left(x-u_{t}^{(0)}(\theta)\right)\right)
$$

where $\phi_{0}(\theta)=1$ for all $\theta \in \Theta$,

$$
u_{t}(\theta)=\sum_{j=-\infty}^{\infty} \phi_{j}(\theta) u_{t-j}, \quad u_{t}^{(\mathcal{I})}(\theta)=u_{t}(\theta)-\sum_{j \in \mathcal{I}} \phi_{j}(\theta) u_{t-j} .
$$

For any $j>0$,

$$
\begin{aligned}
& \mathbb{E}\left(\hat{F}_{\theta, j}(x, y)\right) \\
= & \mathbb{E}\left(I\left(u_{t}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right)\right) \\
= & \mathbb{E}\left(I\left(u_{t} \leq x-u_{t}^{(0)}(\theta)\right) I\left(\phi_{-j}(\theta) u_{t} \leq y-u_{t-j}^{(-j)}(\theta)\right)\right) \\
= & \mathbb{E}\left(I\left(u_{t}+\phi_{j}(\theta) u_{t-j} \leq x-u_{t}^{(0, j)}(\theta)\right) I\left(\phi_{-j}(\theta) u_{t}+u_{t-j} \leq y-u_{t-j}^{(0,-j)}(\theta)\right)\right) \\
= & \mathbb{E}\left(\int I\left(z+\phi_{j}(\theta) w \leq x-u_{t}^{(0, j)}(\theta)\right) I\left(\phi_{-j}(\theta) z+w \leq y-u_{t-j}^{(0,-j)(\theta)}\right) f(z) f(w) d z d w\right) \\
= & \mathbb{E}\left(\int I\left(u \leq x-u_{t}^{(0, j)}(\theta)\right) I\left(v \leq y-u_{t-j}^{(0,-j)}(\theta)\right) f_{u, v}^{(j)}(u, v) d u d v\right) \\
= & \mathbb{E}\left(F_{u, v}^{(j)}\left(x-u_{t}^{(0, j)}(\theta), y-u_{t-j}^{(0,-j)}(\theta)\right)\right)
\end{aligned}
$$

where the last second equality comes from the change of variables

$$
\binom{u}{v}=\binom{z+\phi_{j}(\theta) w}{\phi_{-j}(\theta) z+w}, \quad\binom{z}{w}=\frac{1}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\binom{u-\phi_{j}(\theta) v}{v-\phi_{-j}(\theta) u} .
$$

The Jacobian equals
$\left|\frac{d(z, w)}{d(u, v)}\right|=\frac{1}{\left(1-\phi_{j}(\theta) \phi_{-j}(\theta)\right)^{2}}\left|\begin{array}{cr}1 & -\phi_{j}(\theta) \\ -\phi_{-j}(\theta) & 1\end{array}\right|=\frac{\left|1-\phi_{j}(\theta) \phi_{-j}(\theta)\right|}{\left(1-\phi_{j}(\theta) \phi_{-j}(\theta)\right)^{2}}=\frac{1}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}>0$.
For sufficiently large $j,\left|\phi_{k}(\theta) \phi_{-k}(\theta)\right|<1$ for all $k \geq j$ and for relatively small $j$, there is always a compact set of $\theta \in \Theta$ such that $\left|\phi_{j}(\theta) \phi_{-j}(\theta)\right|<1$ since $\phi_{j}(\theta)$ is zero at the true parameter value for $j \neq 0$.

By applying Mean Value Theorem on $f$, we can get

$$
\begin{aligned}
f_{u, v}^{(j)}(u, v)= & \frac{1}{1-\phi_{j}(\theta) \phi_{-j}(\theta)} f\left(\frac{u-\phi_{j}(\theta) v}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) f\left(\frac{v-\phi_{-j}(\theta) u}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) \\
= & \left(1+O\left(\phi_{j}(\theta) \phi_{-j}(\theta)\right)\right) f\left(\frac{u-\phi_{j}(\theta) v}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) f\left(\frac{v-\phi_{-j}(\theta) u}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) \\
= & f(u) f(v)\left(1+O\left(\phi_{j}(\theta) \phi_{-j}(\theta)\right)\right) \\
& +O\left(u \phi_{-j}(\theta) f\left(\frac{u-\phi_{j}(\theta) v}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) \dot{f}\left(\frac{v-\phi_{-j}(\theta) u \eta}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right)\right) \\
& +O\left(v \phi_{j}(\theta) f\left(\frac{v-\phi_{-j} u}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) \dot{f}\left(\frac{u-\phi_{j}(\theta) v \eta}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right)\right)
\end{aligned}
$$

for $\eta \in(0,1)$. Under Assumption 2.2 with $a=1$,

$$
F_{u, v}^{(j)}(x, y)=F(x) F(y)+O\left(\phi_{j}(\theta)+\phi_{-j}(\theta)\right)
$$

uniformly in $(x, y) \in \mathbb{R}^{2}$.
Before proceeding to the next step, we define truncated versions of $u_{t, m}^{(0, j)}(\theta)$ and $u_{t-j, m}^{(0)}(\theta)$ for some $j / 2 \leq m<j$,

$$
\begin{aligned}
u_{t}^{(0, j)}(\theta) & =u_{t, m}^{(0, j)}(\theta)+\varepsilon_{t, m}^{(0, j)}(\theta) \\
u_{t-j}^{(0,-j)}(\theta) & =u_{t-j, m}^{(0, j)}(\theta)+\varepsilon_{t-j, m}^{(0,-j)}(\theta)
\end{aligned}
$$

where, the truncated residuals

$$
\begin{gathered}
u_{t, m}^{(0, j)}(\theta)=u_{t, m}^{(0)}(\theta)=\sum_{k=-m}^{m} \phi_{k}(\theta) u_{t-k}-u_{t} \\
u_{t-j, m}^{(0,-j)}(\theta)=u_{t-j, m}^{(0)}(\theta)=\sum_{k=-m}^{m} \phi_{k}(\theta) u_{t-j-k}-u_{t-j},
\end{gathered}
$$

and for each $\theta \in \Theta$ and $\mu_{0}>1$, the truncation error

$$
\begin{aligned}
\mathbb{E}\left|\varepsilon_{t, m}^{(0, j)}(\theta)\right| \leq \mathbb{E}\left|u_{t}\right|\left(\sum_{k=m+1}^{\infty}\left|\phi_{k}(\theta)\right|+\sum_{k=-\infty}^{-m-1}\left|\phi_{k}(\theta)\right|\right) \leq C(m+1)^{1-\mu_{0}}<C j^{1-\mu_{0}} \\
\mathbb{E}\left|\varepsilon_{t-j, m}^{(0,-j)}(\theta)\right| \leq \mathbb{E}\left|u_{t-j}\right|\left(\sum_{k=m+1}^{\infty}\left|\phi_{k}(\theta)\right|+\sum_{k=-\infty}^{-m-1}\left|\phi_{k}(\theta)\right|\right) \leq C(m+1)^{1-\mu_{0}}<C j^{1-\mu_{0}} .
\end{aligned}
$$

Using this representation,

$$
\begin{aligned}
& \mathbb{E}\left(F_{u, v}^{(j)}\left(x-u_{t}^{(0, j)}(\theta), y-u_{t-j}^{(0,-j)}(\theta)\right)\right) \\
= & \mathbb{E}\left(F\left(x-u_{t}^{(0, j)}(\theta)\right) F\left(y-u_{t-j}^{(0,-j)}(\theta)\right)+O\left(\phi_{j}(\theta)+\phi_{-j}(\theta)\right)\right) \\
= & \mathbb{E}\left(F\left(x-u_{t, m}^{(0, j)}(\theta)\right) F\left(y-u_{t-j, m}^{(0, j)}(\theta)\right)\right)+\eta_{1} \mathbb{E}\left|\varepsilon_{t, m}^{(0, j)}(\theta) F\left(y-u_{t-j}^{(0,-j)}(\theta)\right)\right| \\
& +\eta_{2} \mathbb{E}\left|\varepsilon_{t-j, m}^{(0,-j)}(\theta) F\left(x-u_{t}^{(0, j)}(\theta)\right)\right|+O\left(\phi_{j}(\theta)+\phi_{-j}(\theta)\right) \quad \text { where } \eta_{1}, \eta_{2} \in(0,1) \\
= & \mathbb{E}\left(F\left(x-u_{t, m}^{(0, j)}(\theta)\right) F\left(y-u_{t-j, m}^{(0,-j)}(\theta)\right)\right)+O\left(j^{1-\mu_{0}}\right) \\
= & \mathbb{E}\left(F\left(x-u_{t, m}^{(0, j)}(\theta)\right)\right) \mathbb{E}\left(F\left(y-u_{t-j, m}^{(0,-j)}(\theta)\right)\right)+O\left(j^{1-\mu_{0}}\right) \\
= & \mathbb{E}\left(F\left(x-u_{t}^{(0, j)}(\theta)\right)\right) \mathbb{E}\left(F\left(y-u_{t-j}^{(0,-j)}(\theta)\right)\right)+O\left(j^{1-\mu_{0}}\right) \\
= & F_{\theta}(x) F_{\theta}(y)+O\left(j^{1-\mu_{0}}\right) .
\end{aligned}
$$

The second equality follows immediately from the application of MVT and the fourth one comes from independence after designed truncation on the residuals. The last second equality is induced from redoing truncation and filling the gaps at lag $j$ and leads $-j$. Therefore, we conclude uniformly in $\theta$ and $(x, y) \in \mathbb{R}^{2}$

$$
\left|\sigma_{\theta, j}(x, y)\right| \leq O\left(j^{1-\mu_{0}}\right)
$$

### 2.7.2. Proof of Lemma 2.3.1

To analyze $\operatorname{Var}\left(\hat{\sigma}_{\theta, j}(x, y)\right)$, we need to compute the most difficult contribution,
$\operatorname{Var}\left(\hat{F}_{\theta, j}(x, y)\right)=\frac{1}{(T-j)^{2}} \sum_{t=1+j}^{T} \sum_{t^{\prime}+j}^{T} \operatorname{Cov}\left(I\left(u_{t-j}(\theta) \leq y\right) I\left(u_{t}(\theta) \leq x\right), I\left(u_{t^{\prime}-j}(\theta) \leq y\right) I\left(u_{t^{\prime}}(\theta) \leq x\right)\right)$,
where the covariance is equal to

$$
\begin{aligned}
& \mathbb{E}\left(I\left(u_{t^{\prime}}(\theta) \leq x\right) I\left(u_{t^{\prime}-j}(\theta) \leq y\right) I\left(u_{t}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right)\right) \\
& -\mathbb{E}\left(I\left(u_{t^{\prime}}(\theta) \leq x\right) I\left(u_{t^{\prime}-j}(\theta) \leq y\right)\right) \mathbb{E}\left(I\left(u_{t}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right)\right)
\end{aligned}
$$

We start with the joint moment of 4 indicators involved in the computation of the covariance in a similar manner to Lemma 2.2.1.

Assume $t>t^{\prime}, j>0$ and $t-t^{\prime}>2 j$, so $t-t^{\prime}-j>\left(t-t^{\prime}\right) / 2$,

$$
\begin{aligned}
& \mathbb{E}\left(I\left(u_{t^{\prime}-j}(\theta) \leq y\right) I\left(u_{t^{\prime}}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right) I\left(u_{t}(\theta) \leq x\right)\right) \\
&= \mathbb{E}\left(\begin{array}{c}
I\left(z_{1}+\phi_{-j}(\theta) z_{2}+\phi_{t^{\prime}-t}(\theta) z_{3}+\phi_{t^{\prime}-t-j}(\theta) z_{4} \leq y-u_{t^{\prime}-j}^{\left(0,-j, t^{\prime}-t-j, t^{\prime}-t\right)}(\theta)\right) \\
I\left(\phi_{j}(\theta) z_{1}+z_{2}+\phi_{t^{\prime}-t+j}(\theta) z_{3}+\phi_{t^{\prime}-t}(\theta) z_{4} \leq x-u_{\left.t^{\prime}, j, t^{\prime}-t+j, t^{\prime}-t\right)}(\theta)\right) \\
I\left(\phi_{t-t^{\prime}}(\theta) z_{1}+\phi_{t-t^{\prime}-j}(\theta) z_{2}+z_{3}+\phi_{-j}(\theta) z_{4} \leq y-u_{t-j}^{\left(0,-j, t-t^{\prime}-j, t-t^{\prime}\right)}(\theta)\right) \\
I\left(\phi_{t-t^{\prime}+j}(\theta) z_{1}+\phi_{t-t^{\prime}}(\theta) z_{2}+\phi_{j}(\theta) z_{3}+z_{4} \leq x-u_{t}^{\left(0, j, t-t^{\prime}+j, t-t^{\prime}\right)}(\theta)\right)
\end{array}\right) f\left(z_{1}\right) f\left(z_{2}\right) f\left(z_{3}\right) f\left(z_{4}\right) \\
& d z_{1} d z_{2} d z_{3} d z_{4} \\
&= \mathbb{E}\left(\int\binom{I\left(u_{1} \leq y-u_{t^{\prime}-j}^{\left(0,-j, t^{\prime}-t-j, t^{\prime}-t\right)}(\theta)\right) I\left(u_{2} \leq x-u_{t}^{\left(0, j, t^{\prime}-t+j, t^{\prime}-t\right)}(\theta)\right)}{I\left(u_{3} \leq y-u_{t-j}^{\left(0,-j, t-t^{\prime}-j, t-t^{\prime}\right)}(\theta)\right) I\left(u_{4} \leq x-u_{t}^{\left(0, j, t-t^{\prime}+j, t-t^{\prime}\right)}(\theta)\right)} f_{u}(u) d u\right) \\
&= \mathbb{E}\left(F _ { u } \left(y-u_{\left.t^{\prime}-j, t-t-t^{\prime}-t\right)}^{\left.\left(0,-j, t^{\prime}-t-j, x-u_{t}^{\left(0, j, t^{\prime}-t+j, t^{\prime}-t\right)}(\theta), y-u_{t-j}^{\left(0,-j, t-t^{\prime}-j, t-t^{\prime}\right)}(\theta), x-u_{t}^{\left(0, j, t-t^{\prime}+j, t-t^{\prime}\right)}(\theta)\right)\right)}\right.\right.
\end{aligned}
$$

where

$$
\begin{aligned}
u & =\left(\begin{array}{l}
u_{1} \\
u_{2} \\
u_{3} \\
u_{4}
\end{array}\right)=\left(\begin{array}{ccc}
z_{1}+\phi_{-j}(\theta) z_{2}+\phi_{t^{\prime}-t}(\theta) z_{3}+\phi_{t^{\prime}-t-j}(\theta) z_{4} \\
\phi_{j}(\theta) z_{1}+z_{2}+\phi_{t^{\prime}-t+j}(\theta) z_{3}+\phi_{t^{\prime}-t}(\theta) z_{4} \\
\phi_{t-t^{\prime}}(\theta) z_{1}+\phi_{t-t^{\prime}-j}(\theta) z_{2}+z_{3}+\phi_{-j}(\theta) z_{4} \\
\phi_{t-t^{\prime}+j}(\theta) z_{1}+\phi_{t-t^{\prime}}(\theta) z_{2}+\phi_{j}(\theta) z_{3}+z_{4}
\end{array}\right) \\
& =\left(\begin{array}{cccc}
1 & \phi_{-j}(\theta) & \phi_{t^{\prime}-t}(\theta) & \phi_{t^{\prime}-t-j}(\theta) \\
\phi_{j}(\theta) & 1 & \phi_{t^{\prime}-t+j}(\theta) & \phi_{t^{\prime}-t}(\theta) \\
\phi_{t-t^{\prime}}(\theta) & \phi_{t-t^{\prime}-j}(\theta) & 1 & \phi_{-j}(\theta) \\
\phi_{t-t^{\prime}+j}(\theta) & \phi_{t-t^{\prime}}(\theta) & \phi_{j}(\theta) & 1
\end{array}\right)\left(\begin{array}{l}
z_{1} \\
z_{2} \\
z_{3} \\
z_{4}
\end{array}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
& \left(\begin{array}{cccc}
1 & \phi_{-j}(\theta) & \phi_{t^{\prime}-t}(\theta) & \phi_{t^{\prime}-t-j}(\theta) \\
\phi_{j}(\theta) & 1 & \phi_{t^{\prime}-t+j}(\theta) & \phi_{t^{\prime}-t}(\theta) \\
\phi_{t-t^{\prime}}(\theta) & \phi_{t-t^{\prime}-j}(\theta) & 1 & \phi_{-j}(\theta) \\
\phi_{t-t^{\prime}+j}(\theta) & \phi_{t-t^{\prime}}(\theta) & \phi_{j}(\theta) & 1
\end{array}\right)^{-1} \\
& =\left(\begin{array}{cc}
\left(\begin{array}{cc}
1 & \phi_{-j}(\theta) \\
\phi_{j}(\theta) & 1
\end{array}\right)^{-1} & \left(\begin{array}{ll}
0 & 0 \\
0 & 0
\end{array}\right) \\
\left(\begin{array}{ll}
0 & 0 \\
0 & 0
\end{array}\right)
\end{array}\left(\begin{array}{cc}
1 & \phi_{-j}(\theta) \\
\phi_{j}(\theta) & 1
\end{array}\right)^{-1}\right)\left(1+O\left(\phi_{t-t^{\prime}}(\theta)+\phi_{t^{\prime}-t}(\theta)\right)\right) \\
& =\frac{1}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\left(\begin{array}{cccc}
1 & -\phi_{j}(\theta) & 0 & 0 \\
-\phi_{-j}(\theta) & 1 & 0 & 0 \\
0 & 0 & 1 & -\phi_{j}(\theta) \\
0 & 0 & -\phi_{-j}(\theta) & 1
\end{array}\right)\left(1+O\left(\phi_{t-t^{\prime}}(\theta)+\phi_{t^{\prime}-t}(\theta)\right)\right) \text {. }
\end{aligned}
$$

Hence the Jacobian is

$$
\frac{1}{\left(1-\phi_{j}(\theta) \phi_{-j}(\theta)\right)^{2}}\left(1+O\left(\phi_{t-t^{\prime}}(\theta)+\phi_{t^{\prime}-t}(\theta)\right)\right)
$$

and by applying MVT to the argument of each $f$ in the linear mappings $z_{a}=$ $z_{a}(u), a=1,2,3,4$, and only keeping components involved with $1, \phi_{j}(\theta), \phi_{-j}(\theta)$,

$$
\begin{aligned}
& f_{u}(u) \\
= & \frac{1}{\left(1-\phi_{j}(\theta) \phi_{-j}(\theta)\right)^{2}} f\left(\frac{u_{1}-\phi_{j} u_{2}}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) f\left(\frac{-\phi_{j} u_{1}+u_{2}}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) f\left(\frac{u_{3}-\phi_{j} u_{4}}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) \\
& \times f\left(\frac{-\phi_{j} u_{3}+u_{4}}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) \times\left(1+O\left(\phi_{t-t^{\prime}}(\theta)+\phi_{t^{\prime}-t}(\theta)\right)\right)+g(u) O\left(\phi_{t-t^{\prime}}(\theta)+\phi_{t^{\prime}-t}(\theta)\right) \\
= & f_{12}\left(u_{1}, u_{2}\right) f_{34}\left(u_{3}, u_{4}\right)+g(u) O\left(\phi_{t-t^{\prime}}(\theta)+\phi_{t^{\prime}-t}(\theta)\right)
\end{aligned}
$$

where

$$
f_{12}\left(u_{1}, u_{2}\right)=\frac{1}{1-\phi_{j}(\theta) \phi_{-j}(\theta)} f\left(\frac{u_{1}-\phi_{j} u_{2}}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right) f\left(\frac{-\phi_{j} u_{1}+u_{2}}{1-\phi_{j}(\theta) \phi_{-j}(\theta)}\right)
$$

$f_{12}=f_{34}$ due to stationarity and $g(u)$ is integrable in $u \in \mathbb{R}^{4}$ under Assumption 2.2 with $a=2$.

By integrating the joint density over $u$, we get

$$
F_{u}(x)=F_{12}\left(x_{1}, x_{2}\right) F_{34}\left(x_{3}, x_{4}\right)+O\left(\phi_{t-t^{\prime}}(\theta)+\phi_{t^{\prime}-t}(\theta)\right)
$$

uniformly in $x=\left(x_{1}, x_{2}, x_{3}, x_{4}\right) \in \mathbb{R}^{4}$.
Therefore,

$$
\begin{aligned}
& \mathbb{E}\left(I\left(u_{t^{\prime}-j}(\theta) \leq y\right) I\left(u_{t^{\prime}}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right) I\left(u_{t}(\theta) \leq x\right)\right) \\
= & \mathbb{E}\binom{F_{12}\left(x-u_{t^{\prime}}^{\left(0, j, t^{\prime}-t, t^{\prime}-t+j\right)}(\theta), y-u^{\left(0,-j, t^{\prime}-t, t^{\prime}-t-j\right)}(\theta)\right)}{F_{34}\left(x-u_{t}^{\left(0, j, t-t, t, t-t^{\prime}+j\right)}(\theta), y-u_{t-j}^{\left(0,-j, t-t^{\prime}, t-t^{\prime}-j\right)}(\theta)\right)}+O\left(\phi_{t-t^{\prime}}(\theta)+\phi_{t^{\prime}-t}(\theta)\right) \\
= & \mathbb{E}\left(F_{12}\left(x-u_{t}^{(0, j)}(\theta), y-u_{t-j}^{(0,-j)}(\theta)\right)\right)^{2}+O\left(\left|t-t^{\prime}\right|^{1-\mu_{0}}\right) \\
= & F_{\theta, j}(x, y)^{2}+O\left(\left|t-t^{\prime}\right|^{1-\mu_{0}}\right) .
\end{aligned}
$$

The second last equality comes from truncation, independence, and refilling the truncation at $m=\left[\left(t-t^{\prime}\right) / 4\right]$, the same techniques used in the previous proof.
Then,

$$
\begin{aligned}
& \mathbb{E}\left|\hat{F}_{\theta, j}(x, y)-F_{\theta, j}(x, y)\right|^{2} \\
\leq & C\left(1 \wedge \frac{(T-j) j}{(T-j)^{2}}\right)+C \frac{1}{(T-j)^{2}} \sum_{t=j+1}^{T} \sum_{t^{\prime}\left|t-t^{\prime}\right|>2 j}\left|t-t^{\prime}\right|^{1-\mu_{0}} \\
\leq & C\left(1 \wedge \frac{j}{(T-j)}\right)+C \frac{j^{2-\mu_{0}}}{T-j}+C \frac{\log T}{(T-j)^{\mu_{0}-1}} .
\end{aligned}
$$

The summation can be divided into two parts. The first component in the first
inequality comes from the summand of $\left|t-t^{\prime}\right| \leq 2 j$, which consists of at maximum $4 j(T-j)$ terms of covariance uniformly bounded by 1 . For the summand with $\left|t-t^{\prime}\right|>2 j$, each term is a $O\left(\left|t-t^{\prime}\right|^{1-\mu_{0}}\right)$. The summation $\sum_{t^{\prime}:\left|t-t^{\prime}\right|>2 j}\left|t-t^{\prime}\right|^{1-\mu_{0}}$ is bounded by an integral $\int_{s>2 j} s^{1-\mu_{0}} d s$.

### 2.7.3. Truncation effect

The representation of the truncated sequences First, we show that the truncation series follows from a similar representation of a residual sequence defined in (2.2.3) with the same condition as Assumption 2.1 satisfied. Say,

$$
\begin{aligned}
\hat{u}_{t}(\theta) & =u_{t}(\theta)-\left(\sum_{j=-\infty}^{t-T-1}+\sum_{j=t}^{\infty}\right) \psi_{j}^{(-1)}(\theta) Y_{t-j} \\
& =\sum_{k=-\infty}^{\infty} \psi_{k}^{(-1)}(\theta) Y_{t-j}-\left(\sum_{j=-\infty}^{t-T-1}+\sum_{j=t}^{\infty}\right) \psi_{j}^{(-1)}(\theta) \sum_{k=-\infty}^{\infty} \psi_{j}\left(\theta_{0}\right) u_{t-j-k} \\
& =\sum_{k=-\infty}^{\infty} \phi_{k}(\theta) u_{t-k}-\sum_{k=-\infty}^{\infty}\left\{\left(\sum_{j=-\infty}^{t-T-1}+\sum_{j=t}^{\infty}\right) \psi_{j}^{(-1)}(\theta) \psi_{k-j}\left(\theta_{0}\right)\right\} u_{t-k} \\
& =\sum_{k=-\infty}^{\infty}\left\{\phi_{k}(\theta)-\left(\sum_{j=-\infty}^{t-T-1}+\sum_{j=t}^{\infty}\right) \psi_{j}^{(-1)}(\theta) \psi_{k-j}\left(\theta_{0}\right)\right\} u_{t-k} \\
& =\sum_{k=-\infty}^{\infty}\left\{\sum_{j=t-T}^{t-1} \psi_{j}^{(-1)}(\theta) \psi_{k-j}\left(\theta_{0}\right)\right\} u_{t-k} \\
& =\sum_{k=-\infty}^{\infty} \phi_{k}^{t, T}(\theta) u_{t-k},
\end{aligned}
$$

where

$$
\phi_{k}(\theta)=\sum_{j=-\infty}^{\infty} \psi_{j}^{(-1)}(\theta) \psi_{k-j}\left(\theta_{0}\right), \quad \phi_{k}^{(t, T)}(\theta)=\sum_{j=t-T}^{t-1} \psi_{j}^{(-1)}(\theta) \psi_{k-j}\left(\theta_{0}\right)
$$

Consider $k>0$, then

$$
\begin{aligned}
& t<k / 2 \rightarrow\left|\left(\sum_{j=-\infty}^{t-T-1}+\sum_{j=t}^{\infty}\right) \psi_{j}^{(-1)}(\theta) \psi_{k-j}\left(\theta_{0}\right)\right| \leq C\left|\psi_{k}(\theta)\right| \\
& k / 2<t<3 k / 2 \rightarrow\left|\left(\sum_{j=-\infty}^{t-T-1}+\sum_{j=t}^{\infty}\right) \psi_{j}^{(-1)}(\theta) \psi_{k-j}\left(\theta_{0}\right)\right| \leq C\left|\psi_{k}^{(-1)}(\theta)\right| \\
& 3 k / 2<t \rightarrow\left|\left(\sum_{j=-\infty}^{t-T-1}+\sum_{j=t}^{\infty}\right) \psi_{j}^{(-1)}(\theta) \psi_{k-j}\left(\theta_{0}\right)\right| \leq C\left|\psi_{k}^{(-1)}(\theta)\right|,
\end{aligned}
$$

while for the case of $k<0$, similar bonds can be obtained by comparing with $t-T<0$. Then, for all $|k|$ large enough, we have

$$
\sup _{\theta \in \Theta}\left|\phi_{k}^{(t, T)}(\theta)\right| \leq C|k|^{-\mu_{0}}
$$

uniformly in $t=1,2, \ldots, T$. Thus, $\hat{u}_{t}(\theta)$ has a similar representation with $u_{t}(\theta)$ in terms of $\phi_{k}^{(t, T)}(\theta)$.

Asymptotic truncation effect Based on the representation of the truncated sequence, we can check that

$$
\begin{aligned}
& \mathbb{E}\left|\hat{F}_{\theta, j}(x, y)-\bar{F}_{\theta, j}(x, y)\right|^{2} \\
= & \frac{1}{(T-j)^{2}} \sum_{t=j+1}^{T} \sum_{t^{\prime}=j+1}^{T} \mathbb{E}\left[I\left(u_{t}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right) I\left(u_{t^{\prime}}(\theta) \leq x\right) I\left(u_{t^{\prime}-j}(\theta) \leq y\right)\right] \\
& -\frac{2}{(T-j)^{2}} \sum_{t=j+1}^{T} \sum_{t^{\prime}=j+1}^{T} \mathbb{E}\left[I\left(u_{t}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right) I\left(\hat{u}_{t^{\prime}}(\theta) \leq x\right) I\left(\hat{u}_{t^{\prime}-j}(\theta) \leq y\right)\right] \\
& +\frac{1}{(T-j)^{2}} \sum_{t=j+1}^{T} \sum_{t^{\prime}=j+1}^{T} \mathbb{E}\left[I\left(\hat{u}_{t}(\theta) \leq x\right) I\left(\hat{u}_{t-j}(\theta) \leq y\right) I\left(\hat{u}_{t^{\prime}}(\theta) \leq x\right) I\left(\hat{u}_{t^{\prime}-j}(\theta) \leq y\right)\right]
\end{aligned}
$$

where $\bar{F}_{\theta, j}(x, y)=\frac{1}{T-j} \sum_{t=j+1}^{T} I\left(\hat{u}_{t}(\theta) \leq x\right) I\left(\hat{u}_{t-j}(\theta) \leq y\right)$ and each expectation can be treated as before in terms of $\phi_{k}(\theta)$ and $\phi_{k}^{(t, T)}(\theta)$. Once the smooth representation in terms of the $c d f F$ is obtained, we can bound the expectation with the assistance of MVT

$$
\mathbb{E} \sup _{\theta \in \Theta}\left|u_{t}(\theta)-\hat{u}_{t}(\theta)\right|^{2} \leq C\left(t^{1-\mu_{0}}+(T+1-t)^{1-\mu_{0}}\right)
$$

then for each $\theta \in \Theta, \mu_{0}>2$, we have

$$
\mathbb{E}\left|\hat{F}_{\theta, j}(x, y)-\bar{F}_{\theta, j}(x, y)\right|^{2} \leq C(T-j)^{-1}
$$

where we can conclude the effect of the truncation due to the finite sample observations can be asymptotically negligible.

### 2.7.4. Proof of Theorem 2.3.2

First we need to show pointwise convergence of $\hat{\mathcal{Q}}_{T}(\theta)$ to $\mathcal{Q}_{0}(\theta)$ for each $\theta \in \Theta$, i.e,

$$
\hat{\mathcal{Q}}_{T}(\theta)-\mathcal{Q}_{0}(\theta)=o_{p}(1) \text { for each } \theta \in \Theta
$$

We first approximate the population loss function by

$$
\mathcal{Q}_{T}(\theta)=\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{(j \pi)^{2}} \sigma_{\theta, j}^{2}(x, y) d W(x, y)
$$

so that,

$$
\begin{aligned}
\left|\mathcal{Q}_{0}(\theta)-\mathcal{Q}_{T}(\theta)\right| & =\left|\sum_{j=T}^{\infty} \int_{\mathbb{R}^{2}} \frac{1}{(j \pi)^{2}} \sigma_{\theta, j}^{2}(x, y) d W(x, y)+\sum_{j=1}^{T-1} \int_{\mathbb{R}^{2}} \frac{j}{T} \frac{1}{(j \pi)^{2}} \sigma_{\theta, j}^{2}(x, y) d W(x, y)\right| \\
& \leq\left|\sum_{j=T}^{\infty} \int_{\mathbb{R}^{2}} \frac{1}{(j \pi)^{2}} \sigma_{\theta, j}^{2}(x, y) d W(x, y)\right|+\left|\frac{1}{\pi^{2}} \sum_{j=1}^{T-1} \int_{\mathbb{R}^{2}} \frac{1}{j T} \sigma_{\theta, j}^{2}(x, y) d W(x, y)\right| .
\end{aligned}
$$

Note that $\left|\sigma_{\theta, j}(x, y)\right|$ is uniformly bounded by 1 in $(\theta, x, y) \in \Theta \times \mathbb{R}^{2}$ for each $j=1,2, \ldots T$. Hence,

$$
\begin{aligned}
\sup _{\theta \in \Theta}\left|\mathcal{Q}_{0}(\theta)-\mathcal{Q}_{T}(\theta)\right| \leq & \frac{1}{\pi^{2}} \sum_{j=T}^{\infty} \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \sup _{(\theta, x, y) \in \Theta \times \mathbb{R}^{2}}\left|\sigma_{\theta, j}(x, y)\right|^{2} d W(x, y) \\
& +\frac{1}{T \pi^{2}} \sum_{j=1}^{T-1} \frac{1}{j} \int_{\mathbb{R}^{2}} \sup _{(\theta, x, y) \in \Theta \times \mathbb{R}^{2}}\left|\sigma_{\theta, j}(x, y)\right|^{2} d W(x, y) \\
\leq & C \sum_{j=T}^{\infty} \frac{1}{j^{2}}+\frac{C}{T} \sum_{j=1}^{T-1} \frac{1}{j} \\
\leq & \frac{C}{T}+\frac{C \ln (T-1)}{T}=o(1)
\end{aligned}
$$

From the above statement, showing $\hat{\mathcal{Q}}_{T}(\theta)-\mathcal{Q}_{0}(\theta)=o_{p}(1)$ is equivalent to showing

$$
\begin{equation*}
\left|\hat{\mathcal{Q}}_{T}(\theta)-\mathcal{Q}_{T}(\theta)\right|=o_{p}(1) \tag{2.7.1}
\end{equation*}
$$

First we define $z_{t}(\theta, x):=I\left(u_{t}(\theta) \leq x\right)-F_{\theta}(x)$, which is a mean zero sequence, and

$$
\bar{\sigma}_{\theta, j}(x, y)=\frac{1}{T-j} \sum_{t=j+1}^{T} z_{t}(\theta, x) z_{t-j}(\theta, y),
$$

where

$$
\begin{aligned}
& \hat{\sigma}_{\theta, j}^{2}(x, y)-\sigma_{\theta, j}^{2}(x, y)=\hat{\sigma}_{\theta, j}^{2}(x, y)-\bar{\sigma}_{\theta, j}^{2}(x, y)+\bar{\sigma}_{\theta, j}^{2}(x, y)-\sigma_{\theta, j}^{2}(x, y) \\
& \hat{\sigma}_{\theta, j}^{2}(x, y)-\bar{\sigma}_{\theta, j}^{2}(x, y)=\left|\hat{\sigma}_{\theta, j}(x, y)-\bar{\sigma}_{\theta, j}(x, y)\right|^{2}+2\left(\hat{\sigma}_{\theta}(x, y)-\bar{\sigma}_{\theta, j}(x, y)\right) \bar{\sigma}_{\theta, j}(x, y) .
\end{aligned}
$$

By Cauchy-Schwarz inequality,

$$
\begin{aligned}
& \mathbb{E}\left|\left(\hat{\sigma}_{\theta, j}(x, y)-\bar{\sigma}_{\theta, j}(x, y)\right) \bar{\sigma}_{\theta, j}(x, y)\right| \\
\leq & \left\{\mathbb{E}\left|\left(\hat{\sigma}_{\theta, j}(x, y)-\bar{\sigma}_{\theta, j}(x, y)\right)\right|^{2} \mathbb{E}\left|\bar{\sigma}_{\theta, j}(x, y)\right|^{2}\right\}^{1 / 2} \\
\leq & C(T-j)^{-1}
\end{aligned}
$$

The last inequality comes from $\left|\bar{\sigma}_{\theta, j}(x, y)\right| \leq 1$ and

$$
\begin{aligned}
(T-j)^{4} \mathbb{E}\left|\hat{\sigma}_{\theta, j}(x, y)-\bar{\sigma}_{\theta, j}(x, y)\right|^{2} & \leq\left\{\mathbb{E}\left|\sum_{t=j+1}^{T} z_{t}(\theta, x)\right|^{4} \mathbb{E}\left|\sum_{t=j+1}^{T} z_{t-j}(\theta, y)\right|^{4}\right\}^{1 / 2} \\
& \leq C(T-j)^{2}
\end{aligned}
$$

The proof of the above result uses the Marcinkiewicz-Zygmund inequality in Theorem 1 of Doukhan and Louhichi (1999) for a sequence of weakly dependent random variables $\left\{X_{n}\right\}_{n \in \mathbb{N}}$ centered at expectation, for which

$$
C_{r, q}:=\left|\operatorname{Cov}\left(X_{t_{1}} \cdots X_{t_{m}}, X_{t_{m+1}} \cdots X_{t_{q}}\right)\right|=O\left(r^{-q / 2}\right) \quad \text { as } r \rightarrow \infty
$$

for some fixed $q \in N, q \geq 2$ and the sup is taken over all $\left\{t_{1}, t_{2}, \ldots, t_{q}\right\}$ such that $1 \leq t_{1} \leq \cdots \leq t_{q}$ and $m, r$ satisfy $t_{m+1}-t_{m}=r$, there exists a positive constant $C$ independent on $n$ for which

$$
\left|\mathbb{E}\left(\sum_{n=1}^{N} X_{n}\right)^{q}\right| \leq C N^{q / 2}
$$

In our context, we apply this result for $X_{n}=\left\{I\left(u_{n}(\theta) \leq x\right)-F_{\theta}(x)\right\}$ for fixed $x$ and $q=4$, considering $m=1,2$.
Similarly,
$\left|\bar{\sigma}_{\theta, j}^{2}(x, y)-\sigma_{\theta, j}^{2}(x, y)\right|=\left|\bar{\sigma}_{\theta, j}(x, y)-\sigma_{\theta, j}(x, y)\right|^{2}+2\left(\bar{\sigma}_{\theta, j}(x, y)-\sigma_{\theta, j}(x, y)\right) \sigma_{\theta, j}(x, y)$
so that

$$
\begin{aligned}
& \quad \mathbb{E}\left|\bar{\sigma}_{\theta, j}^{2}(x, y)-\sigma_{\theta, j}^{2}(x, y)\right| \\
& \leq \mathbb{E}\left|\bar{\sigma}_{\theta, j}(x, y)-\sigma_{\theta, j}(x, y)\right|^{2}+2\left\{\mathbb{E}\left|\bar{\sigma}_{\theta, j}(x, y)-\sigma_{\theta, j}(x, y)\right|^{2} \mathbb{E}\left|\sigma_{\theta, j}(x, y)\right|^{2}\right\}^{1 / 2} \\
& \leq \\
& \leq C\left(1 \wedge \frac{j}{T-j}\right)+C \frac{j^{2-\mu_{0}}}{T-j}+C \frac{\log T}{(T-j)^{\mu_{0}-1}} \\
& \quad+C\left(1 \wedge \frac{j}{T-j}\right)^{1 / 2}+C \frac{j^{1-\mu_{0} / 2}}{(T-j)^{1 / 2}}+C \frac{(\log T)^{1 / 2}}{(T-j)^{\mu_{0} / 2-1 / 2}}
\end{aligned}
$$

by Lemma 2.3.1. Hence,

$$
\begin{aligned}
\mathbb{E}\left|\hat{\mathcal{Q}}_{T}(\theta)-\mathcal{Q}_{T}(\theta)\right| & =\mathbb{E}\left|\sum_{j=1}^{T-1} \int_{\mathbb{R}^{2}}\left(\hat{\sigma}_{\theta, j}^{2}(x, y)-\sigma_{\theta, j}^{2}(x, y)\right)\left(1-\frac{j}{T}\right) \frac{1}{(j \pi)^{2}} d W(x, y)\right| \\
\leq & \frac{1}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}}\left\{\mathbb{E}\left|\hat{\sigma}_{\theta, j}^{2}(x, y)-\bar{\sigma}_{\theta, j}^{2}(x, y)\right|+\mathbb{E}\left|\bar{\sigma}_{\theta, j}^{2}(x, y)-\sigma_{\theta, j}^{2}(x, y)\right|\right\} d W(x, y) \\
\leq & \frac{1}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \mathbb{E}\left|\bar{\sigma}_{\theta, j}^{2}(x, y)-\sigma_{\theta, j}^{2}(x, y)\right| d W(x, y) \\
& +\frac{1}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \mathbb{E}\left|\hat{\sigma}_{\theta, j}(x, y)-\bar{\sigma}_{\theta, j}(x, y)\right|^{2} d W(x, y) \\
& +\frac{2}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \mathbb{E}\left|\left(\hat{\sigma}_{\theta, j}(x, y)-\bar{\sigma}_{\theta, j}(x, y)\right) \bar{\sigma}_{\theta, j}(x, y)\right| d W(x, y) \\
= & A+B+C .
\end{aligned}
$$

For some $\mu_{0}>1$,

$$
\begin{aligned}
A \leq & \frac{C}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}}\left(\left(1 \wedge \frac{j}{T-j}\right)+\frac{j^{2-\mu_{0}}}{T-j}+\frac{\log T}{(T-j)^{\mu_{0}-1}}\right) d W(x, y) \\
& +\frac{C}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}}\left(\left(1 \wedge \frac{j}{T-j}\right)^{1 / 2}+\frac{j^{1-\mu_{0} / 2}}{(T-j)^{1 / 2}}+\frac{(\log T)^{1 / 2}}{(T-j)^{\mu_{0} / 2-1 / 2}}\right) d W(x, y) \\
\leq & \frac{C}{T} \sum_{j=1}^{T-1} \frac{1}{j^{2}}((T-j) \wedge j)+\frac{C}{T} \sum_{j=1}^{T-1}\left(j^{-\mu_{0}}+\frac{1}{j^{2}}(\log T)(T-j)^{2-\mu_{0}}\right) \\
& +\frac{C}{T} \sum_{j=1}^{T-1} \frac{1}{j^{2}}\left((T-j)^{1 / 2} \wedge(j(T-j))^{1 / 2}\right)+\frac{C}{T} \sum_{j=1}^{T-1}(T-j)^{1 / 2} j^{-1-\mu_{0} / 2} \\
& +\frac{C}{T} \sum_{j=1}^{T-1} \frac{1}{j^{2}}(\log T)^{1 / 2}(T-j)^{1 / 2-\mu_{0} / 2} \\
= & o(1) \text { as } T \rightarrow \infty .
\end{aligned}
$$

Next,

$$
B \leq \frac{C}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}}(T-j)^{-1} \leq \frac{C}{T} \sum_{j=1}^{T-1} \frac{1}{j^{2}} \leq \frac{C}{T}=o(1)
$$

and, similarly for $C$,

$$
C \leq \frac{C}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}}(T-j)^{-1} \leq \frac{C}{T}=o(1)
$$

The proof of pointwise convergence of $\hat{\mathcal{Q}}_{T}(\theta)$ to $\mathcal{Q}_{T}(\theta)$ is completed.
As the result of the non-differentiability of the objective function, we need to show stochastic equicontinuity of $\hat{\mathcal{Q}}_{T}(\theta)$. For the definition of stochastic equicontinuity see Chapter 36 Section 2.7 in Newey and McFadden (1994).

By definition, we shall prove

$$
\sup _{\bar{\theta} \in \mathcal{B}(\theta, \epsilon, \eta)}\left|\hat{\mathcal{Q}}_{T}(\bar{\theta})-\hat{\mathcal{Q}}_{T}(\theta)\right|=\Delta(\epsilon, \eta)=o_{p}(1) \text { for each } \theta \in \Theta \text { and } T \geq T_{0}(\epsilon, \eta)
$$

where $\mathcal{B}(\theta, \epsilon, \eta)$ denotes an open set containing $\theta$ that may depend on $\epsilon$ and $\eta$. Equivalently,

$$
\sup _{\bar{\theta} \in \mathcal{B}(\theta, \epsilon, \eta))}\left|\hat{\mathcal{Q}}_{T}(\bar{\theta})-\hat{\mathcal{Q}}_{T}(\theta)\right|=\sup _{\bar{\theta} \in \mathcal{B}(\theta, \epsilon, \eta)}\left|\hat{\mathcal{Q}}_{T}(\bar{\theta})-\mathcal{Q}_{T}(\bar{\theta})+\mathcal{Q}_{T}(\bar{\theta})-\mathcal{Q}_{T}(\theta)+\mathcal{Q}_{T}(\theta)-\hat{\mathcal{Q}}_{T}(\theta)\right|=o_{p}(1) .
$$

The above statement is implied by

$$
\begin{aligned}
& \sup _{\bar{\theta} \in \mathcal{B}(\theta, \epsilon, \eta)}\left|\hat{\mathcal{Q}}_{T}(\bar{\theta})-\mathcal{Q}_{T}(\bar{\theta})\right|=o_{p}(1) \\
& \sup _{\bar{\theta} \in \mathcal{B}(\theta, \epsilon, \eta)}\left|\mathcal{Q}_{T}(\bar{\theta})-\mathcal{Q}_{T}(\theta)\right|=o(1)
\end{aligned}
$$

where the second equation immediately follows from the continuity of the population function $\mathcal{Q}_{T}(\theta)$ in a compact set $\theta$.
Define a metric $d_{k}$ of $\theta$ on $\Theta$ by the Euclidean norm, and we shall show the process indexed by $\theta \in \Theta$

$$
\begin{equation*}
\mathcal{D}_{T}(\bar{\theta}):=\hat{\mathcal{Q}}_{T}(\bar{\theta})-\mathcal{Q}_{T}(\bar{\theta}) \tag{2.7.2}
\end{equation*}
$$

is stochastically equicontinuous by demonstrating the class of function $\mathcal{D}_{T}(\theta)$ is P Donsker.

Theorem 2.1 of Newey (1991) confirms uniform convergence in the probability of objective function in a compact set of parameter $\Theta$. In addition, the identification of $\theta_{0}$ by the population loss function $\mathcal{Q}_{0}(\theta)$ is guaranteed under Assumptions 2.1, 2.3 and 2.4. Since the global minimum of the non-negative function $\mathcal{Q}_{0}(\theta)$ is uniquely attained at $\theta=\theta_{0}$, and for $\theta \neq \theta_{0}, \mathcal{Q}_{0}(\theta)>0$ contributed by the non-zero term(s) $\sigma_{\theta, j}(x, y)$ due to the pairwise dependence of $u_{t}(\theta)$ and $u_{t-j}(\theta)$ for some $j$ and $(x, y) \in$ $\mathbb{R}^{2}$. Then, the fundamental consistency theorem for extremum estimators implies

$$
\hat{\theta}_{T} \longrightarrow p \theta_{0} \text { as } T \rightarrow \infty
$$

### 2.7.5. Proof of Theorem 2.3.3

The consistency of the estimates based on the smoothed $c d f$ transformation, $\tilde{\theta}_{T}^{h}$, easily follows the steps in the proof of Theorem 3 in Velasco (2022) with some modifications by replacing the characteristic function with the smoothed cdf $\Lambda(\cdot)$ in the objective function $\tilde{\mathcal{Q}}_{T}(\theta ; h)$, the sample counterpart of the population loss function

$$
\begin{equation*}
\mathcal{Q}_{0}(\theta ; h):=\frac{2}{\pi^{2}} \sum_{j=1}^{\infty} \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \sigma_{\theta, j}^{2}(x, y ; h) d W(x, y), \tag{2.7.3}
\end{equation*}
$$

where

$$
\begin{aligned}
\sigma_{\theta, j}(x, y ; h) & :=\mathbb{E}\left[\Lambda\left(\frac{x-u_{t}(\theta)}{h}\right) \Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\right]-\mathbb{E}\left(\Lambda\left(\frac{x-u_{t}(\theta)}{h}\right)\right) \mathbb{E}\left(\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\right) \\
& =F_{\theta, j}(x, y ; h)-F_{\theta, j}(x, \infty ; h) F_{\theta, j}(\infty, y ; h)
\end{aligned}
$$

The bound of $\sigma_{\theta, j}(x, y ; h)$ can be derived in a similar way as Lemma 1 in Velasco (2022), so that,

$$
\sup _{\theta \in \Theta}\left|\sigma_{\theta, j}(x, y ; h)\right| \leq C_{h} j^{1-\mu_{0}}
$$

uniformly in $(x, y) \in \mathbb{R}^{2}$ and $C_{h}$ is a positive number depending on $h$ but bounded due to fixed $h>0$.
Similarly, for $j=1,2, \ldots, T-1$, for $\theta \in \Theta$,

$$
\mathbb{E}\left|\sigma_{\theta, j}(x, y ; h)-\tilde{\sigma}_{\theta, j}(x, y ; h)\right|^{2} \leq C_{h}\left(1 \wedge \frac{j}{T-j}\right)+C \frac{j^{2-\mu_{0}}}{T-j}+C \frac{\log T}{(T-j)^{\mu_{0}-1}}
$$

The pointwise convergence of $\tilde{\mathcal{Q}}_{T}(\theta ; h)$ to $\mathcal{Q}_{0}(\theta ; h)$ is shown with the assistance of the centered variable, $z_{t}(\theta, x ; h):=\Lambda\left(\frac{x-u_{t}(\theta)}{h}\right)-\mathbb{E}\left[\Lambda\left(\frac{x-u_{t}(\theta)}{h}\right)\right]$, and

$$
\bar{\sigma}_{\theta, j}(x, y ; h)=\frac{1}{T-j} \sum_{t=j+1}^{T} z_{t}(\theta, x ; h) z_{t-j}(\theta, y ; h) .
$$

Then, the same approach in the proof of Theorem 2.3.2 can be applied here. This convergence is uniform in $\theta \in \Theta$. This result immediately follows from the uniform boundedness of the derivative of $\tilde{\mathcal{Q}}_{T}(\theta ; h)$ given a fixed $h>0$,

$$
\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}(\theta ; h)=\frac{4}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \tilde{\sigma}_{\theta, j}(x, y ; h) \frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta, j}(x, y ; h) d W(x, y),
$$

where

$$
\begin{aligned}
\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta, j}(x, y ; h) & =\frac{1}{T-j} \sum_{t=j+1}^{T}\left[\lambda\left(\frac{x-u_{t}(\theta)}{h}\right) \Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(-\frac{1}{h}\right) u_{t}^{(1)}(\theta)\right] \\
+ & \frac{1}{T-j} \sum_{t=j+1}^{T}\left[\Lambda\left(\frac{x-u_{t}(\theta)}{h}\right) \lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(-\frac{1}{h}\right) u_{t-j}^{(1)}(\theta)\right] \\
& +\frac{1}{T-j} \sum_{t=j+1}^{T}\left[\lambda\left(\frac{x-u_{t}(\theta)}{h}\right)\left(-\frac{1}{h}\right) u_{t}^{(1)}(\theta)\right] \frac{1}{T-j} \sum_{t=j+1}^{T} \Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right) \\
& +\frac{1}{T-j} \sum_{t=j+1}^{T}\left[\lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)\left(-\frac{1}{h}\right) u_{t-j}^{(1)}(\theta)\right] \frac{1}{T-j} \sum_{t=j+1}^{T} \Lambda\left(\frac{x-u_{t}(\theta)}{h}\right),
\end{aligned}
$$

with $u_{t}^{(1)}(\theta):=\sum_{j=-\infty}^{\infty} \phi_{j}^{(1)}(\theta) u_{t-j}$. By simple algebra, we can show

$$
\mathbb{E} \sup _{\theta \in \Theta}\left\|\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta, j}(x, y ; h)\right\| \leq C_{h} \sum_{j=-\infty}^{\infty} \sup _{\theta \in \Theta}\left\|\phi^{(1)}(\theta)\right\| \mathbb{E}\left|u_{t}\right|<\infty,
$$

and

$$
\mathbb{E} \sup _{\theta \in \Theta}\left\|\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}(\theta ; h)\right\| \leq \frac{4}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \mathbb{E} \sup _{\theta \in \Theta}\left\|\tilde{\sigma}_{\theta, j}(x, y ; h) \frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta, j}(x, y ; h)\right\| d W(x, y)<\infty
$$

under $\mathbb{E}\left|u_{t}\right|<\infty$ and Assumption 2.4, 2.6 and the density function $\lambda(\cdot)$ being uniformly bounded in Assumption 2.5.
The identification of $\theta_{0}$ is achieved using $\mathcal{Q}_{0}(\theta ; h)$ since the dependence structure of the residual sequence $\left\{u_{t}(\theta)\right\}_{t=1}^{T}$ remains unchanged after this smooth transformation, indicating the unique minimum is obtained at $\theta=\theta_{0}$ and $\mathcal{Q}_{0}(\theta ; h)>0$ since $\theta \neq \theta_{0}$ under Assumption 2.1, 2.3 and 2.4.
Therefore, the consistency of $\tilde{\theta}_{T}^{h}$ follows.

### 2.7.6. Proof of Theorem 2.3.4

The asymptotic normality of the estimates, $\tilde{\theta}_{T}^{h}$ is derived by showing a CLT on the score of $\tilde{\mathcal{Q}}_{T}(\theta ; h)$ evaluated at the true value in the Taylor expansion of $\tilde{\mathcal{Q}}_{T}(\theta ; h)$, together with the convergence of the Hessian matrix of $\tilde{\mathcal{Q}}_{T}(\theta ; h)$,

$$
\begin{equation*}
0=\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\tilde{\theta}_{T}^{h} ; h\right)=\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)+\frac{\partial^{2}}{\partial \theta \partial \theta^{\prime}} \tilde{\mathcal{Q}}_{T}\left(\bar{\theta}_{T}^{h} ; h\right)\left(\tilde{\theta}_{T}^{h}-\theta_{0}\right), \quad \bar{\theta}_{T}^{h} \in\left(\theta_{0}, \tilde{\theta}_{T}^{h}\right) . \tag{2.7.4}
\end{equation*}
$$

First, we approximate the $\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)$ by the score of the theoretical counterpart of $\mathcal{Q}_{T}\left(\theta_{0} ; h\right)$ by showing

$$
\left\|\frac{\partial}{\partial \theta} \mathcal{Q}_{T}\left(\theta_{0} ; h\right)-\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)\right\|=o_{p}\left(T^{-1 / 2}\right),
$$

where
$\left\|\frac{\partial}{\partial \theta} \mathcal{Q}_{T}\left(\theta_{0} ; h\right)-\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)\right\| \leq\left\|\frac{\partial}{\partial \theta} \mathcal{Q}_{T}\left(\theta_{0} ; h\right)-\frac{\partial}{\partial \theta} \overline{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)\right\|+\left\|\frac{\partial}{\partial \theta} \overline{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)-\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)\right\|$
with $\mathcal{Q}_{T}\left(\theta_{0} ; h\right)$ and $\overline{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)$ defined by $\sigma_{\theta_{0}, j}(x, y ; h)$ and $\bar{\sigma}_{\theta_{0}, j}(x, y ; h)$, respectively.
First,

$$
\begin{aligned}
& \mathbb{E}\left\|\frac{\partial}{\partial \theta} \mathcal{Q}_{T}\left(\theta_{0} ; h\right)-\frac{\partial}{\partial \theta} \overline{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)\right\| \\
\leq & C_{h} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \mathbb{E}\left\|\sigma_{\theta_{0}, j}(x, y ; h) \frac{\partial}{\partial \theta} \sigma_{\theta_{0}, j}(x, y ; h)-\bar{\sigma}_{\theta_{0}, j}(x, y ; h) \frac{\partial}{\partial \theta} \bar{\sigma}_{\theta_{0}, j}(x, y ; h)\right\| d W(x, y) \\
= & o\left(T^{-1 / 2}\right)
\end{aligned}
$$

since

$$
\begin{aligned}
& \mathbb{E}\left\|\sigma_{\theta_{0}, j}(x, y ; h) \frac{\partial}{\partial \theta} \sigma_{\theta_{0}, j}(x, y ; h)-\bar{\sigma}_{\theta_{0}, j}(x, y ; h) \frac{\partial}{\partial \theta} \bar{\sigma}_{\theta_{0}, j}(x, y ; h)\right\| \\
\leq & 2 \mathbb{E}\left\|\frac{\partial}{\partial \theta} \sigma_{\theta_{0}, j}(x, y ; h)-\frac{\partial}{\partial \theta} \bar{\sigma}_{\theta_{0}, j}(x, y ; h)\right\|+C_{h}\left(\mathbb{E}\left|\sigma_{\theta_{0}, j}(x, y ; h)-\bar{\sigma}_{\theta_{0}, j}(x, y ; h)\right|^{2}\right)^{1 / 2} \\
\leq & C_{h}(T-j)^{-1}
\end{aligned}
$$

where the first inequality comes from triangular inequality and Cauchy-Schwarz inequality, and the second inequality follows from

$$
\begin{aligned}
\mathbb{E}\left\|\frac{\partial}{\partial \theta} \sigma_{\theta_{0}, j}(x, y ; h)-\frac{\partial}{\partial \theta} \bar{\sigma}_{\theta_{0}, j}(x, y ; h)\right\| & \leq C_{h}(T-j)^{-1} \\
\mathbb{E}\left|\sigma_{\theta_{0}, j}(x, y ; h)-\bar{\sigma}_{\theta_{0}, j}(x, y ; h)\right|^{2} & \leq C_{h}(T-j)^{-2}
\end{aligned}
$$

Next, with similar strategies,

$$
\begin{aligned}
& \mathbb{E}\left\|\frac{\partial}{\partial \theta} \overline{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)-\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)\right\| \\
\leq & C_{h} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \mathbb{E}\left\|\bar{\sigma}_{\theta_{0}, j}(x, y ; h) \frac{\partial}{\partial \theta} \bar{\sigma}_{\theta_{0}, j}(x, y ; h)-\tilde{\sigma}_{\theta_{0}, j}(x, y ; h) \frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y ; h)\right\| d W(x, y) \\
\leq & C_{h} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \mathbb{E}\left\|\frac{\partial}{\partial \theta} \bar{\sigma}_{\theta_{0}, j}(x, y ; h)-\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y ; h)\right\| d W(x, y) \\
& +C_{h} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}}\left(\mathbb{E}\left|\bar{\sigma}_{\theta_{0}, j}(x, y)-\tilde{\sigma}_{\theta_{0}, j}(x, y)\right|^{2} \mathbb{E}\left\|\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y)\right\|^{2}\right)^{1 / 2} d W(x, y) \\
= & o\left(T^{-1 / 2}\right),
\end{aligned}
$$

which follows

$$
\begin{aligned}
\mathbb{E}\left\|\frac{\partial}{\partial \theta} \bar{\sigma}_{\theta_{0}, j}(x, y ; h)-\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y ; h)\right\| & \leq C_{h}(T-j)^{-1} \\
\mathbb{E}\left|\bar{\sigma}_{\theta_{0}, j}(x, y)-\tilde{\sigma}_{\theta_{0}, j}(x, y)\right|^{2} & \leq C_{h}(T-j)^{-2}
\end{aligned}
$$

Now, we can check that

$$
\begin{aligned}
& T^{1 / 2} \frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0}\right) \\
= & \frac{4 T^{1 / 2}}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}}\left(\tilde{\sigma}_{\theta_{0}, j}(x, y) \frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y)\right) d W(x, y) \\
= & -\frac{4}{\pi^{2}} \frac{1}{T^{1 / 2}} \sum_{t=1}^{T} \int_{\mathbb{R}^{2}}\left\{\Lambda\left(\frac{x-u_{t}}{h}\right)-\psi^{h}(x)\right\} \sum_{j=1}^{\infty}\left\{\Lambda\left(\frac{y-u_{t-j}}{h}\right)-\psi^{h}(y)\right\} \\
& \times\left\{\phi_{j}^{(1)}\left(\theta_{0}\right) \lambda^{h}(x) \mu^{h}(y)+\phi_{-j}^{(1)}\left(\theta_{0}\right) \lambda^{h}(y) \mu^{h}(x)\right\} d W(x, y)+o_{p}(1) \\
= & -\frac{4}{\pi^{2}} \frac{1}{T^{1 / 2}} \sum_{t=1}^{T} \int_{\mathbb{R}}\left\{\Lambda\left(\frac{x-u_{t}}{h}\right)-\psi^{h}(x)\right\} \lambda^{h}(x) d W(x) \\
& \times \sum_{j=1}^{\infty} \phi_{j}^{(1)}\left(\theta_{0}\right) \int_{\mathbb{R}}\left\{\Lambda\left(\frac{y-u_{t-j}}{h}\right)-\psi^{h}(y)\right\} \mu^{h}(y) d W(y) \\
& -\frac{4}{\pi^{2}} \frac{1}{T^{1 / 2}} \sum_{t=1}^{T} \int_{\mathbb{R}}\left\{\Lambda\left(\frac{x-u_{t}}{h}\right)-\psi^{h}(x)\right\} \mu^{h}(x) d W(x) \\
& \times \sum_{j=1}^{\infty} \phi_{-j}^{(1)}\left(\theta_{0}\right) \int_{\mathbb{R}}\left\{\Lambda\left(\frac{y-u_{t-j}}{h}\right)-\psi^{h}(y)\right\} \lambda^{h}(y) d W(y)+o_{p}(1) \\
= & -\frac{4}{\pi^{2}} \frac{1}{T^{1 / 2}} \sum_{t=2}^{T} e_{t}^{h} V_{t-1}^{h}-\frac{4}{\pi^{2}} \frac{1}{T^{1 / 2}} \sum_{t=2}^{T} \nu_{t}^{h} E_{t-1}^{h}+o_{p}(1),
\end{aligned}
$$

with $\left\{e_{t}^{h}, \nu_{t}^{h}, V_{t-1}^{h}, E_{t-1}^{h}\right\}$ defined in the main text. The equality follows

$$
\begin{aligned}
& \mathbb{E}\left\|\tilde{\sigma}_{\theta_{0}, j}(x, y)\left(\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y)+\phi_{j}^{(1)}\left(\theta_{0}\right) \lambda^{h}(x) \mu^{h}(y)+\phi_{-j}^{(1)}\left(\theta_{0}\right) \lambda^{h}(y) \mu^{h}(x)\right)\right\| \\
\leq & \left(\mathbb{E}\left|\tilde{\sigma}_{\theta_{0}, j}(x, y)\right|^{2} \mathbb{E}\left\|\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y)+\phi_{j}^{(1)}\left(\theta_{0}\right) \lambda^{h}(x) \mu^{h}(y)+\phi_{-j}^{(1)}\left(\theta_{0}\right) \lambda^{h}(y) \mu^{h}(x)\right\|^{2}\right)^{1 / 2} \\
\leq & C_{h}(T-j)^{-1}
\end{aligned}
$$

using $\mathbb{E}\left|\tilde{\sigma}_{\theta_{0}, j}(x, y)\right|^{2} \leq(T-j)^{-1}$ and

$$
\mathbb{E}\left\|\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y)+\phi_{j}^{(1)}\left(\theta_{0}\right) \lambda^{h}(x) \mu^{h}(y)+\phi_{-j}^{(1)}\left(\theta_{0}\right) \lambda^{h}(y) \mu^{h}(x)\right\|^{2} \leq C_{h}(T-j)^{-1}
$$

Note that $\left\{e_{t}^{h} V_{t-1}^{h}\right\}$ and $\left\{\nu_{t}^{h} E_{t-1}^{h}\right\}$ are martingale difference sequences given the past information. Denote the variance of $\left(e_{t}^{h} V_{t-1}^{h}+\nu_{t}^{h} E_{t-1}^{h}\right)$ conditional on the $\sigma$-field by the past sequence of $u_{t}$ by $\operatorname{Var}_{t-j}\left(e_{t}^{h} V_{t-1}^{h}+\nu_{t}^{h} E_{t-1}^{h}\right)$, which can be computed by

$$
\begin{aligned}
\operatorname{Var}_{t-1}\left(e_{t}^{h} V_{t-1}^{h}+\nu_{t}^{h} E_{t-1}^{h}\right) & =\mathbb{V} \operatorname{ar}\left(e_{t}^{h} V_{t-1}^{h}+\nu_{t}^{h} E_{t-1}^{h} \mid \sigma\left(u_{t-1}, \ldots\right)\right) \\
& =\sigma_{e ; h}^{2} V_{t-1}^{h}\left(V_{t-1}^{h}\right)^{\prime}+\sigma_{\nu ; h}^{2} E_{t-1}^{h}\left(E_{t-1}^{h}\right)^{\prime}+\sigma_{e, \nu ; h}\left(V_{t-1}^{h}\left(E_{t-1}^{h}\right)^{\prime}+E_{t-1}^{h}\left(V_{t-1}^{h}\right)^{\prime}\right) .
\end{aligned}
$$

It can be shown that

$$
\begin{align*}
& \frac{1}{T} \sum_{t=1}^{T} \operatorname{Var}_{t-1}\left(e_{t}^{h} V_{t-1}^{h}+\nu_{t}^{h} E_{t-1}^{h}\right) \\
= & \sigma_{e ; h}^{2} \frac{1}{T} \sum_{t=1}^{T}\left(V_{t-1}^{h}\left(V_{t-1}^{h}\right)^{\prime}\right)+\sigma_{\nu ; h}^{2} \frac{1}{T} \sum_{t=1}^{T}\left(E_{t-1}^{h}\left(E_{t-1}^{h}\right)^{\prime}\right)+\sigma_{e, \nu ; h} \frac{1}{T} \sum_{t=1}^{T}\left(V_{t-1}^{h}\left(E_{t-1}^{h}\right)^{\prime}+E_{t-1}^{h}\left(V_{t-1}^{h}\right)^{\prime}\right) \\
= & \sigma_{\nu ; h}^{2}\left(\Sigma_{0,2}+\Sigma_{0,2}^{*}\right)+\sigma_{e, \nu ; h}\left(\Sigma_{0,2}^{\dagger}+\Sigma_{0,2}^{\dagger^{\prime}}\right)+o_{p}(1), \tag{2.7.5}
\end{align*}
$$

by showing

$$
E\left(E_{t-1}^{h}\left(E_{t-1}^{h}\right)^{\prime}\right)=\sum_{j=1}^{\infty} j^{-4} \phi_{-j}^{(1)}\left(\theta_{0}\right) \phi_{-j}^{(1)}\left(\theta_{0}\right)^{\prime} \sigma_{e ; h}^{2}, \quad E\left(V_{t-1}^{h}\left(V_{t-1}^{h}\right)^{\prime}\right)=\sum_{j=1}^{\infty} j^{-4} \phi_{-j}^{(1)}\left(\theta_{0}\right) \phi_{-j}^{(1)}\left(\theta_{0}\right)^{\prime} \sigma_{\nu ; h}^{2},
$$

and

$$
E\left(E_{t-1}^{h}\left(V_{t-1}^{h}\right)^{\prime}\right)=\sum_{j=1}^{\infty} j^{-4} \phi_{-j}^{(1)}\left(\theta_{0}\right) \phi_{-j}^{(1)}\left(\theta_{0}\right)^{\prime} \sigma_{e, \nu ; h}
$$

since $\mathbb{E}\left(e_{t-j}^{h} e_{t-k}^{h}\right)=0$ and $\mathbb{E}\left(\nu_{t-j}^{h} \nu_{t-k}^{h}\right)=0$ for $k \neq j$. Besides, we shall show

$$
\begin{equation*}
\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left(\left\|e_{t}^{h} V_{t-1}^{h}+\nu_{t}^{h} E_{t-1}^{h}\right\|^{2} I\left\{\left\|e_{t}^{h} V_{t-1}^{h}+\nu_{t}^{h} E_{t-1}^{h}\right\|>\epsilon \sqrt{T}\right\}\right)=o(1) \text { for any } \epsilon>0 \tag{2.7.6}
\end{equation*}
$$

The sufficient condition for (2.7.6) is

$$
\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left\|e_{t}^{h} V_{t-1}^{h}\right\|^{4}+\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left\|\nu_{t}^{h} E_{t-1}^{h}\right\|^{4}<\infty
$$

For simplicity, we assume $E_{t-1}^{h}$ and $V_{t-1}^{h}$ are of single dimension. Hence, the above expression can be written as

$$
\mathbb{E}\left|e_{t}^{h}\right|^{4} \frac{\sum_{t=1}^{T} \mathbb{E}\left|V_{t-1}^{h}\right|^{4}}{T}+\mathbb{E}\left|\nu_{t}^{h}\right|^{4} \frac{\sum_{t=1}^{T} \mathbb{E}\left|E_{t-1}^{h}\right|^{4}}{T}
$$

which is bounded because

$$
\begin{aligned}
\mathbb{E}\left|V_{t-1}^{h}\right|^{4} & =\sum_{k_{1}=1}^{\infty} \sum_{k_{2}=1}^{\infty} \sum_{k_{3}=1}^{\infty} \sum_{k_{4}=1}^{\infty}\left(k_{1} k_{2} k_{3} k_{4}\right)^{-2} \phi_{k_{1}}^{(1)}\left(\theta_{0}\right) \phi_{k_{1}}^{(2)}\left(\theta_{0}\right) \phi_{k_{3}}^{(1)}\left(\theta_{0}\right) \phi_{k_{4}}^{(1)}\left(\theta_{0}\right) \mathbb{E}\left|\nu_{t-k_{1}}^{h} \nu_{t-k_{2}}^{h} \nu_{t-k_{3}}^{h} \nu_{t-k_{4}}^{h}\right| \\
& \leq C_{h}\left(\sum_{k=1}^{\infty} k^{-2}\left\|\phi_{k}^{(1)}\left(\theta_{0}\right)\right\|\right)^{4}<\infty .
\end{aligned}
$$

With (2.7.5) and (2.7.6), we can apply CLT for martingale difference sequences by Brown (1971) on $e_{t}^{h} V_{t-1}^{h}+\nu_{t}^{h} E_{t-1}^{h}$ and obtain,

$$
T^{1 / 2} \frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right) \rightarrow_{p} \mathcal{N}\left(0, \frac{16}{\pi^{4}} H_{1, h}\right)
$$

The convergence of Hessian matrix, $\frac{\partial^{2}}{\partial \theta \partial \theta^{\prime}} \tilde{\mathcal{Q}}\left(\theta_{0}\right)$ is derived in a similar way,

$$
\begin{aligned}
& \frac{\partial^{2}}{\partial \theta \partial \theta^{\prime}} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right) \\
= & \frac{4}{\pi^{4}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}}\left\{\left(\frac{\partial \tilde{\sigma}_{\theta_{0}, j}(x, y ; h)}{\partial \theta}\right)\left(\frac{\partial \tilde{\sigma}_{\theta_{0}, j}(x, y ; h)}{\partial \theta}\right)^{\prime}+\left(\tilde{\sigma}_{\theta_{0, j}}(x, y ; h) \frac{\partial^{2} \tilde{\sigma}_{\theta_{0}, j}(x, y ; h)}{\partial \theta \partial \theta^{\prime}}\right)\right\} d W(x, y) \\
= & \frac{4}{\pi^{2}} \sum_{j=1}^{\infty} \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \mathbb{E}\left(\frac{\partial \tilde{\sigma}_{\theta_{0}, j}(x, y ; h)}{\partial \theta}\right) \mathbb{E}\left(\frac{\partial \tilde{\sigma}_{\theta_{0}, j}(x, y ; h)}{\partial \theta}\right)^{\prime} d W(x, y)+o_{p}(1) \\
= & \frac{4}{\pi^{2}} \sum_{j=1}^{\infty} \phi_{j}^{(1)}\left(\theta_{0}\right) \phi_{j}^{(1)}\left(\theta_{0}\right)^{\prime} \frac{1}{j^{2}} \int_{\mathbb{R}}\left(\mu^{h}(y)\right)^{2} d W(y) \int_{\mathbb{R}}\left(\lambda^{h}(x)\right)^{2} d W(x) \\
& +\frac{4}{\pi^{2}} \sum_{j=1}^{\infty} \phi_{-j}^{(1)}\left(\theta_{0}\right) \phi_{-j}^{(1)}\left(\theta_{0}\right)^{\prime} \frac{1}{j^{2}} \int_{\mathbb{R}}\left(\mu^{h}(x)\right)^{2} d W(x) \int_{\mathbb{R}}\left(\lambda^{h}(y)\right)^{2} d W(y) \\
& +\frac{4}{\pi^{2}} \sum_{j=1}^{\infty}\left(\phi_{j}^{(1)}\left(\theta_{0}\right) \phi_{-j}^{(1)}\left(\theta_{0}\right)^{\prime}+\phi_{-j}^{(1)}\left(\theta_{0}\right) \phi_{j}^{(1)}\left(\theta_{0}\right)^{\prime}\right) \frac{1}{j^{2}}\left(\int_{\mathbb{R}} \mu^{h}(y) \lambda^{h}(y) d W(y)\right)^{2}+o_{p}(1) \\
= & \frac{4}{\pi^{2}} h_{1}^{h} \rho_{2}^{h}\left(\Sigma_{0,1}+\sum_{0,1}^{*}\right)+\frac{4}{\pi^{2}}\left(\rho_{12}^{h}\right)^{2}\left(\sum_{0,1}^{\dagger}+\sum_{0,1}^{\dagger^{\prime}}\right)+o_{p}(1)=\frac{4}{\pi^{2}} H_{0, h}+o_{p}(1) .
\end{aligned}
$$

The uniform boundedness of the third-order derivative of the objective function is also required for the convergence of the Hessian matrix. For simplicity, we consider the analysis of $\frac{\partial^{3}}{\partial \theta^{3}} \tilde{\mathcal{Q}}_{T}(\theta)$ in single dimension,

$$
\begin{aligned}
& \frac{\partial^{3}}{\partial \theta^{3}} \tilde{Q}_{T}(\theta) \\
= & \frac{4}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}}\left\{3\left(\frac{\partial \tilde{\sigma}_{\theta, j}(x, y ; h)}{\partial \theta}\right) \frac{\partial^{2} \tilde{\sigma}_{\theta, j}(x, y ; h)}{\partial \theta^{2}}+\left(\tilde{\sigma}_{\theta, j}(x, y ; h) \frac{\partial^{3} \tilde{\sigma}_{\theta, j}(x, y ; h)}{\partial \theta^{3}}\right)\right\} d W(x, y) .
\end{aligned}
$$

Some uniform bounds can be derived for the components of $\frac{\partial^{3}}{\partial \theta^{3}} \tilde{Q}_{T}(\theta)$ under Assumption 2.5, 2.7 and $\mathbb{E}\left|u_{t}\right|^{3}$. By applying Minkowski's inequality, we can show

$$
\begin{aligned}
& \mathbb{E} \sup _{\theta \in \Theta}\left\|u_{t}^{(a)}(\theta)\right\|^{b} \\
\leq & \mathbb{E} \sup _{\theta \in \Theta}\left\|\sum_{j=-\infty}^{\infty} \phi_{j}^{(a)}(\theta)\left|u_{t}\right|\right\|^{b} \leq C \sup _{\theta \in \Theta} \sum_{j=-\infty}^{\infty}\left\|\phi_{j}^{(a)}(\theta)\right\|^{b} \mathbb{E}\left|u_{t}\right|^{b}<\infty \text { for } a, b=1,2,3 .
\end{aligned}
$$

Then, by Hölder's inequality, we can get

$$
\begin{aligned}
& \mathbb{E} \sup _{\theta \in \Theta}\left\|\left(u_{t}^{\left(a_{1}\right)}(\theta)\right)\left(u_{t-j}^{\left(a_{2}\right)}(\theta)\right)\right\| \\
\leq & \left(\mathbb{E} \sup _{\theta \in \Theta}\left\|u_{t}^{\left(a_{1}\right)}(\theta)\right\|^{b_{1}}\right)^{1 / b_{1}}\left(\mathbb{E} \sup _{\theta \in \Theta}\left\|u_{t-j}^{\left(a_{2}\right)}(\theta)\right\|^{b_{2}}\right)^{1 / b_{2}}<\infty \text { for } a_{1}, a_{2}=1,2, b_{1}, b_{2}>1 \text { and } j=0, \pm 1, \pm 2, \ldots \\
& \mathbb{E} \sup _{\theta \in \Theta}\left\|\left(u_{t}^{(1)}(\theta)\right)^{m}\left(u_{t-j}^{(1)}(\theta)\right)^{n}\right\| \\
\leq & \left(\mathbb{E} \sup _{\theta \in \Theta}\left\|\left(u_{t}^{(1)}(\theta)\right)^{m}\right\|^{2}\right)^{1 / 2}\left(\mathbb{E} \sup _{\theta \in \Theta}\left\|\left(u_{t-j}^{(1)}(\theta)\right)^{n}\right\|^{2}\right)^{1 / 2} \text { for } m, n=1,2 \text { and } j=0, \pm 1, \pm 2, \ldots
\end{aligned}
$$

Following the above inequalities together with the uniform boundedness of density of $\Lambda$ and its corresponding derivatives of order 1 and 2 , i.e., $\dot{\lambda}$ and $\ddot{\lambda}$, we are able to prove

$$
\begin{aligned}
& \mathbb{E} \sup _{\theta \in \Theta,(x, y) \in \mathbb{R}^{2}}\left\|\frac{\partial^{a} \tilde{F}_{\theta, j}(x, y ; h)}{\partial \theta^{a}}\right\|<\infty, \text { and } \\
& \mathbb{E} \sup _{\theta \in \Theta,(x, y) \in \mathbb{R}^{2}}\left\|\frac{\partial^{a} \tilde{\sigma}_{\theta, j}(x, y ; h)}{\partial \theta^{a}}\right\|<\infty, \text { for } a=1,2,3 \text { and } j= \pm 1, \pm 2, \ldots
\end{aligned}
$$

Then,

$$
\left.\begin{array}{l}
\mathbb{E}\left\|\sup _{\theta \in \Theta} \frac{\partial^{3}}{\partial \theta^{3}} \tilde{Q}_{T}(\theta)\right\| \\
\leq \\
\frac{4}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \mathbb{E} \sup _{\theta \in \Theta,(x, y) \in \mathbb{R}^{2}}\left\|3\left(\frac{\partial \tilde{\sigma}_{\theta, j}(x, y ; h)}{\partial \theta}\right) \frac{\partial^{2} \tilde{\sigma}_{\theta, j}(x, y ; h)}{\partial \theta^{2}}\right\| d W(x, y) \\
\quad+\frac{4}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \mathbb{E} \sup _{\theta \in \Theta,(x, y) \in \mathbb{R}^{2}}\left\|\tilde{\sigma}_{\theta, j}(x, y ; h) \frac{\partial^{3} \tilde{\sigma}_{\theta, j}(x, y ; h)}{\partial \theta^{3}}\right\| d W(x, y) \\
\leq
\end{array}\right] \sum_{j=1}^{T-1} \frac{1}{(j \pi)^{2}}<\infty \quad 10
$$

Last, given consistency of $\tilde{\theta}_{T}^{h}$, so as $\bar{\theta}_{T}^{h} \in\left(\tilde{\theta}_{T}^{h}, \theta_{0}\right)$, together with the Taylor expansion (2.7.4), the theorem follows.

### 2.7.7. Proof of Theorem 2.3 .5

Identification To check the identification of $\theta_{0}$ using $\mathcal{Q}_{0}(\theta ; h)$ is still valid when $h \rightarrow 0$, we need to study the approximation of $\mathcal{Q}_{0}(\theta ; h)$ to $\mathcal{Q}_{0}(\theta)$ in this dynamic process of $h$, where the key component is $\sigma_{\theta, j}(x, y ; h)$. Let $G_{\theta, j}(\cdot, \cdot)$ be the joint probability $c d f$ of $\left(u_{t}(\theta), u_{t-j}(\theta)\right)$ with a uniformly bounded Lebesgue pdf $g_{\theta, j}(\cdot, \cdot)$. Then, using integration by parts,

$$
\begin{aligned}
F_{\theta, j}(x, y ; h)= & \int_{\mathbb{R}} \int_{\mathbb{R}} \Lambda\left(\frac{x-a}{h}\right) \Lambda\left(\frac{y-b}{h}\right) g_{\theta, j}(a, b) d a d b \\
= & {\left[\left[\Lambda\left(\frac{x-a}{h}\right) \Lambda\left(\frac{y-b}{h}\right) G_{\theta, j}(a, b)\right]_{-\infty}^{\infty}\right]_{-\infty}^{\infty} } \\
& -\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{h^{2}} \lambda\left(\frac{x-a}{h}\right) \lambda\left(\frac{y-b}{h}\right) G_{\theta, j}(a, b) d a d b \\
& +\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{h} \lambda\left(\frac{x-a}{h}\right) \Lambda\left(\frac{y-b}{h}\right) \frac{\partial G_{\theta, j}(a, b)}{\partial b} d a d b \\
& +\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{h} \Lambda\left(\frac{x-a}{h}\right) \lambda\left(\frac{y-b}{h}\right) \frac{\partial G_{\theta, j}(a, b)}{\partial a} d a d b
\end{aligned}
$$

where the first term is

$$
\Lambda(-\infty) \Lambda(-\infty) G_{\theta, j}(\infty, \infty)-\Lambda(\infty) \Lambda(\infty) G_{\theta, j}(-\infty,-\infty)=0
$$

and the second term is

$$
-G_{\theta, j}(x, y)+O\left(h^{2}\right),
$$

uniformly in $(x, y) \in \mathbb{R}^{2}$, exploiting the fact that

$$
\begin{aligned}
\int_{\mathbb{R}} \Lambda\left(\frac{y-b}{h}\right) \frac{\partial G_{\theta, j}(a, b)}{\partial b} d b & =\left[\Lambda\left(\frac{y-b}{h}\right) G_{\theta, j}(a, b)\right]_{b=-\infty}^{\infty}+\frac{1}{h} \int_{\mathbb{R}} \lambda\left(\frac{y-b}{h}\right) G_{\theta, j}(a, b) d b \\
& =0+G_{\theta, j}(a, y)+O\left(h^{2}\right)
\end{aligned}
$$

uniformly in $(a, y)$, then the third term is

$$
\int_{\mathbb{R}} \frac{1}{h} \lambda\left(\frac{x-a}{h}\right) G_{\theta, j}(a, y) d a+O\left(h^{2}\right)=G_{\theta, j}(x, y)+O\left(h^{2}\right),
$$

and the same for the fourth term, thus

$$
F_{\theta, j}(x, y ; h)=G_{\theta, j}(x, y)+O\left(h^{2}\right)
$$

uniformly in $(x, y)$ as $h \rightarrow 0$.
Then, the loss function with smoothed $c d f$ can be reexpressed as

$$
\mathcal{Q}_{0}(\theta ; h)=\frac{2}{\pi^{2}} \sum_{j=1}^{\infty} \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \sigma_{\theta, j}^{2}(x, y) d W(x, y)+O\left(h^{2}\right)
$$

where $\sigma_{\theta, j}(x, y)$ is the general covariance based on the indicator function. From the above statement, we can conclude that identification using $\mathcal{Q}_{0}(\theta ; h)$ is guaranteed as $h \rightarrow 0$ since $\mathcal{Q}_{0}(\theta ; h)>0$ when $\theta \neq \theta_{0}$ under Assumptions 2.1, 2.3 and 2.4.

Consistency Consistency of $\tilde{\theta}_{T}^{h}$ as $h \rightarrow 0$ requires the uniform convergence of $\tilde{\mathcal{Q}}_{T}(\theta ; h)$ to $\mathcal{Q}_{0}(\theta)$ as $h \rightarrow 0$. Equivalently, we need to show all the terms on the rhs of the following expression are negligible,

$$
\begin{aligned}
\tilde{\mathcal{Q}}_{T}(\theta ; h)-\mathcal{Q}_{0}(\theta)= & \frac{2}{\pi^{2}} \sum_{j=1}^{T-1} \frac{1}{j^{2}} \int_{\mathbb{R}^{2}}\left\{\tilde{\sigma}_{\theta, j}^{2}(x, y ; h)-\sigma_{\theta, j}^{2}(x, y)\right\} d W(x, y) \\
& -\frac{2}{\pi^{2}} \sum_{j=1}^{T-1} \frac{1}{j^{2}} \frac{j}{T} \int_{\mathbb{R}^{2}} \tilde{\sigma}_{\theta, j}^{2}(x, y ; h) d W(x, y) \\
& -\frac{2}{\pi^{2}} \sum_{j=T}^{\infty} \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \sigma_{\theta, j}^{2}(x, y) d W(x, y) .
\end{aligned}
$$

The last two terms are easy to show to be $o_{p}(1)$ uniformly in $\theta \in \Theta$ as $h \rightarrow 0$ because of $\left|\tilde{\sigma}_{\theta, j}(x, y ; h)\right| \leq 1$ and $\left|\sigma_{\theta, j}(x, y)\right| \leq 1$ regardless of $h$. So the key part is

$$
\mathbb{E}\left|\tilde{\sigma}_{\theta, j}(x, y ; h)-\sigma_{\theta, j}(x, y)\right|^{2}
$$

which needs to be bounded for $(x, y, \theta)$ as $h \rightarrow 0$. We would show it by decomposing the term into the squared bias and the variance.
Repeat the steps of approximating $F_{\theta, j}(x, y ; h)$ with $F_{\theta}(x, \infty ; h) F_{\theta}(\infty, y ; h)$ using the truncation and MVT techniques, we have
$\Lambda\left(\frac{x-u_{t}(\theta)}{h}\right)=\Lambda\left(\frac{x-u_{t}^{m}(\theta)}{h}\right)+\frac{1}{h} \lambda\left(\frac{u_{t}^{m}(\theta)-u_{t}^{m}(\theta)}{h} \tau\right)\left(u_{t}^{m}(\theta)-u_{t}(\theta)\right)$ for some $\tau \in(0,1)$,
so,

$$
\begin{aligned}
\mathbb{E}\left(\Lambda\left(\frac{x-u_{t}(\theta)}{h}\right) \Lambda\left(\frac{x-u_{t-j}(\theta)}{h}\right)\right)= & \mathbb{E}\left(\Lambda\left(\frac{x-u_{t}^{m}(\theta)}{h}\right)\right) \mathbb{E}\left(\Lambda\left(\frac{x-u_{t-j}^{m}(\theta)}{h}\right)\right) \\
& +O\left(h^{-1} \mathbb{E}\left|u_{t}^{m}(\theta)-u_{t}(\theta)\right|\right) \\
& +O\left(h^{-2} \mathbb{E}\left|\left(u_{t}^{m}(\theta)-u_{t}(\theta)\right)\left(u_{t-j}^{m}(\theta)-u_{t-j}(\theta)\right)\right|\right)
\end{aligned}
$$

for $m=j / 2$, for $\mathbb{E}\left(u_{t}^{2}\right)<\infty$,

$$
\begin{aligned}
& \mathbb{E}\left(\Lambda\left(\frac{x-u_{t}^{m}(\theta)}{h}\right)\right) \mathbb{E}\left(\Lambda\left(\frac{x-u_{t-j}^{m}(\theta)}{h}\right)\right)+O\left(h^{-1} j^{1 / 2-\mu_{0}}\right)+O\left(h^{-2} j^{1-2 \mu_{0}}\right) \\
= & \mathbb{E}\left(\Lambda\left(\frac{x-u_{t}(\theta)}{h}\right)\right) \mathbb{E}\left(\Lambda\left(\frac{x-u_{t-j}(\theta)}{h}\right)\right)+O\left(h^{-1} j^{1 / 2-\mu_{0}}\right)+O\left(h^{-2} j^{1-2 \mu_{0}}\right) \\
= & F_{\theta, j}(x, \infty ; h) F_{\theta, j}(\infty, y ; h)+O\left(h^{-1} j^{1 / 2-\mu_{0}}\right)+O\left(h^{-2} j^{1-2 \mu_{0}}\right) \\
= & G_{\theta}(x) G_{\theta}(y)+\underbrace{O\left(h^{-1} j^{1 / 2-\mu_{0}}\right)+O\left(h^{-2} j^{1-2 \mu_{0}}\right)}_{\text {approximation error from the truncation and MVT }}+ \\
& \underbrace{O\left(h^{2}\right)}_{\text {approximation error from the smoothed } c d f}
\end{aligned}
$$

It is worth noting that there would be a tradeoff between $h$ and $T$ when $h \rightarrow 0$, since when $h$ approaches zero too fast, which makes the approximation error of the indicator function from the smoothed $c d f$ sufficiently small, it also leads to a less precise performance of the joint $c d f$ by the product of the marginal $c d f$ due to the terms in $h^{-1}$ and $h^{-2}$. However, an equivalent version of the proof of Theorem 2.3.3 can be justified without further issue when $h \rightarrow 0$, making sure that $h$ converges slowly enough. The convergence rate of $h$ associated with $T$ needs to be considered with extra care through the key component of the score, $\frac{\partial \tilde{\sigma}_{\theta, j}(x, y ; h)}{\partial \theta}$, in the derivation of the asymptotic distribution.

Asymptotic normality To study the effect of $h$ in the asymptotic distribution of $\tilde{\theta}_{T}^{h}$ as $h \rightarrow 0$, we first write

$$
\begin{aligned}
\tilde{\sigma}_{\theta, j}(x, y ; h) & =\frac{1}{T-j} \sum_{t=j+1}^{T}\left\{\Lambda\left(\frac{x-u_{t}(\theta)}{h}\right)-\hat{F}_{\theta, j}(x, \infty ; h)\right\}\left\{\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)-\hat{F}_{\theta, j}(\infty, y ; h)\right\} \\
& \approx \frac{1}{T-j} \sum_{t=j+1}^{T}\left\{\Lambda\left(\frac{x-u_{t}(\theta)}{h}\right)-\hat{G}_{\theta}(x)\right\}\left\{\Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right)-\hat{G}_{\theta}(y)\right\} .
\end{aligned}
$$

To control for the bias arising from centering the variables at $F(x)=G_{\theta_{0}}(x)$, the marginal cdf of the true innovation $u_{t}$, instead of $\psi^{h}(x)=\mathbb{E}\left(\Lambda\left(\frac{x-u_{t}}{h}\right)\right)$ when $\theta=\theta_{0}$ in the asymptotic theory, we need to check that

$$
\begin{aligned}
& \mathbb{E}\left[\Lambda\left(\frac{x-u_{t}}{h}\right) \Lambda\left(\frac{y-u_{t-j}}{h}\right)\right]-F(x) F(y) \\
= & \mathbb{E}\left[\Lambda\left(\frac{x-u_{t}}{h}\right)\right] \mathbb{E}\left[\Lambda\left(\frac{y-u_{t-j}}{h}\right)\right]-F(x) F(y) \\
= & \left(F(x)+O\left(h^{2}\right)\right)\left(F(y)+O\left(h^{2}\right)\right)-F(x) F(y) \\
= & O\left(h^{2}\right)=o\left((T-j)^{-1 / 2}\right),
\end{aligned}
$$

and for this approximation error to be negligible in the asymptotic analysis at the rate of $T^{-1 / 2}$, it is necessary that

$$
\begin{equation*}
h=o\left(T^{-1 / 4}\right) \text {, i.e. } h T^{1 / 4} \rightarrow 0, \tag{2.7.7}
\end{equation*}
$$

which basically restricts $h$ not tending to zero too slowly with the purpose of controlling for bias in the asymptotic distribution.
Next, we focus on the key component of the score,

$$
\begin{aligned}
\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta, j}(x, y ; h) \approx & -\frac{1}{h(T-j)} \sum_{t=j+1}^{T} u_{t}^{(1)}(\theta) \lambda\left(\frac{x-u_{t}(\theta)}{h}\right) \Lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right) \\
& -\frac{1}{h(T-j)} \sum_{t=j+1}^{T} u_{t-j}^{(1)}(\theta) \Lambda\left(\frac{x-u_{t}(\theta)}{h}\right) \lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right) \\
& +\hat{G}_{\theta}(y) \frac{1}{h(T-j)} \sum_{t=j+1}^{T} u_{t}^{(1)}(\theta) \lambda\left(\frac{x-u_{t}(\theta)}{h}\right) \\
& +\hat{G}_{\theta}(x) \frac{1}{h(T-j)} \sum_{t=j+1}^{T} u_{t-j}^{(1)}(\theta) \lambda\left(\frac{y-u_{t-j}(\theta)}{h}\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y ; h) \approx & -\frac{1}{h(T-j)} \sum_{t=j+1}^{T} u_{t}^{(1)}\left(\theta_{0}\right) \lambda\left(\frac{x-u_{t}}{h}\right) \Lambda\left(\frac{y-u_{t-j}}{h}\right) \\
& -\frac{1}{h(T-j)} \sum_{t=j+1}^{T} u_{t-j}^{(1)}\left(\theta_{0}\right) \Lambda\left(\frac{x-u_{t}}{h}\right) \lambda\left(\frac{y-u_{t-j}}{h}\right) \\
& +\hat{G}_{\theta_{0}}(y) \frac{1}{h(T-j)} \sum_{t=j+1}^{T} u_{t}^{(1)}\left(\theta_{0}\right) \lambda\left(\frac{x-u_{t}}{h}\right) \\
& +\hat{G}_{\theta_{0}}(x) \frac{1}{h(T-j)} \sum_{t=j+1}^{T} u_{t-j}^{(1)}\left(\theta_{0}\right) \lambda\left(\frac{y-u_{t-j}}{h}\right) .
\end{aligned}
$$

With $\mathbb{E}\left(u_{t}\right)=0$, we have that

$$
\begin{align*}
\frac{\partial}{\partial \theta} & \tilde{\sigma}_{\theta_{0}, j}(x, y ; h) \\
\rightarrow_{p}- & \frac{1}{h} \mathbb{E}\left[u_{t} \lambda\left(\frac{x-u_{t}}{h}\right)\right] \mathbb{E}\left[\Lambda\left(\frac{y-u_{t}}{h}\right)\right]-\frac{1}{h} \phi_{j}^{(1)}\left(\theta_{0}\right) \mathbb{E}\left[u_{t} \Lambda\left(\frac{y-u_{t}}{h}\right)\right] \mathbb{E}\left[\lambda\left(\frac{x-u_{t}}{h}\right)\right] \\
& -\frac{1}{h} \mathbb{E}\left[u_{t} \lambda\left(\frac{y-u_{t}}{h}\right)\right] \mathbb{E}\left[\Lambda\left(\frac{x-u_{t}}{h}\right)\right]-\frac{1}{h} \phi_{-j}^{(1)}\left(\theta_{0}\right) \mathbb{E}\left[u_{t} \Lambda\left(\frac{x-u_{t}}{h}\right)\right] \mathbb{E}\left[\lambda\left(\frac{y-u_{t}}{h}\right)\right] \\
& +F(x) \frac{1}{h} \mathbb{E}\left[u_{t} \lambda\left(\frac{x-u_{t}}{h}\right)\right]+F(y) \frac{1}{h} \mathbb{E}\left[u_{t} \lambda\left(\frac{y-u_{t}}{h}\right)\right], \tag{2.7.8}
\end{align*}
$$

for $j=1,2, \ldots$.
Then, let $h \rightarrow 0$, assuming $\Lambda$ and $\lambda$ are symmetric around 0 for simplicity, by Taylor expansion,

$$
\begin{align*}
\frac{1}{h} \mathbb{E}\left[\lambda\left(\frac{x-u_{t}}{h}\right)\right] & =\frac{1}{h} \int \lambda\left(\frac{x-u}{h}\right) f(u) d u=f(x)+O\left(h^{2} \sup _{u}\left|f^{(2)}(u)\right|\right) \\
\frac{1}{h} \mathbb{E}\left[u_{t} \lambda\left(\frac{x-u_{t}}{h}\right)\right] & =\frac{1}{h} \int u \lambda\left(\frac{x-u}{h}\right) f(u) d u=x f(x)+O\left(h^{2}\left\{\sup _{u}\left|f^{(1)}(u)\right|+\sup _{u}\left|u f^{(2)}(u)\right|\right\}\right) \\
\mathbb{E}\left[\Lambda\left(\frac{y-u_{t}}{h}\right)\right] & =\int \Lambda\left(\frac{y-u}{h}\right) f(u) d u=\frac{1}{h} \int F(u) \lambda\left(\frac{y-u}{h}\right) d u=F(y)+O\left(h^{2} \sup _{u}\left|f^{(1)}(u)\right|\right) \\
\mathbb{E}\left(u_{t} \Lambda\left(\frac{y-u_{t}}{h}\right)\right) & =\int u \Lambda\left(\frac{y-u}{h}\right) f(u) d u=\frac{1}{h} \int u \lambda\left(\frac{y-u}{h}\right) F(u) d u-\int F(u) \Lambda\left(\frac{y-u}{h}\right) d u \\
& =\mathbb{E}(u I(u \leq y))+O\left(h^{2} \sup _{u}\left\{f(u)+\left|u f^{(1)}(u)\right|\right\}\right) . \tag{2.7.9}
\end{align*}
$$

From (2.7.9), we can see that the bias between $\mathbb{E}\left(\frac{\partial \tilde{\sigma}_{0}, j(x, y ; h)}{\partial \theta}\right)$ and $\frac{\partial \sigma_{\theta_{0}, j}(x, y)}{\partial \theta}=$ $-\phi_{j}^{(1)}\left(\theta_{0}\right) \mu(y) f(x)-\phi_{-j}^{(1)}\left(\theta_{0}\right) \mu(x) f(y)$, is $O\left(h^{2}\right)$ under the uniform boundedness of the density function $f(\cdot)$ and its derivative of order 1 and 2 in Assumption 2.7.

Concerning $\operatorname{Var}\left(\frac{\partial \tilde{\sigma}_{\theta}, j}{}(x, y ; h)\right.$, we have

$$
\begin{aligned}
& \mathbb{V a r}\left[\frac{1}{h(T-j)} \sum_{t=j+1}^{T} u_{t}^{(1)}\left(\theta_{0}\right) \lambda\left(\frac{x-u_{t}}{h}\right) \Lambda\left(\frac{y-u_{t-j}}{h}\right)\right] \\
= & \frac{1}{h^{2}(T-j)} \operatorname{Var}\left[u_{t}^{(1)}\left(\theta_{0}\right) \lambda\left(\frac{x-u_{t}}{h}\right) \Lambda\left(\frac{y-u_{t-j}}{h}\right)\right] \\
+ & \frac{1}{h^{2}(T-j)^{2}} \sum_{t=j+1}^{T} \sum_{t^{\prime}=j+1}^{T} \operatorname{Cov}\left[u_{t}^{(1)}\left(\theta_{0}\right) \lambda\left(\frac{x-u_{t}}{h}\right) \Lambda\left(\frac{y-u_{t-j}}{h}\right), u_{t^{\prime}}^{(1)}\left(\theta_{0}\right) \lambda\left(\frac{x-u_{t^{\prime}}}{h}\right) \Lambda\left(\frac{y-u_{t^{\prime}-j}}{h}\right)\right],
\end{aligned}
$$

where

$$
\begin{aligned}
& \mathbb{V} a r\left[u_{t}^{(1)}\left(\theta_{0}\right) \lambda\left(\frac{x-u_{t}}{h}\right) \Lambda\left(\frac{y-u_{t-j}}{h}\right)\right] \\
= & \mathbb{E}\left[\left(u_{t}^{(1)}\left(\theta_{0}\right)\right)^{2} \lambda^{2}\left(\frac{x-u_{t}}{h}\right) \Lambda^{2}\left(\frac{y-u_{t-j}}{h}\right)\right]-\mathbb{E}\left[u_{t}^{(1)}\left(\theta_{0}\right) \lambda\left(\frac{x-u_{t}}{h}\right) \Lambda\left(\frac{y-u_{t-j}}{h}\right)\right]^{2},
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathbb{E}\left[\left(u_{t}^{(1)}\left(\theta_{0}\right)\right)^{2} \lambda^{2}\left(\frac{x-u_{t}}{h}\right) \Lambda^{2}\left(\frac{y-u_{t-j}}{h}\right)\right] \\
= & \sum_{k} \sum_{k^{\prime}} \phi_{k}^{(1)}\left(\theta_{0}\right) \phi_{k^{\prime}}^{(1)}\left(\theta_{0}\right) \mathbb{E}\left[u_{t-k} u_{t-k^{\prime}} \lambda^{2}\left(\frac{x-u_{t}}{h}\right) \Lambda^{2}\left(\frac{y-u_{t-j}}{h}\right)\right] \\
= & \sum_{k} \phi_{k}^{(1)}\left(\theta_{0}\right) \mathbb{E}\left[u_{t-k}^{2} \lambda^{2}\left(\frac{x-u_{t}}{h}\right) \Lambda^{2}\left(\frac{y-u_{t-j}}{h}\right)\right] \\
& +\sum_{k} \sum_{k^{\prime}} \phi_{k}^{(1)}\left(\theta_{0}\right) \phi_{k^{\prime}}^{(1)}\left(\theta_{0}\right) \mathbb{E}\left[u_{t-k} \lambda^{2}\left(\frac{x-u_{t}}{h}\right)\right] \mathbb{E}\left[u_{t-k^{\prime}} \Lambda^{2}\left(\frac{y-u_{t-j}}{h}\right)\right] \\
& +\sum_{k} \sum_{k^{\prime}} \phi_{k}^{(1)}\left(\theta_{0}\right) \phi_{k^{\prime}}^{(1)}\left(\theta_{0}\right) \mathbb{E}\left[u_{t-k} \Lambda^{2}\left(\frac{y-u_{t-j}}{h}\right)\right] \mathbb{E}\left[u_{t-k^{\prime}} \lambda^{2}\left(\frac{x-u_{t}}{h}\right)\right],
\end{aligned}
$$

so that for $k=0$ ( also applies to $k=j$ in the case of $\left.u_{t-j}\right)$,

$$
\begin{aligned}
\mathbb{E}\left[u_{t-k}^{2} \lambda^{2}\left(\frac{x-u_{t}}{h}\right) \Lambda^{2}\left(\frac{y-u_{t-j}}{h}\right)\right] & \leq C \int_{\mathbb{R}} u^{2} \lambda^{2}\left(\frac{x-u}{h}\right) f(u) d u=O\left(h^{3}\right) \\
\mathbb{E}\left[u_{t-k} \lambda^{2}\left(\frac{x-u}{h}\right)\right] & =\int_{\mathbb{R}} u \lambda^{2}\left(\frac{x-u}{h}\right) f(u) d u=O\left(h^{2}\right)
\end{aligned}
$$

and $k \neq 0, j$,

$$
\begin{aligned}
\mathbb{E}\left[u_{t-k}^{2} \lambda^{2}\left(\frac{x-u_{t}}{h}\right) \Lambda^{2}\left(\frac{y-u_{t-j}}{h}\right)\right] & \leq C \mathbb{E}\left(u_{t-k}^{2}\right) \int_{\mathbb{R}} \lambda^{2}\left(\frac{x-u}{h}\right) f(u) d u=O(h) \\
\mathbb{E}\left[u_{t-k} \lambda^{2}\left(\frac{x-u}{h}\right)\right] & =0 .
\end{aligned}
$$

Then, $\operatorname{Var}\left[u_{t}^{(1)}\left(\theta_{0}\right) \lambda\left(\frac{x-u_{t}}{h}\right) \Lambda\left(\frac{y-u_{t-j}}{h}\right)\right]=O\left(h^{-1}\right)$ and $\operatorname{Var}\left(\frac{\partial \tilde{\sigma}_{\theta}, j(x, y ; h)}{\partial \theta}\right)=O\left(h^{-1}(T-j)^{-1}\right)$ if the covariance terms do not contribute to this order. Compiling terms for bias
square and variance of $\frac{\partial \tilde{\sigma}_{\theta, j}(x, y ; h)}{\partial \theta}$, we have

$$
\begin{equation*}
\mathbb{E}\left[\left\|\frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y ; h)-\frac{\partial}{\partial \theta} \sigma_{\theta_{0}, j}(x, y)\right\|^{2}\right]=O\left(h^{4}+h^{-1}(T-j)^{-1}\right), \tag{2.7.10}
\end{equation*}
$$

which needs to be $o(1)$ to maintain the usual rate of convergence of the estimates if we have shown that the contribution of $\tilde{\sigma}_{\theta_{0}, j}(x, y ; h)$ is of order $T^{-1 / 2}$ for the classical CLT. This implies that,

$$
h^{4}+(T h)^{-1} \rightarrow 0 \quad \text { as } T \rightarrow \infty
$$

i.e. $h$ has to converge to zero when $T$ goes to infinity but slower than $T^{-1}$. Combining the rates obtained from (2.7.7) and (2.7.10), we conclude that, at least, the convergence rate of $h$ needs to satisfy

$$
\begin{equation*}
h^{-1} T^{-1}+h^{4} T \rightarrow 0, \text { as } T \rightarrow \infty, \tag{2.7.11}
\end{equation*}
$$

for the asymptotic normality of $\tilde{\theta}_{T}^{h}$ at a $T^{1 / 2}$ rate as $h \rightarrow 0$.
With all these established, we can show that, with the convergence rate of $h$ being (2.7.11),

$$
\begin{align*}
& \frac{\partial}{\partial \theta} \tilde{\sigma}_{\theta_{0}, j}(x, y ; h) \\
& \rightarrow_{p}-\phi_{j}^{(1)}\left(\theta_{0}\right) \mathbb{E}(u I(u \leq y)) f(x)-\phi_{-j}^{(1)}\left(\theta_{0}\right) \mathbb{E}(u I(u \leq x)) f(y)  \tag{2.7.12}\\
&=-\phi_{j}^{(1)}\left(\theta_{0}\right) \mu(y) f(x)-\phi_{-j}^{(1)}\left(\theta_{0}\right) \mu(x) f(y)=\frac{\partial}{\partial \theta} \sigma_{\theta_{0}, j}(x, y),
\end{align*}
$$

which coincides with the limit obtained using "generalized" derivatives of the indicator function, see equation (2.7.12).
Next, we check the score $\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)$,

$$
\frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right)=-\frac{4}{\pi^{2}} \frac{1}{T} \sum_{t=2}^{T} e_{t} V_{t-1}-\frac{4}{\pi^{2}} \frac{1}{T} \sum_{t=2}^{T} \nu_{t} E_{t-1}+o_{p}\left(T^{-1 / 2}\right) .
$$

Then, by applying CLT on $e_{t} V_{t-1}+\nu_{t} E_{t-1}$, we have

$$
\begin{equation*}
T^{1 / 2} \frac{\partial}{\partial \theta} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right) \rightarrow_{p} \mathcal{N}\left(0, \frac{16}{\pi^{4}} H_{1}\right), \text { as } h^{-1} T^{-1}+h^{4} T \rightarrow 0 \tag{2.7.13}
\end{equation*}
$$

where $H_{1}:=\sigma_{e}^{2} \sigma_{\nu}^{2}\left(\Sigma_{0,2}+\Sigma_{0,2}^{*}\right)+\sigma_{e, \nu}\left(\Sigma_{0,2}^{\dagger}+\Sigma_{0,2}^{\dagger^{\prime}}\right)$ with $\left\{\sigma_{e}^{2}, \sigma_{\nu}^{2}, \sigma_{e, \nu}\right\}$ being the limits of $\left\{\sigma_{e ; h}^{2}, \sigma_{\nu ; h}^{2}, \sigma_{e, \nu ; h}\right\}$.

Similarly, regarding the Hessian matrix, we have

$$
\begin{gather*}
\frac{\partial^{2}}{\partial \theta \partial \theta^{\prime}} \tilde{\mathcal{Q}}_{T}\left(\theta_{0} ; h\right) \rightarrow_{p} \frac{4}{\pi^{2}}\left(\rho_{1} \rho_{2}\left(\Sigma_{0,1}+\Sigma_{0,1}^{*}\right)+\left(\rho_{12}\right)^{2}\left(\Sigma_{0,1}^{\dagger}+\Sigma_{0,1}^{\dagger^{\prime}}\right)\right)=\frac{4}{\pi^{2}} H_{0}  \tag{2.7.14}\\
\text { as } h^{-1} T^{-1}+h^{4} T \rightarrow 0,
\end{gather*}
$$

and the theorem follows from (2.7.13) and (2.7.14).

### 2.8. Appendix B: Numerical Calculus

### 2.8.1. Calculations of Sample Objective Function

The sample objective function based on the indicator function (2.3.2) can be computed as follows:

$$
\hat{\mathcal{Q}}_{T}(\theta)=\frac{2}{\pi^{2}} \sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right) \frac{1}{j^{2}} \int_{\mathbb{R}^{2}} \hat{\sigma}_{\theta, j}^{2}(x, y) d W(x, y)
$$

where

$$
\begin{gathered}
\hat{\sigma}_{\theta, j}^{2}(x, y)=\hat{F}_{\theta, j}^{2}(x, y)-2 \hat{F}_{\theta, j}(x, y) \hat{F}_{\theta, j}(x, \infty) \hat{F}_{\theta, j}(\infty, y)+\hat{F}_{\theta, j}^{2}(x, \infty) \hat{F}_{\theta, j}^{2}(\infty, y), \\
\hat{F}_{\theta, j}(x, y)=\frac{1}{T-j} \sum_{t=j+1}^{T} I\left(u_{t}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right)
\end{gathered}
$$

The integration of $\hat{\sigma}_{\theta, j}^{2}(x, y)$ over weighting function $W$ can be decomposed into following three components:

$$
\begin{aligned}
& \int \hat{F}_{\theta, j}^{2}(x, y) d W(x, y) \\
= & \iint \frac{1}{(T-j)^{2}} \sum_{t=j+1}^{T} \sum_{s=j+1}^{T} I\left(u_{t}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right) I\left(u_{s}(\theta) \leq x\right) I\left(u_{s-j}(\theta) \leq y\right) d W(x) d W(y) \\
= & \frac{1}{(T-j)^{2}} \sum_{t=j+1}^{T} \sum_{s=j+1}^{T} \int I\left(u_{t}(\theta) \leq x\right) I\left(u_{s}(\theta) \leq x\right) d W(x) \int I\left(u_{t-j}(\theta) \leq y\right) I\left(u_{s-j}(\theta) \leq y\right) d W(y) \\
= & \frac{1}{(T-j)^{2}} \sum_{t=j+1}^{T} \sum_{s=j+1}^{T} \int_{\max \left\{u_{t}(\theta), u_{s}(\theta)\right\}}^{\infty} d W(x) \int_{\max \left\{u_{t-j}(\theta), u_{s-j}(\theta)\right\}}^{\infty} d W(y) \\
= & \frac{1}{(T-j)^{2}} \sum_{t=j+1}^{T} \sum_{s=j+1}^{T}\left(1-W\left(\max \left\{u_{t}(\theta), u_{s}(\theta)\right\}\right)\right)\left(1-W\left(\max \left\{u_{t-j}(\theta), u_{s-j}(\theta)\right\}\right)\right)
\end{aligned}
$$

where W is set to be a continuous probability measure.
Similarly,

$$
\begin{aligned}
& \int \hat{F}_{\theta, j}(x, y) \hat{F}_{\theta, j}(x, \infty) \hat{F}_{\theta, j}(\infty, y) d W(x, y) \\
&= \iint \frac{1}{(T-j)^{3}} \sum_{t=j+1}^{T} \sum_{s=j+1}^{T} \sum_{r=j+1}^{T} I\left(u_{t}(\theta) \leq x\right) I\left(u_{t-j}(\theta) \leq y\right) I\left(u_{s}(\theta) \leq x\right) I\left(u_{r-j}(\theta) \leq y\right) d W(x) d W(y) \\
&= \frac{1}{(T-j)^{3}} \sum_{t=j+1}^{T} \sum_{s=j+1}^{T} \sum_{r=j+1}^{T} \int I\left(u_{t}(\theta) \leq x\right) I\left(u_{s}(\theta) \leq x\right) d W(x) \int I\left(u_{t-j}(\theta) \leq y\right) I\left(u_{r-j}(\theta) \leq y\right) d W(y) \\
&= \frac{1}{(T-j)^{3}} \sum_{t=j+1}^{T} \sum_{s=j+1}^{T} \sum_{r=j+1}^{T}\left(1-W\left(\max \left\{u_{t}(\theta), u_{s}(\theta)\right\}\right)\right)\left(1-W\left(\max \left\{u_{t-j}(\theta), u_{r-j}(\theta)\right\}\right)\right) \\
& \text { and }
\end{aligned}
$$

$$
\begin{aligned}
& \int \hat{F}_{\theta, j}^{2}(x, \infty) \hat{F}_{\theta, j}^{2}(\infty, y) d W(x, y) \\
= & \frac{1}{(T-j)^{4}} \sum_{t=j+1}^{T} \sum_{s=j+1}^{T} \sum_{m=j+1}^{T} \sum_{n=j+1}^{T} \iint I\left(u_{t}(\theta) \leq x\right) I\left(u_{s}(\theta) \leq x\right) I\left(u_{m-j}(\theta) \leq y\right) I\left(u_{n-j}(\theta) \leq y\right) d W(x, y) \\
= & \frac{1}{(T-j)^{4}} \sum_{t=j+1}^{T} \sum_{s=j+1}^{T} \sum_{m=j+1}^{T} \sum_{n=j+1}^{T}\left(1-W\left(\max \left\{u_{t}(\theta), u_{s}(\theta)\right\}\right)\right)\left(1-W\left(\max \left\{u_{m-j}(\theta), u_{n-j}(\theta)\right\}\right)\right)
\end{aligned}
$$

## 3. CHAPTER III: DIRECTIONAL PREDICTABILITY TESTS

### 3.1. Introduction

Testing predictability of observed time series and model residuals is fundamental in economic analysis, as in the evaluation of asset pricing models and investment strategies, and in general dynamic modeling. This problem is especially difficult for many financial asset returns which lack strong serial correlation but can be affected by nonlinear dependence.

While consistent tests of predictability directed against general nonparametric alternatives have now a long tradition under the form of martingale difference testing (e.g. Hong (1999), Hong and Lee (2005), Escanciano and Velasco (2006a,b)) in many cases only some specific, typically linear alternatives are considered under a Gaussian framework. These methods include many popular white noise residual tests used to check linear model specifications such as Box and Pierce (1970), Breusch (1978) and Godfrey (1978). In parallel, tests of white noise have been also proposed (Lobato et al. (2002), Nankervis and Savin (2010), Shao (2011)) for which nonlinear predictability and higher-order dependence under non-Gaussianity are allowed, but not modeled.

Lanne and Luoto (2013) investigate tests of predictability based on checking the independence hypothesis against general (linear) ARMA or (nonlinear) all-pass dependence. They also check for general linear ARMA dependence under the null of observed series following an uncorrelated all-pass process based on independent and identically distributed (iid) innovations. Imposing non-invertibility in the ARMA model they solve the identification problem under the null discussed in Andrews and Ploberger (1996) and Nankervis and Savin (2010) as AR and MA filters do not cancel out but lead to an all-pass filter. All-pass rational filters, with causal autoregressive roots being the reciprocal of the non-invertible moving average ones, do not induce further linear dependence, but introduce nonlinear predictability on levels as well as on higher-order moments. Therefore, they are specially well suited to describe the behavior of uncorrelated but possibly dependent and predictable time series.

In fact, Lanne et al. (2013) tests can be regarded as either tests of the null hypothesis of serial independence against two specific linear parameterizations (general ARMA or restricted all-pass ARMA), or tests of linear unpredictability (described by the restrictive class of all-pass ARMA models with iid innovations) against general

ARMA linear dependence. Concentrating on these specific parametric models can enhance the performance of tests in many practical situations but the methodology of Lanne et al. (2013) based on maximum likelihood estimation for iid innovations with known non-Gaussian distribution is based on very strong assumptions, see Lanne and Saikkonen (2011) and Lii and Rosenblatt (1992, 1996). Apart from potential specification errors, their ML methods do not account for the possible higher-order dependence introduced in the series by unpredictable but non-independent innovations and their interaction with all-pass filters.

Following a similar strategy, in this paper, we propose new predictability directional tests extending the analysis of Lanne et al. (2013) in several directions to alleviate the above limitations. The most relevant one from both the methodological and applied perspectives is the consideration of hypothesis tests that are robust to higher-order dependence in the sequence of innovations, so we focus on the more general martingale difference null hypothesis or on the wider class of white noise processes generated by all-pass models driven by such unpredictable, but not serially independent, innovation sequences.

To justify the consistency of our predictability tests against all-pass alternatives we also investigate the general dependence that all-pass filtering can generate. In particular, we show that ARMA all-pass processes with martingale difference innovations, despite remaining serially uncorrelated, are non-linearly predictable in the same way as has been showed by Rosenblatt (2000) for serially independent innovations-driven processes. We obtain this result by characterizing the form of the dynamic higher-order moments of such ARMA processes with innovations which are martingale differences, possibly with dynamics in conditional second and third-order moments, extending the seminal analysis of Rao and Gabr (1980) and Hinich (1982) of the trispectrum of linear processes with iid errors.

When searching for linear predictability as a deviation of an all-pass ARMA model, the dependence introduced in higher-order moments by the all-pass filters (despite they do not affect the uncorrelation of levels) interacts with the one that might be already present in martingale difference innovations due to e.g. dynamic conditional heteroskedasticity. Our tests for linear predictability account properly for these general all-pass dynamics in uncorrelated observed series displaying conditional heteroskedasticity and other forms of nonlinear dependence.

Unlike Lanne et al. (2013), our methodology is not based on maximum likelihood analysis relying on the iid assumption of innovations but on pairwise dependence measures checking for the martingale difference property on observations (or model residuals). Specifically, we propose Lagrange Multiplier (LM) tests based on a discrepancy measure which accounts for higher-order dependence in the martingale difference innovations (and in the preliminary parameter estimation under the null
of all-pass dependence) developed in Velasco (2022) starting from Hong (1999) general dependence tests.

Our dependence analysis allows us to prove the consistency of our unpredictability test against all-pass alternatives which only introduce non-linear predictability, as well as against alternatives inducing linear correlation for non-Gaussian innovations. Our approach is an alternative to the ones based on other dependence measures defined using the joint cumulative distribution function (Jin (2022)) or higher-order moments in either the time (Ramsey and Montenegro (1992)) or frequency domains (Velasco and Lobato (2018)).

We also consider higher order models beyond the $\operatorname{ARMA}(1,1)$ model discussed in Lanne et al. (2013) and also compare with iid testing using dependence measures in our simulation exercises. Further, we investigate tests to detect non-linear dependence in model residuals by checking for an additional all-pass factor in the linear ARMA model. Extensions to non-causal models, which can generate also nonlinear dependence, see e.g. Gouriéroux and Zakoïan (2017), are straightforward under our framework.

The paper is organized as follows. In Section 3.2 we investigate the dependence induced by all-pass filtering of martingale difference processes in higher-order moments, and in particular, in squares dependence, extending the analysis in Lanne et al. (2013). In Section 3.3 we describe the hypothesis tested and the asymptotic properties of our tests of dependence in the observed series. Section 3.4 investigates a test to detect all-pass structure in an ARMA model using residuals from a causal and invertible estimation. In Section 3.5 we report the results of a simulation exercise, while Section 3.6 contains an empirical analysis of predictability of a series of returns.

### 3.2. Predictability, linear and nonlinear dependence

In this section, we first show that $\operatorname{ARMA}(p, p)$ processes generated by all-pass filtered martingale difference sequences ( $m d s$ ), despite remaining uncorrelated processes, are no longer $m d s$, i.e. their levels can be predicted by nonlinear functions of the past, as has been showed by (Rosenblatt, 2000, Section 5.4) for iid innovations-driven allpass models. In particular, we show that they display pairwise dependence, which motivates our dependence tests based on pairwise measures derived from bivariate characteristic functions, see also Lemma 3 in Velasco (2022). Second, we generalize the result of Lanne et al. (2013) on the higher order dependence generated in ARMA $(1,1)$ models by non-invertible all-pass filters and iid innovations, to general order $p>1$ models. Then, we show that the squares of observed series follow a $\operatorname{ARMA}(p, p)$ model which implies that levels have a weak GARCH representation.

### 3.2.1. Predictability introduced by all-pass filters on mds sequences

Consider the all-pass $\operatorname{ARMA}(p, p)$ process $X_{t}$ generated by the $m d s$ innovations $\varepsilon_{t}$ satisfying $\mathbb{E}\left[\varepsilon_{t} \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\right]=0$ a.s.,

$$
\begin{equation*}
X_{t}=\psi(L) \varepsilon_{t} \tag{3.2.1}
\end{equation*}
$$

where $\psi$ is the transfer function of a causal and non-invertible all-pass filter of order p,

$$
\psi(z)=\sum_{j=0}^{\infty} \psi_{j} z^{j}=\prod_{j=1}^{p} \frac{1-m_{j}^{-1} z}{1-m_{j} z},
$$

with $\left|m_{j}\right|<1, j=1, \ldots, p, \psi_{0}=1$ and $\psi_{k} \sim \sum_{j=1}^{p} c_{j} m_{j}^{k}$ as $k \rightarrow \infty$ for some $c_{j} \neq 0$. In particular, for $p=1$, note that $\psi_{0}=1$ and $\psi_{k}=m_{1}^{k}\left(1-m_{1}^{-2}\right), k>0$.

Then the second-order spectral density of the all-pass process $X_{t}, f_{2}^{X}(\lambda)=$ $\left(\sigma_{\varepsilon}^{2} / 2 \pi\right)|\psi(\lambda)|^{2}=\sigma_{\varepsilon}^{2} /(2 \pi)$ is constant redefining $\psi(\lambda)=\psi\left(e^{-i \lambda}\right)$ while the thirdorder spectral density of $X_{t}$ can be written in terms of the third-order spectral density of $\varepsilon_{t}$ as

$$
f_{3}^{X}(\lambda)=f_{3}^{\varepsilon}(\lambda) \psi\left(\lambda_{1}\right) \psi\left(\lambda_{2}\right) \psi\left(-\lambda_{1}-\lambda_{2}\right), \quad \lambda=\left(\lambda_{1}, \lambda_{2}\right) .
$$

To evaluate

$$
f_{3}^{\varepsilon}(\lambda)=\frac{1}{(2 \pi)^{2}} \sum_{j, k=-\infty}^{\infty} \kappa_{3}^{\varepsilon}(j, k) \exp \left(-i\left(j \lambda_{1}+k \lambda_{2}\right)\right)
$$

in terms of the joint skewness coefficients of the zero mean (third-order) stationary sequence $\varepsilon_{t}$

$$
\kappa_{3}^{\varepsilon}(j, k)=\kappa_{3}^{\varepsilon}(k, j)=\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+j} \varepsilon_{t+k}\right], \quad j, k=0, \pm 1, \ldots,
$$

we note that for a $m d s \varepsilon_{t}$ the only joint third-moments that could be different from zero are of the form

$$
\begin{array}{rlrl}
\kappa_{3}^{\varepsilon}(j, j) & =\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+j}^{2}\right], & & j \geq 0 \\
\kappa_{3}^{\varepsilon}(j, 0) & =\mathbb{E}\left[\varepsilon_{t+j} \varepsilon_{t}^{2}\right], & j \leq 0 \\
\kappa_{3}^{\varepsilon}(0, k) & =\mathbb{E}\left[\varepsilon_{t+k} \varepsilon_{t}^{2}\right], & & k \leq 0
\end{array}
$$

where $\kappa_{3}^{\varepsilon}(j, j)=\kappa_{3}^{\varepsilon}(-j, 0)=\kappa_{3}^{\varepsilon}(0,-j)$ for $j \geq 0$, because the squares $\varepsilon_{t}^{2}$ could be predicted by past observations. However, levels can not be predicted by the past, implying that

$$
\kappa_{3}^{\varepsilon}(j, k)=\left\{\begin{array}{lc}
\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+j} \varepsilon_{t+k}\right]=0, & \text { when } \max \{j, k\}>0 \text { and } j \neq k  \tag{3.2.2}\\
\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+j} \varepsilon_{t+k}\right]=0, \quad \text { when } \max \{j, k\}<0 .
\end{array}\right.
$$

Therefore, $f_{3}^{X}(\lambda)$ is nonconstant and nonzero as far as $f_{3}^{\varepsilon}(\lambda)$ is not zero a.e. because the $m d s$ innovation process $\varepsilon_{t}$ displays some nonzero third-order moments. Then, some coefficients

$$
\kappa_{3}^{X}(j, k)=\int_{[-\pi, \pi]^{2}} f_{3}^{X}(\lambda) \exp \left(i\left(j \lambda_{1}+k \lambda_{2}\right)\right) d \lambda_{1} d \lambda_{2}
$$

are different from zero for $|j|+|k|>0$, implying that the sequence $X_{t}$ is dependent. In particular, next theorem shows that the skewness coefficients of $X_{t}$ are in general incompatible with those of a $m d s$ process as characterized in (3.2.2). Since $X_{t}$ is a white noise process, it is linearly unpredictable and to show that $X_{t}$ is not a mds we need to argue that $X_{t}$ is nonlinearly predictable by arguing that $\kappa_{3}^{X}(j, k)=$ $\mathbb{E}\left[X_{t} X_{t+j} X_{t+k}\right] \neq 0$ for some $j, k<0$. For showing $X_{t}$ is not a pairwise martingale, i.e. $\mathbb{E}\left[\varepsilon_{t} \mid \varepsilon_{t-j}\right] \neq 0$ for some $j>0$, we need to check that $\kappa_{3}^{X}(j, j)=\mathbb{E}\left[X_{t} X_{t+j}^{2}\right] \neq$ 0 for some $j<0$. However, if $\kappa_{3}^{X}(j, j)=\mathbb{E}\left[X_{t} X_{t+j}^{2}\right] \neq 0, \quad j>0$, this would imply that $X_{t}$ has dynamic conditional heteroskedasticity but would not rule out that $X_{t}$ is a $m d s$.

Theorem 3.2.1. For a mds $\varepsilon_{t}$ with $\mathbb{E}\left|\varepsilon_{t}\right|^{3}<\infty$ and $\kappa_{3}^{\varepsilon}(0,0) \neq 0$, the process $X_{t}$ defined in (3.2.1) has third-order cumulants given by

$$
\kappa_{3}^{X}(j, k)=\kappa_{X, 3}^{[i i d]}(j, k)+\sum_{m=1}^{\infty} \kappa_{X, 3}^{[m]}(j, k), \quad j, k=0, \pm 1, \ldots
$$

where the contribution from the marginal third cumulant of $\varepsilon_{t}$ is

$$
\kappa_{X, 3}^{[i i d]}(j, k)=\kappa_{3}^{\varepsilon}(0,0) \sum_{c=\max \{0,-j,-k\}}^{\infty} \psi_{c+j} \psi_{c+k} \psi_{c}
$$

and the potential contribution from joint cumulants of $\varepsilon_{t}$ for $m=1,2, \ldots$, is

$$
\kappa_{X, 3}^{[m]}(j, k)=\frac{\kappa_{3}^{\varepsilon}(m, m)}{\kappa_{3}^{\varepsilon}(0,0)}\left\{\kappa_{X}^{[i i d]}(j-m, k-m)+\kappa_{X}^{[i i d]}(j+m, k)+\kappa_{X}^{[i i d]}(j, k+m)\right\} .
$$

All proofs of this section are contained in Appendix A. In particular for $j>0$ and $k=0$, this theorem shows that $\kappa_{3}^{X}(j, 0)=\kappa_{3}^{X}(-j,-j)$,

$$
\kappa_{3}^{X}(j, 0)=\kappa_{X, 3}^{[i i d]}(j, 0)+\sum_{m=1}^{\infty} \kappa_{X, 3}^{[m]}(j, 0), \quad \kappa_{X, 3}^{[i i d]}(j, 0)=\kappa_{3}^{\varepsilon}(0,0) \sum_{c=0}^{\infty} \psi_{c}^{2} \psi_{c+j}
$$

is expected to be different from zero for infinitely many $j>0$, confirming that $X_{t}$ is not a pairwise $m d s$ as when $\varepsilon_{t}$ is $i i d$, in which case $\kappa_{X, 3}^{[m]}(j, 0)=0$ for $m>0$ but $\kappa_{3}^{X}(j, 0)=\kappa_{X, 3}^{[i i d]}(j, 0) \neq 0$ providing an alternative pairwise argument for Rosenblatt (2000) analysis of predictability of all-pass processes under iid.

We now repeat the analysis for fourth order dynamics in terms of the joint kurtosis coefficient for a zero mean (fourth-order) stationary sequence $\varepsilon_{t}$ as

$$
\kappa_{4}^{\varepsilon}(j, k, \ell)=\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+j} \varepsilon_{t+k} \varepsilon_{t+\ell}\right]-\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+j}\right] \mathbb{E}\left[\varepsilon_{t+k} \varepsilon_{t+\ell}\right]-\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+\ell}\right] \mathbb{E}\left[\varepsilon_{t+j} \varepsilon_{t+k}\right]-\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+k}\right] \mathbb{E}\left[\varepsilon_{t+j} \varepsilon_{t+\ell}\right] .
$$

For a stationary $m d s \varepsilon_{t}$ the joint fourth cumulants that could be different from zero apart from $\kappa_{4}^{\varepsilon}:=\kappa_{4}^{\varepsilon}(0,0,0)$ are derived from the fact that future $\varepsilon_{t+\ell}^{3}$ and $\varepsilon_{t+\ell}^{2}$ could be predicted by the past levels $\varepsilon_{t}$ and squares $\varepsilon_{t}^{2}$ for some $\ell>0$.

This leads to the following possibly different from zero joint cumulants of the $m d s \varepsilon_{t}$, apart from the marginal one $\kappa_{4}^{\varepsilon}(0,0,0)$. First,

$$
\begin{array}{rll}
\kappa_{4}^{\varepsilon}(j, j, j) & =\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+j}^{3}\right], & j>0 \\
\kappa_{4}^{\varepsilon}(j, 0,0) & =\mathbb{E}\left[\varepsilon_{t+j} \varepsilon_{t}^{3}\right], & j<0 \\
\kappa_{4}^{\varepsilon}(0, k, 0) & =\mathbb{E}\left[\varepsilon_{t+k} \varepsilon_{t}^{3}\right], & \\
\kappa_{4}^{\varepsilon}(0,0, \ell) & =\mathbb{E}\left[\varepsilon_{t+\ell} \varepsilon_{t}^{3}\right], & \ell<0,
\end{array}
$$

where $\kappa_{4}^{\varepsilon}(j, j, j)=\kappa_{4}^{\varepsilon}(-j, 0,0)=\kappa_{4}^{\varepsilon}(0,-j, 0)=\kappa_{4}^{\varepsilon}(0,0,-j)$ for $j \geq 0$, because cubes could be predicted by past observations, despite levels can not, and second,

$$
\begin{array}{ll}
\kappa_{4}^{\varepsilon}(j, j, k)=\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+k} \varepsilon_{t+j}^{2}\right]-\sigma^{4} 1\{k=0\}, & j \geq 0, k<j \\
\kappa_{4}^{\varepsilon}(j, k, j)=\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+k} \varepsilon_{t+j}^{2}\right]-\sigma^{4} 1\{k=0\}, \quad j \geq 0, k<j \\
\kappa_{4}^{\varepsilon}(k, j, j)=\mathbb{E}\left[\varepsilon_{t} \varepsilon_{t+k} \varepsilon_{t+j}^{2}\right]-\sigma^{4} 1\{k=0\}, \quad j \geq 0, k<j
\end{array}
$$

where $\kappa_{4}^{\varepsilon}(j, j, k)=\kappa_{4}^{\varepsilon}(0,-j, k-j)$, because squares could be predicted by past observations. To investigate pairwise dependence we focus on the case $k=0$ and $j>0$

$$
\begin{array}{ll}
\kappa_{4}^{\varepsilon}(j, j, 0) & =\mathbb{E}\left[\varepsilon_{t}^{2} \varepsilon_{t+j}^{2}\right]-\sigma^{4}, \\
\kappa_{4}^{\varepsilon}(j, 0, j)=\mathbb{E}\left[\varepsilon_{t}^{2} \varepsilon_{t+j}^{2}\right]-\sigma^{4}, & j>0 \\
\kappa_{4}^{\varepsilon}(0, j, j)=\mathbb{E}\left[\varepsilon_{t}^{2} \varepsilon_{t+j}^{2}\right]-\sigma^{4}, & j>0,
\end{array}
$$

where $\kappa_{4}^{\varepsilon}(j, j, 0)=\kappa_{4}^{\varepsilon}(j, 0, j)=\kappa_{4}^{\varepsilon}(0, j, j)=\kappa_{4}^{\varepsilon}(-j,-j, 0)=\kappa_{4}^{\varepsilon}(-j, 0,-j)=$ $\kappa_{4}^{\varepsilon}(0,-j,-j)$ for $j \geq 0$, because squares could also be predicted by the squares of a single past observation.

Therefore, the fourth order spectral density of $X_{t}$,

$$
f_{4}^{X}(\lambda)=f_{4}^{\varepsilon}(\lambda) \psi\left(\lambda_{1}\right) \psi\left(\lambda_{2}\right) \psi\left(\lambda_{3}\right) \psi\left(-\lambda_{1}-\lambda_{2}-\lambda_{3}\right),
$$

is non-constant and nonzero if some kurtosis coefficients of $\varepsilon_{t}$ are nonzero, so that
some $\kappa_{4}^{X}(j, k, \ell)$ are different from zero for some $|j|+|k|+|\ell|>0$, implying that the sequence $X_{t}$ is dependent. Since $X_{t}$ is not linearly predictable, to show that $X_{t}$ is not a $m d s$ we need to show that $X_{t}$ is nonlinearly predictable by checking that $\kappa_{4}^{X}(j, k, \ell) \neq 0$ for some $j, k, \ell<0$ using the characterization provided by our next result. Further, to argue that $X_{t}$ is not a pairwise $m d s$, we can show that $\kappa_{4}^{X}(j, j, j) \neq 0$, for some $j<0$, i.e. $X_{t}$ is predictable by a nonlinear function of a single observation of the past, $X_{t+j}, j<0$, in this case the cubic function, $X_{t+j}^{3}$.

Theorem 3.2.2. For a mds $\varepsilon_{t}$ with $\mathbb{E}\left|\varepsilon_{t}\right|^{4}<\infty$ and $\kappa_{4}^{\varepsilon}(0,0,0) \neq 0$, the process $X_{t}$ defined in (3.2.1) has fourth-order cumulants given by

$$
\begin{aligned}
\kappa_{4}^{X}(j, k, \ell)= & \kappa_{X, 4}^{[i i d]}(j, k, \ell)+\sum_{m=1}^{\infty} \kappa_{X, 4}^{[1, m]}(j, k, \ell) \\
& +\sum_{m=1}^{\infty} \sum_{n=-\infty}^{m-1} \kappa_{X, 4}^{[2, m, n]}(j, k, \ell)+\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \kappa_{X, 4}^{[3, m, n]}(j, k, \ell)
\end{aligned}
$$

where

$$
\kappa_{4, X}^{[i i d]}(j, k, \ell)=\kappa_{4}^{\varepsilon}(0,0,0) \sum_{d=\max \{0,-j,-k,-\ell\}}^{\infty} \psi_{d+j} \psi_{d+k} \psi_{d+\ell} \psi_{d}
$$

and the potential additional terms due to predictability of squares are given in Appendix $A$.

This theorem shows in particular that $\kappa_{4}^{X}(j, 0,0)=\kappa_{4}^{X}(-j,-j,-j), j>0$, would not be all equal to zero for $\psi_{a}$ given by an all-pass $\operatorname{ARMA}(p, p)$ model, so that $X_{t}$ is not a pairwise $m d s$ unlike its innovations $\varepsilon_{t}$. If $\varepsilon_{t}$ were serially independent, then

$$
\kappa_{4}^{X}(j, 0,0)=\kappa_{4}^{\varepsilon}(0,0,0) \sum_{d=0}^{\infty} \psi_{d}^{3} \psi_{d+j}
$$

which for e.g. $p=1$ can be checked immediately to be non-zero for all $j>0$ when $m_{1} \neq 0$.

### 3.2.2. Higher order dependence introduced by all-pass filters

Theorem 2 can be recast for a description of the dependence introduced on the squares of an all-pass $\operatorname{ARMA}(p, p)$ process driven by iid innovations, as has been showed for $p=1$ by Lanne et al. (2013) and argued by Breidt, Davis and Trindade (2001).

Under iid we get at once that

$$
\kappa_{4}^{X}(j, k, \ell)=\kappa_{X, 4}^{[i i d]}(j, k, \ell)
$$

as now the $i i d$ process $\varepsilon_{t}$ cannot display higher dependence. Then, by uncorrelation
of $X_{t}$ we have that for $j>0$ the autocovariance function of $X_{t}^{2}$,

$$
\mathbb{C}\left(X_{t}^{2}, X_{t+j}^{2}\right)=\mathbb{E}\left[X_{t}^{2}, X_{t+j}^{2}\right]-\sigma_{X}^{4}=\kappa_{X, 4}^{[i i d]}(j, j, 0)=\kappa_{4}^{\varepsilon} \sum_{d=0}^{\infty} \psi_{d+j}^{2} \psi_{d}^{2}
$$

which behaves like the ACF of an $\mathrm{MA}(\infty)$ process with filter coefficients $\psi_{j}^{2}$, displaying positive autocorrelation at all lags. Furthermore, it can be showed that the process $X_{t}^{2}$ has an $\operatorname{ARMA}(p, p)$ representation by investigating the spectral density of the squares $X_{t}^{2}$.

Theorem 3.2.3. The process $X_{t}^{2}$, where $X_{t}$ is given in (3.2.1) for a zero mean iid sequence $\varepsilon_{t}$ with $\kappa_{4}^{\varepsilon}(0,0,0) \neq 0$, has a restricted $\operatorname{ARMA}(p, p)$ representation.

In particular, for $p=1$ and causal AR coefficient $\phi=m_{1},|\phi|<1$, the causal ARMA $(1,1)$ representation of the process $X_{t}^{2}$ has autoregressive coefficient given by $\phi^{2}$.

### 3.3. Directional Predictability Hypothesis Tests

In this section, we describe tests of directional predictability of observed series where the null hypothesis of unpredictability or $m d s$ is tested against linear alternatives allowing either for serial correlation (ARMA) or only for nonlinear predictability (all-pass), and tests of the restricted white noise all-pass ARMA model against a general ARMA one displaying serial correlation.

Lanne et al. (2013) considered the following parameterization of the noninvertible $\operatorname{ARMA}(1,1)$ model

$$
\begin{equation*}
Y_{t}=\phi_{0} Y_{t-1}+\varepsilon_{t-1}-\theta_{0} \varepsilon_{t} \tag{3.3.1}
\end{equation*}
$$

where $\left|\phi_{0}\right|<1,\left|\theta_{0}\right|<1$ and $\varepsilon_{t}$ is an uncorrelated error term with zero mean and finite variance $\sigma_{0}^{2}$. Then we can write

$$
\left(1-\phi_{0} L\right) Y_{t}=\left(1-\theta_{0} L^{-1}\right) L \varepsilon_{t}
$$

and when $\theta_{0} \neq 0$

$$
\left(1-\phi_{0} L\right) Y_{t}=\left(1-\theta_{0}^{-1} L\right)\left(-\theta_{0} \varepsilon_{t}\right)
$$

with $\left|\theta_{0}^{-1}\right|>1$ so the non-invertible MA root $\theta_{0}$ satisfies $\left|\theta_{0}\right|<1$, while if $\theta_{0}=0$ we can write $Y_{t}$ as an $\mathrm{AR}(1)$ model,

$$
\left(1-\phi_{0} L\right) Y_{t}=\varepsilon_{t-1}
$$

The AP model is obtained when $\phi_{0}=\theta_{0}$,

$$
\begin{equation*}
Y_{t}=\phi_{0} Y_{t-1}+\varepsilon_{t-1}-\phi_{0} \varepsilon_{t} \tag{3.3.2}
\end{equation*}
$$

where the AR and MA roots are reciprocals, $\phi_{0}^{-1}$ and $\phi_{0}$ respectively. In general we write

$$
Y_{t}=\psi\left(\vartheta_{0} ; L\right) L \varepsilon_{t}
$$

for $\psi(\vartheta ; z)=\left(1-\theta z^{-1}\right) /(1-\phi z)$ and $\vartheta=(\theta, \phi)^{\prime}$ under the ARMA model (3.3.1) and $\psi(\vartheta ; z)=\left(1-\phi z^{-1}\right) /(1-\phi z)$ and $\vartheta=\phi$ under the AP model (3.3.2).

While Lanne et al. (2013) considered that $\varepsilon_{t}$ is $i i d$, we are interested in this paper in $m d s$ versions of their hypotheses. Thus, we study testing the AP hypothesis with $m d s$ innovations against non-invertible ARMA

$$
\begin{aligned}
H_{A P / m d s} & : \phi_{0}=\theta_{0} \text { in model }(3.3 .1) \\
H_{A P / m d s}^{1} & : \phi_{0} \neq \theta_{0}
\end{aligned}
$$

the mds hypothesis against all-pass restricted $\operatorname{ARMA}(1,1)$

$$
\begin{aligned}
H_{m d s}^{(A P)} & : \quad \phi_{0}=0 \text { in model }(3.3 .2) \\
H_{m d s}^{(A P) 1} & : \quad \phi_{0} \neq 0
\end{aligned}
$$

or the mds hypothesis against unrestricted but non-invertible ARMA $(1,1)$

$$
\begin{aligned}
H_{m d s} & : \phi_{0}=\theta_{0}=0 \text { in model } \\
H_{m d s}^{1} & : \phi_{0} \neq 0 \text { and } / \text { or } \theta_{0} \neq 0
\end{aligned}
$$

where we keep the same notation as in Lanne et al. (2013) despite $H_{m d s}^{(A P)}$ and $H_{m d s}$ imply the same data generating process for $Y_{t}=\varepsilon_{t}$ (but not the alternatives $H_{m d s}^{(A P) 1}$ and $H_{m d s}^{1}$ unless $\phi_{0}=\theta_{0} \neq 0$ ).

These $m d s$ versions of the null hypothesis $H_{A P / m d s}, H_{m d s}^{(A P)}$ and $H_{m d s}$, allow for higher order dependence in $\varepsilon_{t}$. Thus, the AP null hypothesis, equivalent to model (3.3.2), implies that $Y_{t}$ is white noise but nonlinear dependent in levels, i.e. is predictable by Rosenblatt (2000) analysis for non-Gaussian iid $\varepsilon_{t}$ or by Section 2 results for $m d s \varepsilon_{t}$. It further implies that $Y_{t}^{2}$ are predictable even for $i i d \varepsilon_{t}$, apart from the additional dependence possibly introduced by the higher order dependence in the $m d s \varepsilon_{t}$. We consider the same parametric alternatives in the direction of linear models, so that our results in Section 2 show that under $H_{m d s}^{(A P) 1}$ the observed series is no longer a $m d s$, despite being white noise, while under $H_{A P / m d s}^{1}$ and $H_{m d s}^{1}$ it also displays nonzero autocorrelation, justifying the consistency of tests. For presentation we focus on the $\operatorname{ARMA}(1,1)$ model but we discuss the general case in

Appendix B and provide results for any fixed $p=1,2, \ldots$.
Lanne et al. (2013) propose testing the AP/iid hypothesis through a Wald test based on unrestricted estimation of the non-invertible ARMA $(1,1)$ model (3.3.1) using non-Gaussian ML estimation following Lanne et al. (2013) under non-invertibility (and causality), which includes the estimation of a scaling parameter together with additional shape parameters for the given (zero mean) distribution. Using also APconstrained ML estimation they further propose LR tests for this hypothesis.

For testing iid against AP (ARMA) dependence, Lanne et al. (2013) similarly propose Wald significance tests on the AP-restricted (unrestricted) estimate of $\phi_{0}$ ( $\phi_{0}$ and $\theta_{0}$ ) and LR tests which require to estimate the corresponding nuisance (scale and shape) parameters of the distribution of $\varepsilon_{t}$. In all cases standard asymptotics are obtained due to imposing non-invertibility and non-Gaussianity, unlike in the standard (causal and invertible) situation considered in Nankervis and Savin (2010), where a nuisance parameter is only present under the alternative which needs to be accounted for using the methods in Davies (1987) and Andrews and Ploberger (1996).

In this paper we use robust methods to test the $m d s$ versions of the three dependence hypothesis based on nonlinear dependence measures defined with the joint characteristic function of the residuals, accounting for non-Gaussianity and higher order dependence when $\varepsilon_{t}$ is a $m d s$, see e.g. Velasco (2022), though methods based on the cdf could also be used, see Jin (2022). This approach avoids imposing a particular non-Gaussian distribution and estimating of the corresponding shape parameters under serial independence, though versions of the dependence measures under independence can be also pursued (see Section 5 for a finite sample comparison).

Similarly to Lanne et al. (2013) our tests for the AP/mds hypothesis $H_{A P / m d s}$ against general ARMA dependence are based on AP-restricted estimation of the ARMA $(1,1)$ model imposing non-invertibility, while the tests for the $m d s$ hypotheses against general ARMA or all-pass restricted models, $H_{m d s}$ and $H_{m d s}^{(A P)}$ respectively, can be considered as directional variants of the $m d s$ version of Hong (1999) omnibus tests for first-order dependence based on the derivative of the characteristic function. This approach improves the rate of convergence of test statistics (and would allow to detect alternatives departing from the null at the usual parametric rate), but at the cost of having less power against alternatives departing from the null in other directions.

### 3.3.1. Tests of the $m d s$ hypothesis

Our robust LM tests are derived from the following objective function based on the generalized spectral density function

$$
Q_{T}^{m d s}(\vartheta):=\frac{2}{\pi} \sum_{j=1}^{T-1} k^{2}\left(\frac{j}{m}\right)\left(1-\frac{|j|}{T}\right) \int\left|\hat{\sigma}_{\vartheta, j}^{(1,0)}(0, v)\right|^{2} d W(v)
$$

of residuals

$$
\hat{\varepsilon}_{t}(\vartheta)=\psi^{-1}(\vartheta ; L)\left(Y_{t}-\bar{Y}_{T}\right) 1\{1 \leq t \leq T\}=\frac{1-\phi L}{1-\theta L^{-1}}\left(Y_{t}-\bar{Y}_{T}\right) 1\{1 \leq t \leq T\}
$$

where $\vartheta:=(\phi, \theta)^{\prime}$ for $\operatorname{ARMA}(1,1)$ model and $\vartheta:=\phi$ for $\operatorname{AP}(1,1)$ model, and

$$
\hat{\sigma}_{\vartheta, j}^{(1,0)}(0, v):=\left.\frac{\partial}{\partial u} \hat{\sigma}_{\vartheta, j}(u, v)\right|_{u=0}
$$

is the derivative of the residuals' generalized autocovariance function

$$
\hat{\sigma}_{\vartheta, j}(u, v)=\hat{\varphi}_{\vartheta, j}(u, v)-\hat{\varphi}_{\vartheta, j}(u, 0) \hat{\varphi}_{\vartheta, j}(0, v),
$$

based on their empirical pairwise joint characteristic function

$$
\hat{\varphi}_{\vartheta, j}(u, v)=\frac{1}{T-j} \sum_{t=1+j}^{T} \exp \left\{i u \hat{\varepsilon}_{t}(\vartheta)+i v \hat{\varepsilon}_{t-j}(\vartheta)\right\} .
$$

The kernel function $k$ with $k(0)=1$ together with the lag $m \rightarrow \infty$ as $T \rightarrow \infty$ guarantees that asymptotically dependence at all lags is considered in the $L_{2}$ distance $Q_{T}^{m d s}(\vartheta)$ while the weighting measure $W$ aggregates information for all $v \in \mathbb{R}$. Note that we have relabel the time index of the residuals to match that of the observations without loss of generality.

The $m d s$ hypothesis can be tested against the two given parametric alternatives with LM statistics of the form

$$
L M_{T}=T \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right)^{\prime} \widehat{A V a r}\left(T^{1 / 2} \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right)\right)^{-1} \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right)
$$

where $\vartheta_{0}=\phi_{0}=0$ when testing against AP and $\vartheta_{0}=\left(\phi_{0}, \theta_{0}\right)^{\prime}=(0,0)^{\prime}$ when testing against ARMA, so that inference could be conducted in the usual way as there are no nuisance parameters under either null.

In practice we propose to use approximations for the score function of $Q_{T}^{m d s}(\vartheta)$,

$$
\frac{\partial}{\partial \vartheta} Q_{T}^{m d s}(\vartheta)=\frac{2}{\pi} \sum_{j=1}^{T-1} k^{2}\left(\frac{j}{m}\right)\left(1-\frac{|j|}{T}\right) \int 2 \operatorname{Re}\left\{\hat{\sigma}_{\vartheta, j}^{(1,0)}(0, v) \frac{\partial}{\partial \vartheta} \overline{\hat{\sigma}_{\vartheta, j}^{(1,0)}(0, v)}\right\} d W(v)
$$

derived in Velasco (2022) under the null $\vartheta=\vartheta_{0}$,

$$
\frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right)=\frac{1}{T} \sum_{t=2}^{T} \varepsilon_{t} Z_{t-1}\left(\vartheta_{0}\right)+o_{p}\left(T^{-1 / 2}\right)
$$

for $Z_{t-1}\left(\vartheta_{0}\right):=R_{t-1}\left(\vartheta_{0}\right)+S_{t-1}\left(\vartheta_{0}\right)$, where $z_{t}^{0}=z_{t}\left(\vartheta_{0} ; v\right)$ for $z_{t}(\vartheta ; v):=e^{i u \varepsilon_{t}(\vartheta)}-$ $\varphi_{\vartheta}(v), \varphi_{\vartheta}(v):=\mathbb{E}\left[e^{i v \varepsilon_{t}(\vartheta)}\right], R_{t-1}\left(\vartheta_{0}\right)$ and $S_{t-1}\left(\vartheta_{0}\right)$ are given by

$$
R_{t-1}\left(\vartheta_{0}\right):=\sum_{j=1}^{\infty} i \int z_{t-j}^{0}(v) \zeta_{j}^{0}(-v) d W(v), \quad \zeta_{j}^{0}(v):=-\sum_{n=j}^{\infty} \delta_{n}\left(\vartheta_{0}\right) \varphi_{j-n}^{(1,0)}(0, v)
$$

noting that for $\varphi_{j}(u, v)=\mathbb{E}\left[e^{i u \varepsilon_{t}} e^{i v \varepsilon_{t-j}}\right], \varphi_{j}^{(1,0)}(0, v)=\left.(\partial / \partial u) \varphi_{j}(u, v)\right|_{u=0}=$ $i \mathbb{E}\left[\varepsilon_{t} e^{i v \varepsilon_{t-j}}\right]$, and by

$$
S_{t-1}\left(\vartheta_{0}\right):=\sum_{j=1}^{\infty} i \int z_{t-j}^{0}(v) \beta_{j}^{0}(-v) d W(v), \quad \beta_{j}^{0}(v):=-\delta_{-j}\left(\vartheta_{0}\right) v \varphi_{j}^{(2,0)}(0, v),
$$

for $\varphi_{j}^{(2,0)}(0, v)=\left.\left(\partial^{2} / \partial u^{2}\right) \varphi_{j}(u, v)\right|_{u=0}=-\mathbb{E}\left[\varepsilon_{t}^{2} e^{i v \varepsilon_{t-j}}\right]$. The coefficients $\delta_{j}(\vartheta)$ are defined by

$$
\delta(\vartheta ; z):=-\frac{\partial}{\partial \vartheta} \log \psi(\vartheta ; z)=\sum_{j=-\infty}^{\infty} \delta_{j}(\vartheta) z^{j}
$$

with e.g. for $p=1$

$$
\begin{aligned}
\psi\left(\vartheta_{0} ; z\right) & =\frac{1-\theta_{0} z^{-1}}{1-\phi_{0} z} \text { for } \operatorname{ARMA}(1,1) \text { model } \\
\psi\left(\vartheta_{0} ; z\right) & =\frac{1-\phi_{0} z^{-1}}{1-\phi_{0} z} \text { for } \operatorname{AP}(1,1) \text { model }
\end{aligned}
$$

so that $\delta\left(\vartheta_{0} ; z\right)=\left(z^{-1},-z\right)^{\prime}$ when $\vartheta_{0}=\left(\theta_{0}, \phi_{0}\right)^{\prime}=(0,0)^{\prime}$ for the $\operatorname{ARMA}(1,1)$ model and $\delta\left(\vartheta_{0} ; z\right)=z^{-1}-z$ when $\vartheta_{0}=\phi_{0}=0$ for the $\operatorname{AP}(1,1)$ model.

Then, our feasible test statistics for any $p=1,2, \ldots$ are given by

$$
L M_{T}^{A R M A}=\sum_{t=2}^{T} \hat{\varepsilon}_{t}^{0} \hat{Z}_{t-1}^{\prime}\left(\sum_{t=2}^{T}\left(\hat{\varepsilon}_{t}^{0}\right)^{2} \hat{Z}_{t-1} \hat{Z}_{t-1}^{\prime}\right)^{-1} \sum_{t=2}^{T} \hat{\varepsilon}_{t}^{0} \hat{Z}_{t-1}
$$

and for $\hat{U}_{t-1}=\mathbf{1}_{2}^{\prime} \hat{Z}_{t-1}, \mathbf{1}_{2}=(11)^{\prime} \otimes I_{p}$,

$$
L M_{T}^{A P}=\sum_{t=2}^{T} \hat{\varepsilon}_{t}^{0} \hat{U}_{t-1}^{\prime}\left(\sum_{t=2}^{T}\left(\hat{\varepsilon}_{t}^{0}\right)^{2} \hat{U}_{t-1} \hat{U}_{t-1}^{\prime}\right)^{-1} \sum_{t=2}^{T} \hat{\varepsilon}_{t}^{0} \hat{U}_{t-1}
$$

which are the robust Wald tests of joint significance of the regression of $\hat{\varepsilon}_{t}^{0}=\hat{\varepsilon}_{t}\left(\vartheta_{0}\right)=$ $Y_{t}-\bar{Y}_{T}$ on the $2 p$-vector $\hat{Z}_{t-1}$ and on the $p$-vector $\hat{U}_{t-1}=\mathbf{1}_{2}^{\prime} \hat{Z}_{t-1}$, respectively, defined by

$$
\hat{Z}_{t-1}:=\hat{R}_{t-1}+\hat{S}_{t-1}=i \sum_{j=1}^{t-1} \int \hat{z}_{t-j}^{0}(v)\left\{\hat{\zeta}_{j}(-v)+\hat{\beta}_{j}(-v)\right\} d W(v)
$$

for $\hat{z}_{t}^{0}(v):=e^{i u \hat{\varepsilon}_{t}^{0}}-T^{-1} \sum_{s=1}^{T} e^{i u \hat{\varepsilon}_{s}^{0}}$ and

$$
\hat{\zeta}_{j}(v):=-\sum_{n=j}^{T+j-1} \delta_{n}\left(\vartheta_{0}\right) \hat{\varphi}_{\vartheta_{0}, j-n}^{(1,0)}(0, v) \text { and } \hat{\beta}_{j}(v):=-\delta_{-j}\left(\vartheta_{0}\right) v \hat{\varphi}_{\vartheta_{0}, j}^{(2,0)}(0, v) .
$$

Note that $\hat{\zeta}_{j}$ and $\hat{\beta}_{j}$ converge to zero with $j$ exponentially fast, as $\delta_{ \pm j}\left(\vartheta_{0}\right)$ do for ARMA models, so no additional discount by $k\left(\frac{j}{m}\right)$ is needed in the computation of test statistics unlike for parameter estimation by minimizing $Q_{T}^{m d s}(\vartheta)$.

In particular, for $p=1$ using the form of $\delta(\vartheta ; z)$ for the $\operatorname{ARMA}(1,1)$ model we obtain

$$
\hat{Z}_{t-1}:=\binom{-i \int \hat{z}_{t-1}^{0}(v) v \hat{\varphi}_{\vartheta_{0}, 1}^{(2,0)}(0,-v) d W(v)}{i \int \hat{z}_{t-1}^{0}(v) \hat{\varphi}_{\vartheta_{0}}^{(1)}(-v) d W(v)}=\binom{\hat{Z}_{t-1,1}}{\hat{Z}_{t-1,2}}, \text { say },
$$

where the series $\hat{Z}_{t-1}, t=2, \ldots, T$, has closed-form expressions for $c d f \mathrm{~s} W$ with explicit characteristic function, so that for $W$ Gaussian, we obtain

$$
\hat{Z}_{t-1,1}=-\frac{1}{T-1} \sum_{r=2}^{T}\left(\hat{\varepsilon}_{r}^{0}\right)^{2}\left[\begin{array}{c}
\left(\hat{\varepsilon}_{t-1}^{0}-\hat{\varepsilon}_{r-1}^{0}\right) \exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{t-1}^{0}-\hat{\varepsilon}_{r-1}^{0}\right)^{2}\right\} \\
-\frac{1}{T-1} \sum_{s=2}^{T}\left(\hat{\varepsilon}_{s-1}^{0}-\hat{\varepsilon}_{r-1}^{0}\right) \exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{s-1}^{0}-\hat{\varepsilon}_{r-1}^{0}\right)^{2}\right\}
\end{array}\right]
$$

and

$$
\hat{Z}_{t-1,2}=\frac{1}{T} \sum_{r=1}^{T} \hat{\varepsilon}_{r}^{0}\left[\exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{t-1}^{0}-\hat{\varepsilon}_{r}^{0}\right)^{2}\right\}-\frac{1}{T-1} \sum_{s=2}^{T} \exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{s-1}^{0}-\hat{\varepsilon}_{r}^{0}\right)^{2}\right\}\right] .
$$

For our asymptotic theory we need to impose a moment and a mixing assumption on the $m d s \varepsilon_{t}$ to control the higher order dependence, together with non-Gaussianity to achieve consistency against uncorrelated AP alternatives. Non-Gaussianity, characterized by the existence of nonzero higher-order cumulants like in Section 2, is also key in Velasco (2022) for unrestricted estimation of ARMA models by identifying the correct location of AR and MA roots with respect the complex unit circle using the (pairwise) predictability of residuals under wrong location. However, non-Gaussianity is not necessary for root-restricted estimation, e.g. imposing non-invertibility and causality, of ARMA or all-pass models. Assumptions are compiled and further discussed in Appendix B. Extension to other parametric models is
straightforward under standard smoothness conditions on the model (implying weak dependence in the observed data and model residuals) and parameter identification. Next theorem shows that LM tests have standard null asymptotic distribution.

Theorem 3.3.1. Under Assumptions 3.1 and 3.2 as $T \rightarrow \infty$,

$$
\begin{aligned}
L M_{T}^{A R M A} & \rightarrow_{d} \chi_{2 p}^{2} \text { under } H_{m d s} \\
L M_{T}^{A P} & \rightarrow_{d} \quad \chi_{p}^{2} \text { under } H_{m d s}^{(A P)}
\end{aligned}
$$

and $L M_{T}^{A R M A} \rightarrow_{p} \infty$ under $H_{m d s}^{1}$ and $L M_{T}^{A P} \rightarrow_{p} \infty$ under $H_{m d s}^{(A P) 1}$.

### 3.3.2. Tests of the AP model

For the LM testing of the AP model against $\operatorname{ARMA}(1,1)$, i.e. $H_{A P / m d s}: \phi_{0}=\theta_{0}$ in model (3.3.1), we parameterize $\gamma_{0}:=\phi_{0}-\theta_{0}$ and test $H_{A P}^{*}: \gamma_{0}=0$ in

$$
\left(1-\phi_{0} L\right) Y_{t}=\left(1-\phi_{0} L^{-1}+\gamma_{0} L^{-1}\right) \varepsilon_{t-1}
$$

so that keeping the same notation for the model transfer function with the new parameters $\vartheta^{*}:=(\gamma, \phi)^{\prime}$,

$$
\psi^{*}\left(\vartheta_{0}^{*} ; z\right)=\frac{\left(1-\phi_{0} z^{-1}+\gamma_{0} z^{-1}\right)}{\left(1-\phi_{0} z\right)} \text { for ARMA }(1,1) \text { model }
$$

and similarly,

$$
\delta^{*}\left(\vartheta^{*} ; z\right)=-\frac{\partial}{\partial \vartheta^{*}} \ln \psi^{*}\left(\vartheta^{*} ; z\right)
$$

so that

$$
\delta^{*}\left(\vartheta_{0}^{*} ; z\right)=\left(-\sum_{j=0}^{\infty} \phi_{0}^{j} z^{-j-1}, \quad \sum_{j=0}^{\infty} \phi_{0}^{j} z^{-j-1}-\sum_{j=0}^{\infty} \phi_{0}^{j} z^{j+1}\right)^{\prime}
$$

with

$$
\delta^{*}\left(\vartheta_{0}^{*} ; z\right)=A \delta\left(\vartheta_{0} ; z\right), \quad A:=\binom{A_{1}}{A_{2}}=\left(\begin{array}{cc}
-1 & 0 \\
1 & 1
\end{array}\right) \otimes I_{p}
$$

for $\vartheta^{*}=A^{\prime} \vartheta$ and $\vartheta=A^{\prime} \vartheta^{*}$.
The LM statistic based on the score $\left(\partial / \partial \vartheta^{*}\right) Q_{T}^{m d s}\left(\tilde{\vartheta}_{T}^{*}\right)$ evaluated at the all-pass restricted estimate $\tilde{\vartheta}_{T}^{*}=\left(0^{\prime}, \tilde{\phi}_{T}^{\prime}\right)^{\prime}$ is approximated by

$$
\widetilde{L M}_{T}^{A P}=T\left(\frac{1}{T} \sum_{t=2}^{T} \tilde{\varepsilon}_{t} \tilde{V}_{t-1}\right)^{\prime} \tilde{V}_{T}^{-1}\left(\frac{1}{T} \sum_{t=2}^{T} \tilde{\varepsilon}_{t} \tilde{V}_{t-1}\right)
$$

where $\tilde{\varepsilon}_{t}=\hat{\varepsilon}_{t}\left(\tilde{\vartheta}_{T}^{*}\right)$ are the restricted residuals and $\tilde{V}_{t-1}=A_{1} \tilde{Z}_{t-1}$, respectively,
with

$$
\begin{equation*}
\tilde{Z}_{t-1}:=\tilde{R}_{t-1}+\tilde{S}_{t-1}=i \sum_{j=1}^{T-1} \int \tilde{z}_{t-j}(v)\left\{\tilde{\zeta}_{j}(-v)+\tilde{\beta}_{j}(-v)\right\} d W(v) \tag{3.3.3}
\end{equation*}
$$

for $\tilde{z}_{t}(v):=z_{t}\left(v, \tilde{\vartheta}_{T}^{*}\right)=e^{i u \tilde{\varepsilon}_{t}}-T^{-1} \sum_{s=1}^{T} e^{i u \tilde{\varepsilon}_{s}}$ and

$$
\tilde{\zeta}_{j}(v):=-\sum_{n=j}^{T+j-1} \delta_{n}\left(\tilde{\vartheta}_{T}^{*}\right) \hat{\varphi}_{\hat{\vartheta}_{T}^{*}, j-n}^{(1,0)}(0, v) \text { and } \tilde{\beta}_{j}(v):=-\delta_{-j}\left(\tilde{\vartheta}_{T}^{*}\right) v \hat{\varphi}_{\hat{\vartheta}_{T}^{*}, j}^{(2,0)}(0, v),
$$

with the empirical characteristic function evaluated at restricted residuals. The integrals have again closed-form expressions not requiring numerical integration.

The asymptotic variance estimate

$$
\tilde{V}_{T}:=\tilde{\Xi}_{1} A \tilde{V}_{0}^{m d s} A^{\prime} \tilde{\Xi}_{1}^{\prime}
$$

accounts for the estimation of $\phi_{0}$ through

$$
\tilde{\Xi}_{1}:=\left[\begin{array}{ll}
I_{p} & \left.\vdots-A_{1} \tilde{H}_{0}^{m d s} A_{2}^{\prime}\left\{A_{2} \tilde{H}_{0}^{m d s} A_{2}^{\prime}\right\}^{-1}\right]
\end{array}\right.
$$

with

$$
\begin{equation*}
\tilde{H}_{0}^{m d s}=-\sum_{j=1}^{T-1} \int\left\{\tilde{\zeta}_{j}(-v)+\tilde{\beta}_{j}(-v)\right\}\left\{\tilde{\zeta}_{j}(v)+\tilde{\beta}_{j}(v)\right\}^{\prime} d W(v), \tag{3.3.4}
\end{equation*}
$$

and for conditional heteroskedasticity of general form by means of

$$
\begin{equation*}
\tilde{V}_{0}^{m d s}:=\frac{1}{T-1} \sum_{t=2}^{T} \tilde{\varepsilon}_{t}^{2} \tilde{Z}_{t-1} \tilde{Z}_{t-1}^{\prime}, \tag{3.3.5}
\end{equation*}
$$

see Appendix B for explicit formulae.
The asymptotic distribution of the LM statistics is provided in next theorem, which shares similar properties as the previous tests against simple alternatives. We now require regularity conditions on the weighting function $W$, the lag kernel $k$ and bandwidth $m$ because estimation of nuisance parameters is needed under $H_{A P / m d s}$. These are satisfied for many standard choices. Note that $m$ does not play a role in first order asymptotic properties of tests as far as it grows with $T$ slower than $T^{1 / 2}$ for a smooth $k$ at the origin.

Theorem 3.3.2. Under Assumptions 3.1, 3.2 and 3.3 as $T \rightarrow \infty$,

$$
\widetilde{L M}_{T}^{A P} \rightarrow_{d} \chi_{p}^{2} \text { under } H_{A P / m d s}
$$

and $\widetilde{L M}_{T}^{A P} \rightarrow_{p} \infty$ under $H_{A P / m d s}^{1}$.

### 3.4. Directional Nonlinear Dependence Tests for Residuals

In this section, we adapt the previous non-linear dependence tests based on an allpass filter to residuals of a linear $\operatorname{ARMA}\left(q_{1}, q_{2}\right)$ model which has already accounted for the linear dependence in observed data by an appropriate choice of the model orders. To this end, we design LM tests that allow, as in the previous section, for higher order dependence in the $m d s$ true errors and account for the estimation effect of the preliminary $\operatorname{ARMA}\left(q_{1}, q_{2}\right)$ model. To impose the additional all-pass filter in the model dynamics, we use a multiplicative $\operatorname{ARMA}\left(q_{1}, q_{2}\right)-\mathrm{AP}(p, p)$ structure which guarantees that the extra roots in the AR and MA lag polynomials are reciprocals, i.e.

$$
\begin{equation*}
\psi\left(\vartheta_{0} ; z\right)=\frac{\beta(z)}{\alpha(z)} \frac{1-\phi_{0} z^{-1}}{1-\phi_{0} z} \tag{3.4.1}
\end{equation*}
$$

for $p=1$, where the, non necessarily causal and invertible, $\operatorname{ARMA}\left(q_{1}, q_{2}\right)$ filter, $q_{1}+q_{2}>0$, has lag polynomials

$$
\begin{aligned}
\alpha(z) & =1-\alpha_{1} z-\cdots-\alpha_{q_{1}} z^{q_{1}} \\
\beta(z) & =1-\beta_{1} z-\cdots-\beta_{q_{2}} z^{q_{2}}
\end{aligned}
$$

with no common roots, all outside the unit circle, with $\vartheta_{0}=\left(\phi_{0}, \alpha_{1}, \ldots, \alpha_{q_{1}}, \beta_{1}, \ldots, \beta_{q_{2}}\right)^{\prime}$.
Then we focus on testing the mds hypothesis on the $\operatorname{ARMA}\left(q_{1}, q_{2}\right)$ residuals

$$
\varepsilon_{t}^{*}=\frac{\alpha(L)}{\beta(L)} Y_{t},
$$

against all-pass restricted $\operatorname{ARMA}(1,1)$ dependence, i.e.

$$
\begin{aligned}
& H_{m d s}^{(A R M A-A P)}: \phi_{0}=0 \text { in model (3.4.1), } \\
& H_{m d s}^{(A R M A-A P) 1}: \quad \phi_{0} \neq 0,
\end{aligned}
$$

but discuss general $p \geq 1$. This alternative can not be detected consistently by usual residual serial correlation tests, while tests checking for ARCH effects in residuals could detect it indirectly due to the higher order dependence generated by the all pass filter.

The LM statistic to test $H_{m d s}^{(A R M A-A P)}$ is based on the score of the filter $\psi\left(\vartheta_{0} ; z\right)$ with respect to $\phi, \delta^{A P}\left(\vartheta_{0} ; z\right):=-(\partial / \partial \phi) \log \psi\left(\vartheta_{0} ; z\right)=z^{-1}-z$ for $\vartheta_{0}=\left(\phi_{0}, \alpha_{0}^{\prime}, \beta_{0}^{\prime}\right)^{\prime}=$ $\left(\phi_{0}, \vartheta_{0}^{*}\right)^{\prime}, \vartheta_{0}^{*}=\left(\alpha_{0}^{\prime}, \beta_{0}^{\prime}\right)^{\prime}$, which for $p=1$ is $\delta^{A P}\left(\vartheta_{0} ; z\right)=z^{-1}-z$ as when testing the $\operatorname{AP}(1,1)$ model but now $\varepsilon_{t}^{*}$ has to be replaced by the residuals $\tilde{\varepsilon}_{t}$ after estimation of
$\left(\alpha_{0}, \beta_{0}\right)$ whose effect depends on

$$
\begin{aligned}
\delta^{\mathrm{ARMA}}\left(\vartheta_{0} ; z\right) & =-\frac{\partial}{\partial \vartheta^{*}} \log \psi\left(\vartheta_{0} ; z\right) \\
& =-\frac{\partial}{\partial \vartheta^{*}} \log \left(1-\beta_{1} z-\cdots-\beta_{q_{2}} z^{q_{2}}\right)+\frac{\partial}{\partial \vartheta^{*}} \log \left(1-\alpha_{1} z-\cdots-\alpha_{q_{1}} q^{q_{1}}\right) \\
& =\left(\frac{z}{\beta(z)}, \cdots, \frac{z^{q_{2}}}{\beta(z)},-\frac{z}{\alpha(z)}, \cdots,-\frac{z^{q_{1}}}{\alpha(z)}\right)^{\prime} .
\end{aligned}
$$

Then the LM statistic based on the score $(\partial / \partial \phi) Q_{T}^{m d s}\left(\tilde{\vartheta}_{T}^{*}\right)$ evaluated at the $\operatorname{ARMA}\left(q_{1}, q_{2}\right)$ estimates $\tilde{\vartheta}_{T}^{*}=\left(\tilde{\alpha}_{T}^{\prime}, \tilde{\beta}_{T}^{\prime}\right)^{\prime}$ is approximated by

$$
\widetilde{L M}_{T}^{A R M A-A P}:=T\left(\frac{1}{T} \sum_{t=2}^{T} \tilde{\varepsilon}_{t} \tilde{Z}_{t-1}^{(1)}\right) \tilde{V}_{T}^{-1}\left(\frac{1}{T} \sum_{t=2}^{T} \tilde{\varepsilon}_{t} \tilde{Z}_{t-1}^{(1)}\right)
$$

where $\tilde{\varepsilon}_{t}=\hat{\varepsilon}_{t}\left(\tilde{\vartheta}_{T}^{*}\right)$ are the ARMA residuals and $\tilde{Z}_{t-1}^{(1)}$ is defined using the form of $\delta^{A P}(z)$, so that for $p=1$ equals

$$
\tilde{Z}_{t-1}^{(1)}:=i \int \tilde{z}_{t-1}(v)\left\{\tilde{\varphi}_{\vartheta_{0}}^{(1)}(-v)-\tilde{\varphi}_{\vartheta_{0}, 1}^{(2,0)}(0,-v)\right\} d W(v)
$$

for $\tilde{z}_{t}(v):=z_{t}\left(v, \tilde{\vartheta}_{T}^{*}\right)=e^{i u \tilde{\varepsilon}_{t}}-T^{-1} \sum_{s=1}^{T} e^{i u \tilde{\varepsilon}_{s}}$ and has the same closed-form expression as before.

The asymptotic variance estimate,

$$
\tilde{V}_{T}:=\tilde{\Xi}_{1} \tilde{V}_{0}^{m d s} \tilde{\Xi}_{1}^{\prime},
$$

accounts for the pre-estimation of $\vartheta_{0}^{*}=\left(\alpha_{0}^{\prime}, \beta_{0}^{\prime}\right)^{\prime}$ corresponding to the restricted model under $H_{m d s}^{(A R M A-A P)}$ through the $p \times\left(p+q_{1}+q_{2}\right)$ matrix

$$
\tilde{\Xi}_{1}:=\left[\begin{array}{ccc}
I_{p} & \vdots & -\tilde{H}_{0,12}^{m d s}\left\{\tilde{H}_{0,22}^{m d s}\right\}^{-1}
\end{array}\right]
$$

with $\tilde{H}_{0}^{m d s}=\left[\tilde{H}_{0, a b}^{m d s}\right]_{a b=1,2}$ as in (3.3.4) and $\zeta_{j}(v)$ and $\beta_{j}(v)$ defined in terms of the coefficients of the joint expansion of $\delta(z):=-(\partial / \partial \vartheta) \log \psi\left(\vartheta_{0} ; z\right)=\left(\delta^{A P}\left(\vartheta_{0} ; z\right), \delta^{A R M A}\left(\vartheta_{0} ; z\right)^{\prime}\right)^{\prime}$.
These are also used for $\tilde{Z}_{t-1}=\left(\tilde{Z}_{t-1}^{(1)}, \tilde{Z}_{t-1}^{(2) \prime}\right)^{\prime}$ in (3.3.3) for the estimate $\tilde{V}_{0}^{m d s}$ of the asymptotic variance of the joint score $(\partial / \partial \vartheta) Q_{T}^{m d s}\left(\vartheta_{0}\right)$ in (3.3.5). Then, it is immediate to argue that the residuals LM test shares the same asymptotic properties as when applied to observed data, with consistency relying on non-Gaussianity.

Theorem 3.4.1. Under Assumptions 3.1, 3.2 and 3.3 as $T \rightarrow \infty$,

$$
\widetilde{L M}_{T}^{A R M A-A P} \rightarrow_{d} \chi_{p}^{2} \text { under } H_{m d s}^{(A R M A-A P)}
$$

and $\widetilde{L M}_{T}^{A R M A-A P} \rightarrow_{p} \infty$ under $H_{m d s}^{(A R M A-A P) 1}$.

### 3.5. Monte Carlo Simulations

In this section, we report the results of a simulation exercise to investigate the finite sample properties of our new directional tests. We follow the experiment in Lanne et al. $(2013)$ and simulate $\operatorname{ARMA}(1,1)$ models with innovations distributed as (standardized) $\chi_{4}^{2}$ and $t_{4}$ random variates, to consider both asymmetric and symmetric distributions, with lighter and heavier tails, respectively. We impose either independence or just $m d s$ with conditional heteroskedasticity generated by a $\operatorname{GARCH}(1,1)$ model with parameters $(1,0.8,0.1)$. We simulate 10,000 independent replications with three sample sizes, $T=100,200$ and 1,000 replications for $T=500$, choose $m=\left\lfloor T^{1 / 5}\right\rfloor+1$. The model parameters are the same as in Lanne et al. (2013) simulations, where MLE and test statistics used that the data was iid and $t$ distributed.

For iid innovations we also consider tests based on an independence loss function $Q_{T}^{i i d}$ using the joint characteristic function itself instead of its derivative for measuring first order dependence with $Q_{T}^{m d s}$. The test statistics based on the iid criteria are described in Appendix B and share the same asymptotic properties as those of Section 3 but exploit the independence assumption to construct a simpler standardization.

We summarize the results for $\chi^{2}$ and $t$ distributions in Tables 1 and 2, respectively, for the the three combinations of $Q_{T}^{i i d}$ and $Q_{T}^{m d s}$ and of models with iid and $m d s$ innovations which are well specified under the null. In terms of size, the LM tests of simple hypothesis $H_{m d s}^{(A P)}$ and $H_{m d s}$ have empirical size closer to the nominal level compared to those testing the composite hypothesis $H_{A P}$, which need to estimate the restricted model under the null. While tests under iid have reasonable performance for both innovation distributions, tests under $m d s$ display substantial over-rejection for both parameters for smaller sample sizes, but improving with sample size for $\chi_{4}^{2}$ innovations. However, tests for the AP-mds hypothesis for series with $t_{4}$ innovations are severely over-sized, even for the largest sample size.

However, regarding power, LM tests against simple hypothesis for series with $m d s t_{4}$ innovations have very little power when compared to the $m d s \chi_{4}^{2}$ innovations for all models and sample sizes, and also compared with the reported results for the composite hypothesis $H_{A P}$. By contrast, for the $\chi^{2}$ innovations, LM tests provide comparable rejection rates for both types of null hypothesis. Perhaps due to the local nature of LM tests, the power functions of simple hypotheses tests seem not to be monotone for large departures from the null, specially for the smaller sample sizes and iid innovations.

LM tests imposing the (true) iid condition through the checking function $Q_{T}^{i i d}$ show more power than those just exploiting the $m d s$ condition through $Q_{T}^{m d s}$, as could be expected from using a weaker restriction and from the need to account for higher order dependence in the standardization of test statistics. Similarly, when imposing the $m d s$ assumption in innovations, conditional heteroskedasticity makes more difficult to reject all tested versions of the null hypothesis of conditional independence in mean.

Compared to Lanne et al. (2013) procedures exploiting knowledge of the true distribution of iid innovations, the results based on our LM tests of the mds hypothesis report much lower power. However, when compared with tests based on $Q_{T}^{i i d}$, results improve substantially, getting closer to theirs for $\chi^{2}$ innovations, whose strong skewness seems to provide more information to our semiparametric tests than the symmetric $t$ innovations. This is related to the representation of tests statistics based on generalized autocovariances in terms of correlations of higher-order power transformations of observations, see Escanciano and Velasco (2006a) and Velasco (2022).

Table 3.5.1: Rejection rates of $5 \%$ LM tests: $\operatorname{ARMA}(1,1)$ models, $\varepsilon_{t} \sim \chi_{4}^{2}$.

| $\phi_{0}$ | $\theta_{0}$ | $Q_{T}^{i i d}, \varepsilon_{t} \sim$ iid |  |  | $Q_{T}^{m d s}, \varepsilon_{t} \sim$ iid |  |  | $Q_{T}^{m d s}, \varepsilon_{t} \sim m d s$ |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $H_{A P}: \phi_{0}=\theta_{0} \text { in ARMA }(1,1)$ |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 0.00 | 0.00 | 4.65 | 5.32 | 6.5 | 11.83 | 5.03 | 5.3 | 8.60 | 4.21 | 3.4 |
| 0.80 | 0.80 | 4.54 | 5.16 | 5.4 | 7.22 | 6.71 | 6.8 | 9.10 | 7.50 | 6.3 |
| 0.80 | 0.85 | 4.02 | 7.43 | 28.4 | 14.01 | 20.40 | 43.2 | 10.39 | 12.72 | 48.0 |
| 0.80 | 0.90 | 5.84 | 19.01 | 72.3 | 23.61 | 41.90 | 85.5 | 15.35 | 23.74 | 19.2 |
| 0.80 | 0.95 | 7.88 | 28.33 | 84.8 | 31.42 | 54.38 | 90.2 | 19.75 | 33.79 | 66.8 |
| 0.80 | 0.75 | 11.93 | 20.06 | 40.4 | 6.61 | 10.06 | 26.5 | 8.04 | 10.63 | 19.0 |
| 0.80 | 0.70 | 27.75 | 51.83 | 87.7 | 14.59 | 34.98 | 77.5 | 12.82 | 23.61 | 49.9 |
| 0.80 | 0.65 | 47.20 | 78.80 | 97.4 | 30.39 | 65.94 | 97.2 | 21.25 | 41.07 | 77.6 |
| $H_{m d s}^{(A P)}: \phi_{0}=0$ in all-pass $(1,1)$ |  |  |  |  |  |  |  |  |  |  |
| 0.00 | 0.00 | 4.92 | 4.93 | 5.4 | 4.83 | 5.06 | 4.5 | 4.53 | 6.24 | 6.2 |
| 0.10 | 0.10 | 24.14 | 41.54 | 79.6 | 14.07 | 27.84 | 62.1 | 4.85 | 4.96 | 7.3 |
| 0.20 | 0.20 | 59.81 | 87.95 | 100.0 | 39.66 | 73.21 | 98.8 | 7.60 | 11.02 | 21.2 |
| 0.40 | 0.40 | 85.07 | 99.19 | 100.0 | 71.20 | 95.52 | 100.0 | 16.62 | 27.27 | 50.6 |
| 0.60 | 0.60 | 58.75 | 79.00 | 89.8 | 54.10 | 76.33 | 93.8 | 26.32 | 44.30 | 71.7 |
| $H_{m d s}: \phi_{0}=\theta_{0}=0$ in $\operatorname{ARMA}(1,1)$ |  |  |  |  |  |  |  |  |  |  |
| 0.00 | 0.00 | 4.49 | 5.03 | 5.5 | 5.06 | 5.55 | 6.0 | 3.99 | 6.14 | 5.5 |
| 0.10 | 0.10 | 17.96 | 32.74 | 71.0 | 10.77 | 20.32 | 54.3 | 4.85 | 4.91 | 6.9 |
| 0.20 | 0.20 | 48.89 | 80.84 | 99.6 | 30.66 | 64.28 | 98.1 | 6.77 | 9.43 | 18.7 |
| 0.40 | 0.40 | 76.61 | 98.15 | 100.0 | 67.78 | 96.52 | 100.0 | 15.87 | 24.87 | 48.5 |
| 0.60 | 0.60 | 50.93 | 74.96 | 89.3 | 59.19 | 83.78 | 95.4 | 25.39 | 43.36 | 70.1 |

Note: The model innovations are (standardized) $\chi_{4}^{2}$ and $t_{4}$ random variates, either iid or transformed to $m d s$ by multiplication of the conditional standard deviation generated by a $\operatorname{GARCH}(1,1)$ model with parameters $(1,0.8,0.1) \cdot m=\left\lfloor T^{1 / 5}\right\rfloor+1$ and 10,000 replications are used for $T=100,200$ or 1,000 for $T=500$. Null hypotheses parameters in bold. See text for test statistics computation.

Table 3.5.2: Rejection rates of $5 \%$ LM tests: $\operatorname{ARMA}(1,1)$ models, $\varepsilon_{t} \sim t_{4}$.

| $\phi_{0}$ | $\theta_{0}$ |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 0.00 | 0.00 | 3.22 | 3.36 | 4.0 | 11.83 | 11.95 | 13.9 | 8.60 | 7.04 | 6.9 |
| 0.80 | 0.80 | 3.75 | 3.28 | 3.6 | 7.22 | 5.19 | 8.5 | 9.10 | 11.03 | 14.8 |
| 0.80 | 0.85 | 2.44 | 2.43 | 6.3 | 13.24 | 18.19 | 28.3 | 10.66 | 12.04 | 18.1 |
| 0.80 | 0.90 | 3.05 | 6.15 | 24.9 | 21.08 | 30.61 | 47.4 | 14.29 | 18.68 | 28.1 |
| 0.80 | 0.95 | 3.68 | 10.11 | 43.9 | 26.48 | 39.25 | 55.8 | 17.14 | 25.07 | 38.5 |
| 0.80 | 0.75 | 10.12 | 17.12 | 41.7 | 6.00 | 8.76 | 16.5 | 11.58 | 17.00 | 31.6 |
| 0.80 | 0.70 | 24.23 | 47.23 | 86.7 | 12.90 | 25.56 | 59.4 | 18.69 | 32.69 | 60.5 |
| 0.80 | 0.65 | 42.25 | 74.11 | 96.6 | 26.87 | 53.13 | 89.9 | 29.15 | 51.25 | 82.9 |
| $H_{m d s}^{(A P)}: \phi_{0}=0$ in all-pass $(1,1)$ |  |  |  |  |  |  |  |  |  |  |
| 0.00 | 0.00 | 5.28 | 5.05 | 3.9 | 4.83 | 5.19 | 5.2 | 4.53 | 4.50 | 5.2 |
| 0.10 | 0.10 | 8.99 | 13.45 | 24.8 | 6.12 | 7.75 | 12.8 | 4.70 | 4.89 | 4.7 |
| 0.20 | 0.20 | 16.64 | 27.74 | 55.7 | 7.34 | 9.91 | 20.3 | 5.09 | 5.26 | 4.8 |
| 0.40 | 0.40 | 16.74 | 27.03 | 59.8 | 7.08 | 7.86 | 9.5 | 5.66 | 5.79 | 6.8 |
| 0.60 | 0.60 | 7.13 | 8.02 | 14.2 | 7.24 | 7.35 | 7.3 | 5.53 | 5.60 | 5.7 |
| $H_{m d s}: \phi_{0}=\theta_{0}=0$ in ARMA(1,1) |  |  |  |  |  |  |  |  |  |  |
| 0.00 | 0.00 | 4.50 | 5.09 | 4.3 | 5.06 | 4.69 | 5.2 | 3.99 | 3.83 | 4.4 |
| 0.10 | 0.10 | 7.16 | 10.50 | 18.4 | 5.70 | 6.73 | 10.5 | 4.27 | 4.26 | 4.9 |
| 0.20 | 0.20 | 12.63 | 21.07 | 47.4 | 6.48 | 8.98 | 19.1 | 4.54 | 4.88 | 6.2 |
| 0.40 | 0.40 | 12.20 | 20.83 | 49.5 | 5.62 | 6.98 | 10.9 | 5.13 | 5.51 | 6.3 |
| 0.60 | 0.60 | 6.74 | 8.46 | 17.3 | 5.36 | 5.87 | 6.1 | 5.37 | 5.64 | 5.1 |

Note: See note for Table 1.

### 3.6. Empirical Application

In this section, we replicate the empirical analysis of Lanne et al. (2013) on four series of quarterly US returns for market and value-weighted size-ordered portfolios using all NYSE, AMEX, and NASDAQ stocks obtained from Kenneth French web page in June 2022. We use the same time span (1947:1 to 2007:12) to collect monthly data, which is transformed into quarterly returns $(T=732)$. This frequency choice of Lanne et al. (2013) is motivated in terms of suitability of the $\operatorname{ARMA}(1,1)$ specification because otherwise higher frequency return series display strong conditional heteroskedasticity and very weak autocorrelation. Further, higher kurtosis motivates the choice of the $t$ distribution for the innovations, but we remain agnostic about the distribution of shocks by using our methods based on general dependence measures.

Our iid causal and non-invertible ARMA $(1,1)$ estimation results based on minimizing $Q_{T}^{i i d}$ in Table 3 are only similar to those of Lanne et al. (2013) for the Bottom $30 \%$ series, but differ for the other three series, with the ARMA specification with iid errors being seriously questioned for all return series by Hong (1999) independence test applied to model residuals, though from our residual specification test, no
further lags would improve the fitting. In parallel, the AP model ( $H_{A P / i i d}$ ) cannot be rejected in favor of the general ARMA specification using LM and Wald tests, nor its residuals seem to display serial dependence for all series, in agreement with Lanne et al. (2013) results. These potential problems of specification imposing iid errors can also explain the discrepancy between the restricted and unrestricted parameter estimates for some series.

However, both LM and Wald tests confirm that the AP model is able to detect the serial dependence of observations (though the LM test cannot reject $H_{\text {iid }}^{(A P)}$ for Bottom 30\% returns). Again, this conclusion matches the qualitative results of Lanne et al. (2013) based on LR and Wald tests for the $t$ distribution. The direct single-step LM tests of $H_{i i d}$ against $H_{i i d}^{1}$, despite arriving to the same conclusion, have less power than Wald tests, as expected, and also than the omnibus Hong's test in terms of p -values.

The results of Table 4 when only $m d s$ is imposed on model errors, show that both ARMA and AP models provide a good fit according to this identification hypothesis, with restricted AP estimates very similar to unrestricted ARMA $(1,1)$ ones for all return series. The AP model $\left(H_{A P / m d s}\right)$ is never rejected in favour of the general ARMA one as under iid while Wald tests reject the $m d s$ hypothesis $H_{m d s}^{(A P)}$ in favour of the AP model for all four returns with the LM tests rejecting also for all returns at the $10 \%$ level, except for Bottom $30 \%$ returns. Finally, the $m d s$ hypothesis $H_{m d s}$ is not rejected at any usual significance level in favour of the general ARMA model using Hong tests (and only for two series using the LM tests), showing that this testing strategy can be less effective than the one testing $m d s$ against the AP model under the alternative when autocorrelation is weak but there is nonlinear dependence.

### 3.7. Conclusions

In this paper, we have shown how to construct tests of predictability of observed series and model residuals which allow for higher-order dependence on observations or model errors. We use a parametric $\operatorname{ARMA}(1,1)$ specification to describe both linear and non-linear dependence using its all-pass variant and propose Lagrange Multiplier tests which avoid estimation of the complete model. Our finite sample analysis shows that the performance of the new tests is reasonable for moderate sample sizes, but power depends crucially on the true distribution of innovations, so it seems worth exploring alternative characterizations of the past information to those implied by the characteristic function as well as optimal weighting an scaling. Further, in the same way as our empirical application found that Wald tests could be more powerful than LM tests, also pseudo likelihood ratio tests based on $Q_{T}^{m d s}$

Table 3.6.1: iid estimation and tests for quarterly returns.

|  | Portfolio |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Market | Bottom 30\% | Middle 40\% | Top 30\% |
|  | ARMA(1, 1) |  |  |  |
| $\phi_{0}$ | -0.448 | 0.657 | -0.443 | -0.462 |
|  | (.102) | (.076) | (.103) | (.116) |
| $\theta_{0}$ | -0.547 | 0.673 | -0.552 | -0.559 |
|  | (.101) | (.076) | (.103) | (.111) |
| $Q_{T}^{\text {iid }}$ ARMA-AP | 2.706 | 3.915 | 4.022 | 2.444 |
| $\widetilde{L M}_{T}^{A R M A-A P}$ | 0.990 | 0.572 | 0.500 | 0.943 |
| Res. Hong ${ }^{\text {iid }}$ | 0.003 | 0.114 | 0.000 | 0.003 |
|  | AP(1, 1) |  |  |  |
| $\phi_{0}$ | -0.472 | 0.663 | 0.665 | -0.493 |
|  | (.101) | (.071) | (.063) | (.109) |
| $Q_{T}^{i i d}{ }_{\text {ARMA-AP }}$ | 3.015 | 3.922 | 4.048 | 2.764 |
| $\widetilde{L M}_{T}^{A F}$ <br> Res. Hong ${ }^{\text {iid }}$ | 0.880 | 0.015 | 0.024 | 0.967 |
|  | 0.492 | 0.250 | 0.040 | 0.621 |
|  |  | ${ }_{\text {iid }}: \phi_{0}=\theta_{0}$ | vs $H_{A P / i i d}^{1}$ : | $\neq \theta_{0}$ |
| $\begin{aligned} & \widetilde{L M}_{T}^{A P} \\ & \text { Wald } \end{aligned}$ | 0.438 | 0.522 | 0.529 | 0.428 |
|  | 0.104 | 0.754 | 0.073 | 0.117 |
|  | $H_{i i d}^{(A P)}: \phi_{0}=0$ |  | $H_{i i d}^{(A P) 1}: \phi_{0} \neq 0$ |  |
| $L M_{T}^{A P}$ | 0.002 | 0.881 | 0.029 | 0.002 |
| Wald | 0.000 | 0.000 | 0.000 | 0.000 |
|  | $H_{\text {iid }}:\left(\phi_{0}, \theta_{0}\right)=0$ |  | $H_{i i d}^{1}:\left(\phi_{0}, \theta_{0}\right) \neq 0$ |  |
| $L M_{T}^{A R M A}$ | 0.006 | 0.966 | 0.092 | 0.006 |
| Wald | 0.000 | 0.000 | 0.000 | 0.000 |
| Hongid | 0.001 | 0.102 | 0.000 | 0.002 |

Note: See text for data description. Estimates and tests based on $Q_{T}^{i i d}$ described in Appendix B with $m=\left\lfloor T^{1 / 3}\right\rfloor+1=6$. Standard error of estimates in parenthesis computed under the iid assumption. For tests p-values are reported, in bold when lower than 0.05 .
and $Q_{T}^{i i d}$ could be considered.

Table 3.6.2: mds estimation and tests for quarterly returns.

|  | Portfolio |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Market | Bottom 30\% | Middle 40\% |  | Top 30\% |
|  |  | ARMA(1, 1) |  |  |  |
| $\phi_{0}$ | 0623 | 0.604 |  | 0.532 | 0.682 |
|  | (.120) | (.151) |  | (.132) | (.120) |
| $\theta_{0}$ | 0.646 | 0.675 |  | 0.630 | 0.663 |
|  | (.113) | (.143) |  | (.123) | (.111) |
| $Q_{T}^{m d s}$ | 4.606 | 5.462 |  | 5.830 | 4.455 |
| $\begin{aligned} & \widetilde{L M}_{T}^{A R M A-A P} \\ & \text { Res. Hong }{ }^{m d s} \end{aligned}$ | 0.775 | 0.000 |  | 0.712 | 0.668 |
|  | 0.826 | 0.514 |  | 0.474 | 0.374 |
|  | AP(1, 1) |  |  |  |  |
| $\phi_{0}$ | 0.657 | 0.622 |  | 0.601 | 0.660 |
|  | (.103) | (.144) |  | (.124) | (.097) |
|  | 4.693 | 6.039 |  | 6.921 | 4.517 |
| $\stackrel{i}{L M}_{T}^{A R}$ <br> Res. Hong ${ }^{m d s}$ | 0.918 | 0.709 |  | 0.805 | 0.742 |
|  | 0.598 | 0.870 |  | 0.548 | 0.552 |
|  | $H_{A P / m d s}: \phi_{0}=\theta_{0}$ |  | vs | $H_{A P / m d s}^{1}: \phi_{0} \neq \theta_{0}$ |  |
| $\begin{aligned} & \widetilde{L M}_{T}^{A P} \\ & \text { Wald } \end{aligned}$ | 0.923 | 0.383 |  | 0.323 | 0.546 |
|  | 0.650 | 0.223 |  | 0.123 | 0.710 |
|  | $H_{m d s}^{(A P)}: \phi_{0}=0$ |  | vs $H_{m d s}^{(A P) 1}: \phi_{0} \neq 0$ |  |  |
| $L M_{T}^{A P}$ | 0.023 | 0.363 |  | 0.055 | 0.028 |
| Wald | 0.000 | 0.000 |  | 0.000 | 0.000 |
|  | $H_{m d s}:\left(\phi_{0}, \theta_{0}\right)=0$ |  | vs | $H_{m d s}^{1}:\left(\phi_{0}, \theta_{0}\right) \neq 0$ |  |
| $L M_{T}^{A R M A}$ | 0.049 | 0.267 |  | 0.151 | 0.039 |
| Wald | 0.000 | 0.000 |  | 0.000 | 0.000 |
| Hong ${ }^{\text {mds }}$ | 0.583 | 0.728 |  | 0.324 | 0.551 |

Note: See text for data description. Estimates and tests based on $Q_{T}^{m d s}$ with $m=\left\lfloor T^{1 / 3}\right\rfloor+1=6$. Standard error of estimates in parenthesis. For tests, p-values are reported, in bold when lower than 0.05.

### 3.8. Appendix A: Proofs of Section 2

Proof of Theorem 1. The third-order spectral density of the $m d s \varepsilon_{t}$ satisfies

$$
\begin{aligned}
f_{3}^{\varepsilon}(\lambda)= & \frac{1}{(2 \pi)^{2}} \sum_{j, k=-\infty}^{\infty} \kappa_{3}^{\varepsilon}(j, k) \exp \left(-i\left(j \lambda_{1}+k \lambda_{2}\right)\right) \\
= & \frac{1}{(2 \pi)^{2}} \sum_{j=0}^{\infty} \kappa_{3}^{\varepsilon}(j, j) \exp \left(-i j\left(\lambda_{1}+\lambda_{2}\right)\right) \\
& +\frac{1}{(2 \pi)^{2}} \sum_{j=-\infty}^{-1} \kappa_{3}^{\varepsilon}(j, 0)\left\{\exp \left(-i j \lambda_{1}\right)+\exp \left(-i j \lambda_{2}\right)\right\}, \\
= & \frac{\kappa_{3}^{\varepsilon}(0,0)}{(2 \pi)^{2}}+\sum_{j=1}^{\infty} \frac{\kappa_{3}^{\varepsilon}(j, j)}{(2 \pi)^{2}}\left\{\exp \left(-i j\left(\lambda_{1}+\lambda_{2}\right)\right)+\exp \left(i j \lambda_{1}\right)+\exp \left(i j \lambda_{2}\right)\right\}
\end{aligned}
$$

and the joint cumulants $\kappa_{3}^{X}(r, s)$ are obtained inverting $f_{3}^{X}(\lambda)$ as

$$
\begin{aligned}
\kappa_{3}^{X}(r, s) & =\int f_{3}^{X}(\lambda) \exp \left(i\left(r \lambda_{1}+s \lambda_{2}\right)\right) d \lambda_{1} d \lambda_{2} \\
& =\int f_{3}^{\varepsilon}(\lambda) \psi\left(\lambda_{1}\right) \psi\left(\lambda_{2}\right) \psi\left(-\lambda_{1}-\lambda_{2}\right) \exp \left(i\left(r \lambda_{1}+s \lambda_{2}\right)\right) d \lambda_{1} d \lambda_{2} \\
& =\sum_{j=0}^{\infty} \frac{\kappa_{3}^{\varepsilon}(j, j)}{(2 \pi)^{2}} \int \exp \left(-i(j-r) \lambda_{1}-i(j-s) \lambda_{2}\right) \psi\left(\lambda_{1}\right) \psi\left(\lambda_{2}\right) \psi\left(-\lambda_{1}-\lambda_{2}\right) d \lambda_{1} d \lambda_{2} \\
& +\sum_{j=-\infty}^{-1} \frac{\kappa_{3}^{\varepsilon}(j, 0)}{(2 \pi)^{2}} \int \exp \left(-i(j-r) \lambda_{1}\right)+\exp \left(-i(j-s) \lambda_{2}\right) \psi\left(\lambda_{1}\right) \psi\left(\lambda_{2}\right) \psi\left(-\lambda_{1}-\lambda_{2}\right) d \lambda_{1} d \lambda_{2} .
\end{aligned}
$$

Then for a general $m d s \varepsilon_{t}$ with $\kappa_{3}^{\varepsilon}(0,0) \neq 0$ and for which the future $\varepsilon_{t+\ell}^{2}$ could be predicted by the past level $\varepsilon_{t}$, so that it is only possible that for some $m>0$

$$
\kappa_{3}^{\varepsilon}(m, m)=\kappa_{3}^{\varepsilon}(-m, 0)=\kappa_{3}^{\varepsilon}(0,-m) \neq 0 .
$$

Then, we can decompose $\kappa_{3}^{X}(j, k)$ into the contributions of $\kappa_{3}^{\varepsilon}(0,0)$ and $\kappa_{3}^{\varepsilon}(m, m)$ for $m>0$, i.e.

$$
\kappa_{3}^{X}(j, k)=\kappa_{X, 3}^{[i d]}(j, k)+\sum_{m=1}^{\infty} \kappa_{X, 3}^{[m]}(j, k),
$$

where the additional terms wrt the iid case due to predictability of squares are, $m=1,2, \ldots$,

$$
\begin{aligned}
\kappa_{X, 3}^{[m]}(j, k) & =\frac{\kappa_{3}^{\varepsilon}(m, m)}{(2 \pi)^{2}} \int\left\{e^{-i\left(\lambda_{1}+\lambda_{2}\right) m}+e^{i \lambda_{1} m}+e^{i \lambda_{2} m}\right\} \psi\left(e^{-i \lambda_{1}}\right) \psi\left(e^{-i \lambda_{2}}\right) \psi\left(e^{i\left(\lambda_{1}+\lambda_{2}\right)}\right) e^{i\left(j \lambda_{1}+k \lambda_{2}\right)} d \lambda_{1} d \lambda_{2} \\
& =\frac{\kappa_{3}^{\varepsilon}(m, m)}{(2 \pi)^{2}} \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \psi_{a} \psi_{b} \psi_{c} \int\left\{\begin{array}{c}
e^{i \lambda_{1}(-a+c+j-m)} e^{i \lambda_{2}(-b+c+k-m)} \\
+e^{i \lambda_{1}(-a+c+j+m)} e^{i \lambda_{2}(-+c+c+k)} \\
+e^{\lambda_{1}(-a+c+j)} e^{i \lambda_{2}(-b+c+k+m)}
\end{array}\right\} d \lambda_{1} d \lambda_{2} \\
& =\kappa_{3}^{\varepsilon}(m, m)\left\{\begin{array}{c}
\sum_{c=\max \{0,-j-m,-k-m\}}^{\infty} \psi_{c+j-m} \psi_{c+k-m} \psi_{c}+\sum_{c=\max \{0,-j+m,-k\}}^{\infty} \psi_{c+j+m} \psi_{c+k} \psi_{c} \\
+\sum_{c=\max \{0,-j,-k+m\}}^{\infty} \psi_{c+j}^{\infty} \psi_{c+k+m}^{\infty} \psi_{c}
\end{array}\right\} \\
= & \frac{\kappa_{3}^{\varepsilon}(m, m)}{\kappa_{3}^{\varepsilon}(0,0)}\left\{\kappa_{X}^{[i i d]}(j-m, k-m)+\kappa_{X}^{[i i d]}(j+m, k)+\kappa_{X}^{[i i d]}(j, k+m)\right\},
\end{aligned}
$$

and the theorem follows.

The additional terms in the fourth-order joint cumulant of an all-pass model as described in Theorem 2 are, $m=1,2, \ldots$,

$$
\begin{aligned}
& \kappa_{X, 4}^{[1, m]}(j, k, \ell) \\
= & \frac{\kappa_{4}^{\varepsilon}(m, m, m)}{\kappa_{4}^{\varepsilon}(0,0,0)}\left\{\kappa_{X, 4}^{[i i d]}(j-m, k-m, \ell-m)+\kappa_{X, 4}^{[i i d]}(j+m, k, \ell)+\kappa_{X, 4}^{[i i d]}(j, k+m, \ell)+\kappa_{X, 4}^{[i i d]}(j, k, \ell+m)\right\}
\end{aligned}
$$

while $\kappa_{X, 4}^{[2, m, n]}(j, k, \ell)$ for $m>0, n<m$ is given by

$$
\begin{aligned}
& \kappa_{X, 4}^{[2, m, n]}(j, k, \ell) \\
= & \frac{\kappa_{4}^{\varepsilon}(m, m, n)}{\kappa_{4}^{\varepsilon}(0,0,0)}\left\{\kappa_{X, 4}^{[i i d]}(j-m, k-m, \ell-n)+\kappa_{X, 4}^{[i i d]}(j-m, k-n, \ell-m)+\kappa_{X, 4}^{[i i d]}(j-n, k-m, \ell-m)\right\},
\end{aligned}
$$

and $\kappa_{X, 4}^{[3, m, n]}(j, k, \ell)$ for $m>0, n>0$ is given by

$$
\begin{aligned}
& \kappa_{X, 4}^{[3, m, n]}(j, k, \ell) \\
= & \frac{\kappa_{4}^{\varepsilon}(0,-m,-n)}{\kappa_{4}^{\varepsilon}(0,0,0)}\left\{\kappa_{X, 4}^{[i i d]}(j, k+m, \ell+n)+\kappa_{X, 4}^{[i i d]}(j+m, k, \ell+n)+\kappa_{X, 4}^{[i i d]}(j+m, k+n, \ell)\right\} .
\end{aligned}
$$

Proof of Theorem 2. The fourth-order spectral density of a $m d s \varepsilon_{t}, f_{4}^{\varepsilon}(\lambda)$, is given by

$$
\begin{aligned}
& \frac{1}{(2 \pi)^{3}} \sum_{j, k, \ell=-\infty}^{\infty} \kappa_{4}^{\varepsilon}(j, k, \ell) \exp \left(-i\left(j \lambda_{1}+k \lambda_{2}+\ell \lambda_{3}\right)\right) \\
= & \sum_{j=0}^{\infty} \frac{\kappa_{4}^{\varepsilon}(j, j, j)}{(2 \pi)^{3}} \exp \left(-i j\left(\lambda_{1}+\lambda_{2}+\lambda_{3}\right)\right)+\sum_{j=-\infty}^{-1} \frac{\kappa_{4}^{\varepsilon}(j, 0,0)}{(2 \pi)^{3}}\left\{\begin{array}{c}
\exp \left(-i j \lambda_{1}\right)+\exp \left(-i j \lambda_{2}\right) \\
+\exp \left(-i j \lambda_{3}\right)
\end{array}\right\} \\
+ & \sum_{j=1}^{\infty} \sum_{k=-\infty}^{j-1} \frac{\kappa_{4}^{\varepsilon}(j, j, k)}{(2 \pi)^{3}}\left\{\exp \left(-i j\left(\lambda_{1}+\lambda_{2}\right)-i k \lambda_{3}\right)+\exp \left(-i j\left(\lambda_{1}+\lambda_{3}\right)-i k \lambda_{2}\right)+\exp \left(-i j\left(\lambda_{2}+\lambda_{3}\right)-i k \lambda_{1}\right)\right\} \\
+ & \sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{-1} \frac{\kappa_{4}^{\varepsilon}(0, j, k)}{(2 \pi)^{3}}\left\{\exp \left(-i j \lambda_{2}-i k \lambda_{3}\right)+\exp \left(-i j \lambda_{1}-i k \lambda_{3}\right)+\exp \left(-i j \lambda_{1}-i k \lambda_{2}\right)\right\} \\
= & \frac{\kappa_{4}^{\varepsilon}(0,0,0)}{(2 \pi)^{3}}+\frac{1}{(2 \pi)^{3}} \sum_{j=1}^{\infty} \frac{\kappa_{4}^{\varepsilon}(j, j, j)}{(2 \pi)^{3}}\left\{\exp \left(-i j\left(\lambda_{1}+\lambda_{2}+\lambda_{3}\right)\right)+\exp \left(i j \lambda_{1}\right)+\exp \left(i j \lambda_{2}\right)+\exp \left(i j \lambda_{3}\right)\right\} \\
+ & \sum_{j=1}^{\infty} \sum_{k=1-j}^{\infty} \frac{\kappa_{4}^{\varepsilon}(j, j,-k)}{(2 \pi)^{3}}\left\{\exp \left(-i j\left(\lambda_{1}+\lambda_{2}\right)+i k \lambda_{3}\right)+\exp \left(-i j\left(\lambda_{1}+\lambda_{3}\right)+i k \lambda_{2}\right)+\exp \left(-i j\left(\lambda_{2}+\lambda_{3}\right)+i k \lambda_{1}\right)\right\} \\
+ & \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{\kappa_{4}^{\varepsilon}(0,-j,-k)}{(2 \pi)^{3}}\left\{\exp \left(i j \lambda_{2}+i k \lambda_{3}\right)+\exp \left(i j \lambda_{1}+i k \lambda_{3}\right)+\exp \left(i j \lambda_{1}+i k \lambda_{2}\right)\right\},
\end{aligned}
$$

and denoting $\psi^{[4]}:=\psi\left(\lambda_{1}\right) \psi\left(\lambda_{2}\right) \psi\left(\lambda_{3}\right) \psi\left(-\lambda_{1}-\lambda_{2}-\lambda_{3}\right)$ for $\psi(\lambda)=\psi\left(e^{-i \lambda}\right)$ and
$d \lambda^{[4]}:=d \lambda_{1} d \lambda_{2} d \lambda_{3}$, its inverse is

$$
\begin{aligned}
\kappa_{4}^{X}(r, s, t) & =\int f_{4}^{X}(\lambda) \exp \left(i\left(r \lambda_{1}+s \lambda_{2}+t \lambda_{3}\right)\right) d \lambda_{1} d \lambda_{2} d \lambda_{3} \\
& =\int f_{3}^{\varepsilon}(\lambda) \psi\left(\lambda_{1}\right) \psi\left(\lambda_{2}\right) \psi\left(\lambda_{3}\right) \psi\left(-\lambda_{1}-\lambda_{2}-\lambda_{3}\right) \exp \left(i\left(r \lambda_{1}+s \lambda_{2}+t \lambda_{3}\right)\right) d \lambda_{1} d \lambda_{2} d \lambda_{3} \\
& =\sum_{j=0}^{\infty} \frac{\kappa_{4}^{\varepsilon}(j, j, j)}{(2 \pi)^{3}} \int \exp \left(-i(j-r) \lambda_{1}-i(j-s) \lambda_{2}-i(j-t) \lambda_{3}\right) \psi^{[4]} d \lambda^{[4]} \\
& +\sum_{j=-\infty}^{-1} \frac{\kappa_{4}^{\varepsilon}(j, 0,0)}{(2 \pi)^{3}} \int\left\{\exp \left(-i(j-r) \lambda_{1}\right)+\exp \left(-i(j-s) \lambda_{2}+\exp \left(-i(j-t) \lambda_{3}\right)\right)\right\} \psi^{[4]} d \lambda^{[4]} \\
& +\sum_{j=1}^{\infty} \sum_{k=-\infty}^{j-1} \frac{\kappa_{4}^{\varepsilon}(j, j, k)}{(2 \pi)^{3}} \int\left\{\begin{array}{c}
\exp \left(-i(j-r) \lambda_{1}-i(j-s) \lambda_{2}-i(k-t) \lambda_{3}\right) \\
+\exp \left(-i(j-r) \lambda_{1}-i(j-t) \lambda_{3}-i(k-s) \lambda_{2}\right) \\
+\exp \left(-i(j-s) \lambda_{2}-i(j-t) \lambda_{3}-i(k-r) \lambda_{1}\right)
\end{array}\right\} \psi^{[4]} d \lambda^{[4]} \\
& +\sum_{j=-\infty}^{-1} \sum_{k=-\infty}^{-1} \frac{\kappa_{4}^{\varepsilon}(0, j, k)}{(2 \pi)^{3}} \int\left\{\begin{array}{c}
\exp \left(-i(j-s) \lambda_{2}-i(k-t) \lambda_{3}+i r \lambda_{1}\right) \\
+\exp \left(-i(j-r) \lambda_{1}-i(k-t) \lambda_{3}+i s \lambda_{2}\right) \\
+\exp \left(-i(j-r) \lambda_{1}-i(k-s) \lambda_{2}+i t \lambda_{3}\right)
\end{array}\right\} \psi^{[4]} d \lambda^{[4]} .
\end{aligned}
$$

Then we can decompose $\kappa_{4}^{X}(j, k, \ell)$ into the contributions of $\kappa_{4}^{\varepsilon}:=\kappa_{4}^{\varepsilon}(0,0,0)$ (iid contribution) and those of $\kappa_{4}^{\varepsilon}(m, m, m)$ for $m>0, \kappa_{4}^{\varepsilon}(m, m, n)$ for $m>0, n<m$ and $\kappa_{4}^{\varepsilon}(0,-m,-n)$ for $m>0, n>0$ i.e.

$$
\begin{aligned}
\kappa_{4}^{X}(j, k, \ell)= & \kappa_{X, 4}^{[i i d]}(j, k, \ell)+\sum_{m=1}^{\infty} \kappa_{X, 4}^{[1, m]}(j, k, \ell) \\
& +\sum_{m=1}^{\infty} \sum_{n=-\infty}^{n-1} \kappa_{X, 4}^{[2, m, n]}(j, k, \ell)+\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \kappa_{X, 4}^{[3, m, n]}(j, k, \ell)
\end{aligned}
$$

where the additional terms wrt the iid case due to predictability of squares are,
$\kappa_{X, 4}^{[1, m]}(j, k, \ell), m=1,2, \ldots$, given by

$$
\left.\begin{array}{rl} 
& \kappa_{X, 4}^{[1, m]}(j, k, \ell) \\
= & \frac{\kappa_{4}^{\varepsilon}(m, m, m)}{(2 \pi)^{3}} \int\left\{e^{-i\left(\lambda_{1}+\lambda_{2}+\lambda_{3}\right) m}+e^{i \lambda_{1} m}+e^{i \lambda_{2} m}+e^{i \lambda_{3} m}\right\} e^{i\left(j \lambda_{1}+k \lambda_{2}+\ell \lambda_{3}\right)} \psi^{[4]} d \lambda^{[4]} \\
= & \frac{\kappa_{4}^{\varepsilon}(m, m, m)}{(2 \pi)^{3}} \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \sum_{d=0}^{\infty} \psi_{a} \psi_{b} \psi_{c} \psi_{d} \int\left\{\begin{array}{c}
\left.e^{i \lambda_{1}(-a+d+j-m)} \begin{array}{c}
i \lambda_{2}(-b+d+k-m) \\
+e^{i \lambda_{3}(-c+d+\ell-m)} \\
+e^{i \lambda_{1}(-a+d+j+m)} e^{i \lambda_{2}(-b+d+k)} e^{i \lambda_{3}(-c+d+\ell)} \\
+e^{i \lambda_{1}(-a+d+j)} e^{i \lambda_{2}(-b+d+k)} e^{i \lambda_{3}(-c+d+\ell+m)}
\end{array}\right\} d \lambda^{[4]} \\
= \\
= \\
= \\
\kappa_{4}^{\varepsilon}(m, m, m)\left\{\begin{array}{c}
\sum_{d=\max \{0,-j+m,-k+m,-\ell+m\}}^{\infty} \psi_{d+j-m} \psi_{d+k-m} \psi_{d+\ell-m} \psi_{d} \\
+\sum_{d=\max \{0,-j-m,-k,-\ell\}}^{\infty} \psi_{d+j+m} \psi_{d+k} \psi_{d+\ell} \psi_{d} \\
+\sum_{d=\max \{0,-j,-k-m,-\ell\}}^{\infty} \psi_{d+j} \psi_{d+k+m} \psi_{d+\ell} \psi_{d} \\
+\sum_{d=\max \{0,-j,-k,-\ell-m\}}^{\infty} \psi_{d+j} \psi_{d+k} \psi_{d+\ell+m} \psi_{d}
\end{array}\right\}
\end{array}\right\} \\
\kappa_{4}^{\infty}(0,0,0)
\end{array} \kappa_{X, 4}^{[i i d]}(j-m, k-m, \ell-m)+\kappa_{X, 4}^{[i i d]}(j+m, k, \ell)+\kappa_{X, 4}^{[i i d]}(j, k+m, \ell)+\kappa_{X, 4}^{[i i d]}(j, k, \ell+m)\right\} .
$$

while $\kappa_{X, 4}^{[2, m, n]}(j, k, \ell)$ for $m>0, n<m$ is given by

$$
\left.\begin{array}{rl} 
& \kappa_{X, 4}^{[2, m, n]}(j, k, \ell) \\
= & \frac{\kappa_{4}^{\varepsilon}(m, m, n)}{(2 \pi)^{3}} \int\left\{e^{-i\left(\lambda_{1}+\lambda_{2}\right) m-i \lambda_{3} n}+e^{-i\left(\lambda_{1}+\lambda_{3}\right) m-i \lambda_{2} n}+e^{-i\left(\lambda_{2}+\lambda_{3}\right) m-i \lambda_{1} n}\right\} e^{i\left(j \lambda_{1}+k \lambda_{2}+\ell \lambda_{3}\right)} \psi^{[4]} d \lambda^{[4]} \\
= & \frac{\kappa_{4}^{\varepsilon}(m, m, n)}{(2 \pi)^{3}} \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \sum_{d=0}^{\infty} \psi_{a} \psi_{b} \psi_{c} \psi_{d} \int\left\{\begin{array}{c}
e^{-i \lambda_{1}(m+a-d-j)-i \lambda_{2}(m+b-d-k)-i \lambda_{3}(n+c-d-\ell)} \\
+e^{-i \lambda_{1}(m+a-d-j)-i \lambda_{2}(n+b-d-k)-i \lambda_{3}(m+c-d-\ell)} \\
+e^{-i \lambda_{1}(n+a-d-j)-i \lambda_{2}(m+b-d-k)-i \lambda_{3}(m+c-d-\ell)}
\end{array}\right\} d \lambda^{[4]}
\end{array}\right\} \begin{aligned}
& =\left\{\begin{array}{l}
\sum_{d=\max \{0,-j+m,-k+m,-\ell+n\}}^{\infty} \psi_{d+j-m} \psi_{d+k-m}^{\varepsilon} \psi_{d+\ell-n} \psi_{d} \\
+\sum_{d=\max \{0,-j+m,-k+n,-\ell+m\}}^{\infty} \psi_{d+j-m} \psi_{d+k-n} \psi_{d+\ell-m} \psi_{d} \\
+\sum_{d=\max \{0,-j+n,-k+m,-\ell+m\}}^{\infty} \psi_{d+j-n} \psi_{d+k-m} \psi_{d+\ell-m} \psi_{d}
\end{array}\right\} \\
& = \\
& =\frac{\kappa_{3}^{\varepsilon}(m, m, n)}{\kappa_{4}^{\varepsilon}(0,0,0)}\left\{\kappa_{X, 4}^{[i i d]}(j-m, k-m, \ell-n)+\kappa_{X, 4}^{[i i d]}(j-m, k-n, \ell-m)+\kappa_{X, 4}^{[i i d]}(j-n, k-m, \ell-m)\right\},
\end{aligned}
$$

and $\kappa_{X, 4}^{[3, m, n]}(j, k, \ell)$ for $m>0, n>0$ is given by

$$
\left.\begin{array}{rl} 
& \kappa_{X, 4}^{[3, m, n]}(j, k, \ell) \\
= & \frac{\kappa_{4}^{\varepsilon}(0,-m,-n)}{(2 \pi)^{3}} \int\left\{e^{i \lambda_{2} m+i \lambda_{3} n}+e^{i \lambda_{1} m+i \lambda_{3} n}+e^{i \lambda_{1} m+i \lambda_{2} n}\right\} e^{i\left(j \lambda_{1}+k \lambda_{2}+\ell \lambda_{3}\right)} \psi^{[4]} d \lambda^{[4]} \\
= & \frac{\kappa_{4}^{\varepsilon}(0,-m,-n)}{(2 \pi)^{3}} \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \sum_{c=0}^{\infty} \sum_{d=0}^{\infty} \psi_{a} \psi_{b} \psi_{c} \psi_{d} \int\left\{\begin{array}{c}
e^{i \lambda_{1}(j-a+d)+i \lambda_{2}(m+k-b+d)+i \lambda_{3}(n+\ell-c+d)} \\
+e^{i \lambda_{1}(m+j-a+d)+i \lambda_{2}(k-b+d) i \lambda_{3}(n+\ell-c+d)} \\
+e^{i \lambda_{1}(m+j-a+d)+i \lambda_{2}(n+k-b+d)+i \lambda_{3}(\ell-c+d)}
\end{array}\right\} d \lambda^{[4]}
\end{array}\right\} \begin{aligned}
& \left.=\begin{array}{l}
\sum_{d=\max \{0,-j,-k-m,-\ell-n\}}^{\infty} \psi_{d+j} \psi_{d+k+m} \psi_{d+\ell+n} \psi_{d} \\
+\sum_{d=\max \{0,-j-m,-k,-\ell-n\}}^{\infty} \psi_{d+j+m} \psi_{d+k} \psi_{d+\ell+n} \psi_{d} \\
+\sum_{d=\max \{0,-j-m,-k-n,-\ell\}}^{\infty} \psi_{d+j+m} \psi_{d+k+n} \psi_{d+\ell} \psi_{d}
\end{array}\right\} \\
& = \\
& =\frac{\kappa_{4}^{\varepsilon}(0,-m,-n)}{\kappa_{4}^{\varepsilon}(0,0,0)}\left\{\kappa_{X, 4}^{[i i d]}(j, k+m, \ell+n)+\kappa_{X, 4}^{[i i d]}(j+m, k, \ell+n)+\kappa_{X, 4}^{[i i d]}(j+m, k+n, \ell)\right\}
\end{aligned}
$$

and the theorem follows. $\square$

Proof of Theorem 3. The squares of $\varepsilon_{t}$ have autocorrelation function given for $j \neq 0$ by

$$
\begin{aligned}
\mathbb{C}\left(\varepsilon_{t}^{2}, \varepsilon_{t+j}^{2}\right) & =E\left[\varepsilon_{t}^{2} \varepsilon_{t+j}^{2}\right]-\sigma^{4} \\
& =\underbrace{E\left[\varepsilon_{t}^{2}\right]}_{=\sigma^{2}} \underbrace{E\left[\varepsilon_{t+j}^{2}\right]}_{=\sigma^{2}}+2 \underbrace{E\left[\varepsilon_{t} \varepsilon_{t+j}\right]^{2}}_{=0}+\operatorname{cum}\left[\varepsilon_{t}, \varepsilon_{t}, \varepsilon_{t+j}, \varepsilon_{t+j}\right]-\sigma^{4} \\
& =\operatorname{cum}\left[\varepsilon_{t}, \varepsilon_{t}, \varepsilon_{t+j}, \varepsilon_{t+j}\right] \\
& =\kappa_{4}^{\varepsilon}(0, j, j),
\end{aligned}
$$

while for $j=0, \mathbb{C}\left(\varepsilon_{t}^{2}, \varepsilon_{t}^{2}\right)=\mathbb{V}\left(\varepsilon_{t}^{2}\right)=2 \sigma_{\varepsilon}^{4}+\kappa_{4}^{\varepsilon}(0,0,0)$, so that

$$
f_{2}^{\varepsilon^{2}}(\lambda)=\frac{1}{2 \pi} \sum_{j=-\infty}^{\infty} \kappa_{4}^{\varepsilon}(0, j, j) \exp (-i j \lambda)+\frac{2 \sigma_{\varepsilon}^{4}}{2 \pi}
$$

which is closely related to the 4 -th order spectral density of the levels $\varepsilon_{t}$,

$$
f_{4}^{\varepsilon}\left(\lambda_{1}, \lambda_{2}, \lambda_{3}\right)=\frac{1}{(2 \pi)^{3}} \sum_{a, b, c=-\infty}^{\infty} \kappa_{4}^{\varepsilon}(a, b, c) \exp \left(-i a \lambda_{1}-i b \lambda_{2}-i c \lambda_{3}\right)
$$

because

$$
\begin{aligned}
\int_{\Pi^{2}} f_{4}^{\varepsilon}\left(\lambda_{1}, \lambda_{2}, \lambda-\lambda_{2}\right) d \lambda_{1} d \lambda_{2} & =\sum_{a, b, c=-\infty}^{\infty} \frac{\kappa_{4}^{\varepsilon}(a, b, c)}{(2 \pi)^{3}} \int_{\Pi^{2}} \exp \left(-i a \lambda_{1}-i b \lambda_{2}-i c\left(\lambda-\lambda_{2}\right)\right) d \lambda_{1} d \lambda_{2} \\
& =\sum_{a, b, c=-\infty}^{\infty} \frac{\kappa_{4}^{\varepsilon}(a, b, c)}{(2 \pi)^{3}} \exp (-i c \lambda) \int_{\Pi^{2}} \exp \left(-i a \lambda_{1}-i(b-c) \lambda_{2}\right) d \lambda_{1} d \lambda_{2} \\
& =\sum_{a, b, c=-\infty}^{\infty} \frac{\kappa_{4}^{\varepsilon}(a, b, c)}{2 \pi} \exp (-i c \lambda)(2 \pi)^{2} 1\{a=0\} 1\{b=c\} \\
& =\sum_{c=-\infty}^{\infty} \frac{\kappa_{4}^{\varepsilon}(0, c, c)}{2 \pi} \exp (-i c \lambda) \\
& =f_{2}^{\varepsilon^{2}}(\lambda)-\frac{2 \sigma_{\varepsilon}^{4}}{2 \pi}
\end{aligned}
$$

We now proceed for $p=2$ to reduce notation complexity. Assume that the causal AR inverse roots satisfy $\left|\phi_{j}\right|<1$ and the noninvertible MA ones are given by $\psi_{j}=1 / \phi_{j}$, so that

$$
\begin{aligned}
& f_{4}^{\varepsilon}\left(\lambda_{1}, \lambda_{2}, \lambda_{3}\right) \\
= & \frac{\kappa_{4}}{(2 \pi)^{3}} \prod_{j=1}^{2} \frac{\left(1+\psi_{j} \exp \left(i \lambda_{1}\right)\right)\left(1+\psi_{j} \exp \left(i \lambda_{2}\right)\right)\left(1+\psi_{j} \exp \left(i \lambda_{3}\right)\right)\left(1+\psi_{j} \exp \left(-i\left(\lambda_{1}+\lambda_{2}+\lambda_{3}\right)\right)\right)}{\left(1+\phi_{j} \exp \left(i \lambda_{1}\right)\right)\left(1+\phi_{j} \exp \left(i \lambda_{2}\right)\right)\left(1+\phi_{j} \exp \left(i \lambda_{3}\right)\right)\left(1+\phi_{j} \exp \left(-i\left(\lambda_{1}+\lambda_{2}+\lambda_{3}\right)\right)\right)}
\end{aligned}
$$

and, $\rho:=\phi_{1} \phi_{2}^{-1}, 0<|\rho|<1$ or $0<\left|\phi_{1}\right|<\left|\phi_{2}\right|<1$,

$$
\begin{aligned}
& f_{2}^{\varepsilon^{2}}(\lambda)-\frac{\sigma^{4}}{2 \pi} \\
= & \int_{\Pi^{2}} f_{4}^{X}\left(\lambda_{1}, \lambda_{2}, \lambda-\lambda_{2}\right) d \lambda_{1} d \lambda_{2} \\
= & \frac{\kappa_{4}}{2 \pi} \frac{1}{2 \pi} \int_{\Pi} \prod_{j=1}^{2} \frac{\left(1+\psi_{j} \exp \left(i \lambda_{2}\right)\right)\left(1+\psi_{j} \exp \left(i\left(\lambda-\lambda_{2}\right)\right)\right)}{\left(1+\phi_{j} \exp \left(i \lambda_{2}\right)\right)\left(1+\phi_{j} \exp \left(i\left(\lambda-\lambda_{2}\right)\right)\right)} d \lambda_{2} \\
& \times \frac{1}{2 \pi} \int_{\Pi} \prod_{j=1}^{2} \frac{\left(1+\psi_{j} \exp \left(i \lambda_{1}\right)\right)\left(1+\psi_{j} \exp \left(-i\left(\lambda+\lambda_{1}\right)\right)\right)}{\left(1+\phi_{j} \exp \left(i \lambda_{1}\right)\right)\left(1+\phi_{j} \exp \left(-i\left(\lambda+\lambda_{1}\right)\right)\right)} d \lambda_{1} \\
= & \frac{\kappa_{4}}{2 \pi\left(1-\left(\phi_{1} \phi_{2}^{-1}\right)\right)^{2}} \frac{1+\psi_{1} \psi_{2}\left\{\psi_{2}^{2}-\rho-1\right\}+\exp (i \lambda)\left\{\psi_{2}^{2}+\psi_{1} \psi_{2}-\rho^{-1}-1\right\}}{+\exp (i 2 \lambda)\left\{\psi_{1}^{2} \psi_{2}^{2}-\psi_{1}\left(\psi_{1}+\psi_{2}\right)+\rho^{-1}\right\}} \begin{array}{|c|l|l|l}
2
\end{array} \\
& \quad\left(1-\left(\phi_{2} \phi_{1}\right) \exp (i \lambda)\right)\left(1-\left.\left(\phi_{2}\right)^{2} \exp (i \lambda)\right|^{2}\right.
\end{aligned}
$$

because

$$
\begin{aligned}
& \frac{1}{2 \pi} \int_{\Pi} \prod_{j=1}^{2} \frac{\left(1+\psi_{j} \exp \left(i \lambda_{2}\right)\right)\left(1+\psi_{j} \exp \left(i\left(\lambda-\lambda_{2}\right)\right)\right)}{\left(1+\phi_{j} \exp \left(i \lambda_{2}\right)\right)\left(1+\phi_{j} \exp \left(i\left(\lambda-\lambda_{2}\right)\right)\right)} d \lambda_{2} \\
= & \frac{1}{2 \pi} \int_{\Pi}\left(1+\psi_{1} \exp \left(i \lambda_{2}\right)\right)\left(1+\psi_{2} \exp \left(i \lambda_{2}\right)\right) \\
& \times\left(1+\psi_{1} \exp \left(i\left(\lambda-\lambda_{2}\right)\right)\right)\left(1+\psi_{2} \exp \left(i\left(\lambda-\lambda_{2}\right)\right)\right) \\
& \times\left(\sum_{j=0}^{\infty}\left(-\phi_{1}\right)^{j} \exp \left(i j \lambda_{2}\right)\right)\left(\sum_{j=0}^{\infty}\left(-\phi_{2}\right)^{j} \exp \left(i j \lambda_{2}\right)\right) \\
& \times\left(\sum_{j=0}^{\infty}\left(-\phi_{1}\right)^{j} \exp \left(i j\left(\lambda-\lambda_{2}\right)\right)\right)\left(\sum_{j=0}^{\infty}\left(-\phi_{2}\right)^{j} \exp \left(i j\left(\lambda-\lambda_{2}\right)\right)\right) d \lambda_{2} \\
= & \frac{1+\psi_{1} \psi_{2}\left\{\psi_{2}^{2}-\rho-1\right\}+\exp (i \lambda)\left\{\psi_{2}^{2}+\psi_{1} \psi_{2}-\rho^{-1}-1\right\}+\exp (i 2 \lambda)\left\{\psi_{1}^{2} \psi_{2}^{2}-\psi_{1}\left(\psi_{1}+\psi_{2}\right)+\rho^{-1}\right\}}{\left(1-\left(\phi_{1} \phi_{2}^{-1}\right)\right)^{2}\left(1-\left(\phi_{2} \phi_{1}\right) \exp (i \lambda)\right)\left(1-\left(\phi_{2}\right)^{2} \exp (i \lambda)\right)}
\end{aligned}
$$

which can be showed studying the $16=2^{4}$ terms obtained after making the multiplication of the 4 binomials in the numerator, while the denominator corresponds to $\mathrm{AR}(2)$ dynamics.

The first six terms not depending on $\lambda_{2}$ are,

$$
\begin{aligned}
&\left(1+\left(\psi_{1}+\psi_{2}\right)^{2} \exp (i \lambda)+\psi_{1}^{2} \psi_{2}^{2} \exp (i 2 \lambda)\right) \\
& \times \frac{1}{2 \pi} \int_{\Pi}\left(\sum_{j_{1}=0}^{\infty}\left(-\phi_{1}\right)^{j_{1}} \exp \left(i j_{1} \lambda_{2}\right)\right)\left(\sum_{j_{2}=0}^{\infty}\left(-\phi_{2}\right)^{j_{2}} \exp \left(i j_{2} \lambda_{2}\right)\right) \\
& \times\left(\sum_{j_{3}=0}^{\infty}\left(-\phi_{1}\right)^{j_{3}} \exp \left(i j_{3}\left(\lambda-\lambda_{2}\right)\right)\right)\left(\sum_{j_{4}=0}^{\infty}\left(-\phi_{2}\right)^{j_{4}} \exp \left(i j_{4}\left(\lambda-\lambda_{2}\right)\right)\right) d \lambda_{2} \\
&=\left(1+\left(\psi_{1}+\psi_{2}\right)^{2} \exp (i \lambda)+\psi_{1}^{2} \psi_{2}^{2} \exp (i 2 \lambda)\right) \\
& \times\left(\sum_{j_{1}=0}^{\infty}\left(\phi_{2} \phi_{1}\right)^{j_{1}} \exp \left(i j_{1} \lambda\right)\right)\left(\sum_{j_{2}=0}^{\infty}\left(\phi_{2}\right)^{2 j_{2}} \exp \left(i j_{2} \lambda\right)\right)\left(\sum_{j_{3}=0}^{\infty}\left(\phi_{1} \phi_{2}^{-1}\right)^{j_{3}}\right) \\
&=\frac{\left(1+\left(\psi_{1}+\psi_{2}\right)^{2} \exp (i \lambda)+\psi_{1}^{2} \psi_{2}^{2} \exp (i 2 \lambda)\right)}{\left(1-\left(\phi_{2} \phi_{1}\right) \exp (i \lambda)\right)\left(1-\left(\phi_{2}\right)^{2} \exp (i \lambda)\right)\left(1-\left(\phi_{1} \phi_{2}^{-1}\right)\right)}
\end{aligned}
$$

The next 10 terms depending on powers of $\exp \left(i \lambda_{2}\right)$ are

$$
\begin{aligned}
= & \frac{1}{2 \pi} \int_{\Pi}\left\{\psi_{1}+\psi_{2}+\psi_{1} \psi_{2}\left(\psi_{1}+\psi_{2}\right) \exp (i \lambda)\right\} \exp \left(i \lambda_{2}\right) \\
& +\psi_{1} \psi_{2} \exp \left(i 2 \lambda_{2}\right) \\
& +\left\{\left(\psi_{1}+\psi_{2}\right) \exp (i \lambda)+\psi_{1} \psi_{2}\left(\psi_{1}+\psi_{2}\right) \exp (i 2 \lambda)\right\} \exp \left(-i \lambda_{2}\right) \\
& +\psi_{1} \psi_{2} \exp (i 2 \lambda) \exp \left(-i 2 \lambda_{2}\right) \\
& \times\left(\sum_{j_{1}=0}^{\infty}\left(-\phi_{1}\right)^{j_{1}} \exp \left(i j_{1} \lambda_{2}\right)\right)\left(\sum_{j_{2}=0}^{\infty}\left(-\phi_{2}\right)^{j_{2}} \exp \left(i j_{2} \lambda_{2}\right)\right) \\
& \times\left(\sum_{j_{3}=0}^{\infty}\left(-\phi_{1}\right)^{j_{3}} \exp \left(i j_{3}\left(\lambda-\lambda_{2}\right)\right)\right)\left(\sum_{j_{4}=0}^{\infty}\left(-\phi_{2}\right)^{j_{4}} \exp \left(i j_{4}\left(\lambda-\lambda_{2}\right)\right)\right) d \lambda_{2}
\end{aligned}
$$

which is equal to

$$
\begin{aligned}
& \frac{1}{\left(1-\left(\phi_{2} \phi_{1}\right) \exp (i \lambda)\right)\left(1-\left(\phi_{2}\right)^{2} \exp (i \lambda)\right)\left(1-\left(\phi_{1} \phi_{2}^{-1}\right)\right)} \\
& \times\left\{\begin{array}{c}
-\phi_{2}\left\{\psi_{1}+\psi_{2}+\psi_{1} \psi_{2}\left(\psi_{1}+\psi_{2}\right) \exp (i \lambda)\right\} \exp (i \lambda) \\
+\phi_{2}^{2} \psi_{1} \psi_{2} \exp (i 2 \lambda) \\
-\phi_{2}^{-1}\left\{\left(\psi_{1}+\psi_{2}\right) \exp (i \lambda)+\psi_{1} \psi_{2}\left(\psi_{1}+\psi_{2}\right) \exp (i 2 \lambda)\right\} \exp (-i \lambda) \\
+\phi_{2}^{-2} \psi_{1} \psi_{2} \exp (i 2 \lambda) \exp (-i 2 \lambda)
\end{array}\right\}
\end{aligned}
$$

where the numerator is, $\rho:=\phi_{1} \phi_{2}^{-1}=\psi_{2} \psi_{1}^{-1}$,

$$
\psi_{2} \psi_{1}\left\{\psi_{2}^{2}-\rho-1\right\}+\exp (i \lambda)\left\{-\psi_{2}^{2} \psi_{1}\left(\psi_{1}+\psi_{2}\right)-\rho^{-1}-1\right\}+\exp (i 2 \lambda)\left\{-\psi_{1}\left(\psi_{1}+\psi_{2}\right)+\rho^{-1}\right\}
$$

leading to causal $\operatorname{ARMA}(2,2)$ dynamics together with the previous terms, giving overall numerator of the rational part of $f_{2}^{\varepsilon^{2}}$ equal to

$$
\begin{aligned}
& 1+\left(\phi_{2} \phi_{1}\right)^{-1}\left\{\phi_{2}^{-2}-\rho-1\right\}+\exp (i \lambda)\left\{\left(\psi_{1}+\psi_{2}\right)^{2}-\psi_{2}^{2} \psi_{1}\left(\psi_{1}+\psi_{2}\right)-\rho^{-1}-1\right\} \\
& +\exp (i 2 \lambda)\left\{\psi_{1}^{2} \psi_{2}^{2}-\psi_{1}\left(\psi_{1}+\psi_{2}\right)+\rho^{-1}\right\}
\end{aligned}
$$

which is
$1+\psi_{1} \psi_{2}\left\{\psi_{2}^{2}-\rho-1\right\}+\exp (i \lambda)\left\{\psi_{1}\left(\psi_{1}+\psi_{2}\right)\left(1+\rho-\psi_{2}^{2}\right)-\rho^{-1}-1\right\}+\exp (i 2 \lambda)\left\{\psi_{1}^{2} \psi_{2}^{2}-\psi_{1}\left(\psi_{1}+\psi_{2}\right)+\rho^{-1}\right\}$,
which has to be complemented with the constant term $\frac{2 \sigma_{\varepsilon}^{4}}{2 \pi}$, which corresponds to an additive white noise and does not contribute additionally to the order of the $\operatorname{ARMA}(2,2)$ representation.

For a general $\operatorname{ARMA}(p, p)$ all pass model, a similar solution is always found

$$
f_{2, \varepsilon^{2}}(\lambda)=\kappa_{4}^{\varepsilon}(0,0,0)\left|g\left(e^{i \lambda}\right)\right|^{2}+\frac{2 \sigma_{\varepsilon}^{4}}{2 \pi}
$$

where $g\left(e^{i \lambda}\right)$ is the rational transfer function of an $\operatorname{ARMA}(p, p)$ model, because all terms of the integrals have the same denominator as a factorization of $p$ terms $1-c_{j} e^{i \lambda}$ (originating from the $p$ infinite series depending on $e^{i \lambda}$ ) which lead to a numerator which is a polynomial in $e^{i \lambda}$ of order $p$.

### 3.9. Appendix B

### 3.9.1. General ARMA(p,p) model

The general (non-invertible) ARMA $(p, p)$ model extending model (3.3.2) can be written as

$$
\Pi_{j=1}^{p}\left(1-\phi_{j} L\right) Y_{t}=\Pi_{j=1}^{p}\left(1-\theta_{j} L^{-1}\right) L^{p} \varepsilon_{t}
$$

where $0<\left|\phi_{j}\right|<1$ and $0<\left|\theta_{j}\right|<1$ for all $j$, or equivalently

$$
\begin{equation*}
\alpha(L) Y_{t}=\beta\left(L^{-1}\right) L^{p} \varepsilon_{t} \tag{3.9.1}
\end{equation*}
$$

where $\alpha(z)=1-\alpha_{1} z-\cdots-\alpha_{p} z^{p}=\Pi_{j=1}^{p}\left(1-\phi_{j} z\right)$ and $\beta(z)=1-\beta_{1} z-\cdots-\beta_{p} z^{p}=$ $\Pi_{j=1}^{p}\left(1-\theta_{j} L^{-1}\right)$ with $\alpha_{p} \neq 0$ and $\beta_{p} \neq 0$. When the model is all-pass so that $\phi_{j}=\theta_{j}$ for each $j=1, \ldots, p$ then $\alpha_{j}=\beta_{j}, j=1, \ldots, p$, and we can write

$$
\begin{equation*}
\alpha(L) Y_{t}=\alpha\left(L^{-1}\right) L^{p} \varepsilon_{t} . \tag{3.9.2}
\end{equation*}
$$

In particular, for $m d s \varepsilon_{t}$, we consider testing the $\mathbf{A P}(p, p)$ hypothesis against general non-invertible ARMA models,

$$
\begin{array}{ll}
H_{A P / m d s} & : \alpha_{j}=\beta_{j} \text { for all } j=1, \ldots, p \text { in model }(3.9 .1), \\
H_{A P / m d s}^{1} & : \alpha_{j} \neq \beta_{j}, \text { for at least one } j=1, \ldots, p
\end{array}
$$

and testing the mds hypothesis against $\mathrm{AP}(p, p)$ and

$$
\begin{aligned}
H_{m d s}^{(A P)} & : \alpha_{j}=0 \text { for all } j=1, \ldots, p \text { in model }(3.9 .2), \\
H_{m d s}^{(A P) 1} & : \alpha_{j} \neq 0, \text { for at least one } j=1, \ldots, p
\end{aligned}
$$

against unrestricted but non-invertible $\operatorname{ARMA}(p, p)$ models

$$
\begin{aligned}
H_{m d s} & : \alpha_{j}=\beta_{j}=0 \text { for all } j=1, \ldots, p \text { in model }(3.9 .1), \\
H_{m d s}^{1} & : \alpha_{j} \neq 0 \mathrm{and} / \text { or } \beta_{j} \neq 0 \text { for at least one } j=1, \ldots, p .
\end{aligned}
$$

Then, the testing will follow in a similar way, but adapting the scores to the general
model with

$$
\begin{aligned}
& \psi\left(\vartheta_{0} ; z\right)=\frac{\beta\left(z^{-1}\right)}{\alpha(z)}=\frac{1-\beta_{1} z^{-1}-\cdots-\beta_{p} z^{-p}}{1-\alpha_{1} z-\cdots-\alpha_{p} z^{p}} \text { for ARMA }(p, p) \text { model } \\
& \psi\left(\vartheta_{0} ; z\right)=\frac{\alpha\left(z^{-1}\right)}{\alpha(z)}=\frac{1-\alpha_{1} z^{-1}-\cdots-\alpha_{p} z^{-p}}{1-\alpha_{1} z-\cdots-\alpha_{p} z^{p}} \text { for } \operatorname{AP}(p, p) \text { model }
\end{aligned}
$$

for $\vartheta_{0}=\left(\alpha_{1}, \ldots, \alpha_{p}, \beta_{1}, \ldots, \beta_{p}\right)^{\prime}$ and $\vartheta_{0}=\left(\alpha_{1}, \ldots, \alpha_{p}\right)^{\prime}$, respectively, and therefore

$$
\begin{align*}
\delta\left(\vartheta_{0} ; z\right) & =-\frac{\partial}{\partial \vartheta} \log \psi\left(\vartheta_{0} ; z\right)  \tag{3.9.3}\\
& =-\frac{\partial}{\partial \vartheta} \log \left(1-\beta_{1} z^{-1}-\cdots-\beta_{p} z^{-p}\right)+\frac{\partial}{\partial \vartheta} \log \left(1-\alpha_{1} z-\cdots-\alpha_{p} z^{p}\right) \\
& =\left(\frac{z^{-1}}{\beta\left(z^{-1}\right)}, \cdots, \frac{z^{-p}}{\beta\left(z^{-1}\right)},-\frac{z}{\alpha(z)}, \cdots,-\frac{z^{p}}{\alpha(z)}\right)^{\prime} \\
& =\left(z^{-1}, \cdots, z^{-p},-z, \cdots,-z^{p}\right)^{\prime} \quad \text { when } \vartheta_{0}=\left(\alpha_{1}, \ldots, \alpha_{p}, \beta_{1}, \ldots, \beta_{p}\right)^{\prime}=\mathbf{0} \text { for ARMA }(p, p)
\end{align*}
$$

while

$$
\begin{aligned}
\delta\left(\vartheta_{0} ; z\right) & =-\frac{\partial}{\partial \vartheta} \log \psi\left(\vartheta_{0} ; z\right) \\
& =-\frac{\partial}{\partial \vartheta} \log \left(1-\alpha_{1} z^{-1}-\cdots-\alpha_{p} z^{-p}\right)+\frac{\partial}{\partial \vartheta} \log \left(1-\alpha_{1} z-\cdots-\alpha_{p} z^{p}\right) \\
& =\left(\frac{z^{-1}}{\alpha\left(z^{-1}\right)}-\frac{z}{\alpha(z)}, \cdots, \frac{z^{-p}}{\alpha\left(z^{-1}\right)}-\frac{z^{p}}{\alpha(z)}\right) \\
& =\left(z^{-1}-z, \cdots, z^{-p}-z^{p}\right)^{\prime} \quad \text { when } \vartheta_{0}=\left(\alpha_{1}, \ldots, \alpha_{p}\right)^{\prime}=\mathbf{0} \text { for AP }(p, p) .
\end{aligned}
$$

Therefore, for ARMA processes only $\zeta_{j}^{0}(v) \neq 0$ and $\beta_{j}^{0}(v) \neq 0$ for $j=1, \ldots, p$.

The LM testing of the AP model against ARMA, $H_{A P}: \phi_{0}=\theta_{0}$ in model (3.9.1) requires to parametrize $\gamma_{j}:=\alpha_{j}-\beta_{j}$ for $j=1, \ldots, p$ and test $H_{A P}^{*}: \gamma_{j}=0$ all $j=1, \ldots, p$ in

$$
\left(1-\alpha_{1} L-\cdots-\alpha_{p} L^{p}\right) Y_{t}=\left(1-\alpha_{1} L^{-1}+\gamma_{1} L^{-1}-\cdots-\alpha_{p} L^{-p}-\gamma_{p} L^{-p}\right) \varepsilon_{t-1}
$$

so that for $\vartheta_{0}^{*}=\left(\gamma_{1}, \ldots, \gamma_{p}, \alpha_{1}, \ldots, \alpha_{p}\right)^{\prime}$,

$$
\psi\left(\vartheta_{0}^{*} ; z\right)=\frac{1-\alpha_{1} z^{-1}+\gamma_{1} z^{-1}-\cdots-\alpha_{p} z^{-p}-\gamma_{p} z^{-p}}{1-\alpha_{1} z-\cdots-\alpha_{p} z^{p}} \text { for ARMA }(p, p) \text { model }
$$

and therefore

$$
\delta\left(\vartheta^{*} ; z\right)=\frac{\partial}{\partial \vartheta} \phi\left(\vartheta^{*} ; z\right)=-\frac{\frac{\partial}{\partial \vartheta} \psi\left(\vartheta^{*} ; z\right)}{\psi\left(\vartheta^{*} ; z\right)} \frac{\psi\left(\vartheta_{0}^{*} ; z\right)}{\psi\left(\vartheta^{*} ; z\right)}
$$

and

$$
\begin{aligned}
\delta\left(\vartheta_{0}^{*} ; z\right) & =-\frac{\partial}{\partial \vartheta^{*}} \log \psi\left(\vartheta_{0}^{*} ; z\right) \\
& =-\frac{\partial}{\partial \vartheta^{*}} \log \left(1-\alpha_{1} z^{-1}+\gamma_{1} z^{-1}-\cdots-\alpha_{p} z^{-p}-\gamma_{p} z^{-p}\right)+\frac{\partial}{\partial \vartheta^{*}} \log \left(1-\alpha_{1} z-\cdots-\alpha_{p} z^{p}\right) \\
& =\left(-\frac{z^{-1}}{(\alpha+\gamma)\left(z^{-1}\right)}, \cdots,-\frac{z^{-p}}{(\alpha+\gamma)\left(z^{-1}\right)}, \frac{z^{-1}}{(\alpha+\gamma)\left(z^{-1}\right)}-\frac{z}{\alpha(z)}, \cdots, \frac{z^{-p}}{(\alpha+\gamma)\left(z^{-1}\right)}-\frac{z^{p}}{\alpha(z)}\right)^{\prime} \\
& =\left(-\frac{z^{-1}}{\alpha\left(z^{-1}\right)}, \cdots,-\frac{z^{-p}}{\alpha\left(z^{-1}\right)}, \frac{z^{-1}}{\alpha\left(z^{-1}\right)}-\frac{z}{\alpha(z)}, \cdots, \frac{z^{-p}}{\alpha\left(z^{-1}\right)}-\frac{z^{p}}{\alpha(z)}\right)^{\prime}
\end{aligned}
$$

when $\vartheta_{0}^{*}=\left(\mathbf{0}^{\prime}, \alpha_{1}, \ldots, \alpha_{p}\right)^{\prime}$ for $\operatorname{AP}(p, p)$ model, so that

$$
\delta\left(\vartheta_{0}^{*} ; z\right)=A^{(p)} \delta^{A R M A}\left(\vartheta_{0} ; z\right) \text { for } A^{(p)}=\left(A \otimes I_{p}\right)=\left(\begin{array}{cc}
-I_{p} & 0 \\
I_{p} & I_{p}
\end{array}\right)
$$

with $\vartheta^{*}=A^{(p) \prime} \vartheta$ and $\vartheta=A^{(p) \prime} \vartheta^{*}$.

### 3.9.2. Technical Assumptions

Assumption 3.1. For non-Gaussian $\varepsilon_{t}$ (i.e. with nonzero third or fourth order marginal cumulant) and some $\nu \geq 5$ :

1. $\varepsilon_{t}$ is stationary mds, $\mathbb{E}\left[\varepsilon_{t} \mid I_{t-1}\right]=0$ and $\mathbb{E}\left|\varepsilon_{t}\right|^{\nu}<\infty$.
2. $\varepsilon_{t}$ is strong mixing with mixing coefficients satisfying $\sum_{j=1}^{\infty} j^{2} \alpha(j)^{\frac{\nu-4}{\nu}}<\infty$.

Assumption 3.2. $W: \mathbb{R} \rightarrow \mathbb{R}^{+}$is continuous, symmetric and increasing with unbounded support and $\int|u|^{3} d W(u)<\infty$.

## Assumption 3.3.

1. $k: \mathbb{R} \rightarrow[-1,1]$ is symmetric and continuous at 0 and all but a finite number of points, with $k(1)=1$ and $|k(x)| \leq C|x|^{-b}, b \geq 1$, for large $x$, and $1-k(x)=$ $k_{\tau}|x|^{\tau}+o(x)$ as $x \rightarrow 0$ for some $\tau \in(0, \infty)$ and $k_{\tau}>0$.
2. $1 / m+m^{2} / T \rightarrow 0$ as $T \rightarrow \infty$.

Assumption 3.1 is key for identification and to ensure standard local properties of the score of the loss function $Q_{T}^{m d s}$ under the null hypothesis of $m d s$. Assumption 3.1.2 also implies the mixing condition used in Andrews (1991) for summability of the fourth-order cumulants and the conditions in Yoshihara (1978) to bound the fourth moment of sums of mixing processes.

Assumption 3.2 on $W$ is similar to the corresponding one used in Hong (1999) to argue for the consistency of serial dependence tests and is stronger than the nondecreasing with bounded total variation condition he used to derive the null asymptotic distribution of test statistics. We also introduce a moment condition on $W$, to control fluctuations of $\left|\sigma_{\theta, j}^{(1,0)}(u, v)\right|$ in $u$ and $v$ when using derivatives of the $c f$, and a factorization, to simplify numerical calculations and asymptotic analysis.

Assumption 3.3.1 was used by Hong (1999) for the analysis of dependence tests and is standard in the related literature of smoothed spectral density estimation. Assumption 3.3.2 allows to choose $m$ for optimal MSE estimation of the generalized spectral densities for standard kernels, but our theory does not provide a rule for the choice of $m$ because first order asymptotic properties of scores of $Q_{T}^{m d s}(\vartheta)$ do not depend on $m$ once Assumption 3.3.2 holds.

### 3.9.3. Proofs of Section 3

Proof of Theorem 4. It follows directly from the CLT for the scores of the loss function $Q_{T}^{m d s}$ in Velasco (2022).

Proof of Theorem 5. From the first-order condition of the restricted AP estimation,

$$
\begin{aligned}
0 & =\frac{\partial}{\partial \vartheta_{2}^{*}} Q_{T}^{m d s}\left(\tilde{\vartheta}_{T}^{*}\right)=\frac{\partial}{\partial \phi} Q_{T}^{m d s}\left(\tilde{\vartheta}_{T}^{*}\right) \\
& =\frac{2}{\pi} \sum_{j=1}^{T-1} k^{2}\left(\frac{j}{m}\right)\left(1-\frac{|j|}{T}\right) \int 2 \operatorname{Re}\left\{\hat{\sigma}_{\tilde{\vartheta}_{T}^{*}, j}^{(1,0) \dagger}(0, v) \overline{\dot{\sigma}_{\phi_{0}, j}^{(1,0)}(0, v)}\right\} d W(v)+o_{p}\left(T^{-1 / 2}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\sigma}_{\tilde{\vartheta}_{T}^{*}, j}^{(1,0) \dagger}(0, v) & =\hat{\sigma}_{\vartheta_{0}^{*}, j}^{(1,0) \dagger}(0, v)+\left(\tilde{\vartheta}_{T}^{*}-\vartheta_{0}^{*}\right)^{\prime} \dot{\sigma}_{\vartheta_{0}^{*}, j}^{(1,0)}(0, v)+o_{p}\left(T^{-1 / 2}\right) \\
& =\hat{\sigma}_{\vartheta_{0}^{*}, j}^{(1,0) \dagger}(0, v)+\left(\tilde{\phi}_{T}-\phi_{0}\right) \dot{\sigma}_{\phi_{0}, j}^{(1,0)}(0, v)+o_{p}\left(T^{-1 / 2}\right),
\end{aligned}
$$

we obtain up to $o_{p}\left(T^{-1 / 2}\right)$ terms,

$$
\begin{aligned}
0 & =\frac{\partial}{\partial \vartheta_{2}^{*}} Q_{T}^{m d s}\left(\vartheta_{0}^{*}\right)+\frac{2}{\pi} \sum_{j=1}^{T-1} k^{2}\left(\frac{j}{m}\right)\left(1-\frac{|j|}{T}\right) \int 2 \operatorname{Re}\left\{\dot{\sigma}_{\phi_{0}, j}^{(1,0)}(0, v) \overline{\dot{\sigma}_{\phi_{0}, j}^{(1,0)}(0, v)}\right\} d W(v)\left(\tilde{\phi}_{T}-\phi_{0}\right) \\
& =\frac{\partial}{\partial \vartheta_{2}^{*}} Q_{T}^{m d s}\left(\vartheta_{0}^{*}\right)+H_{22,0}^{*}\left(\tilde{\phi}_{T}-\phi_{0}\right)+o_{p}\left(T^{-1 / 2}\right) \\
& =A_{2} \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right)+A_{2} H_{0}^{m d s} A_{2}^{\prime}\left(\tilde{\phi}_{T}-\phi_{0}\right)+o_{p}\left(T^{-1 / 2}\right)
\end{aligned}
$$

where

$$
H_{0}^{m d s}:=-\sum_{j=1}^{\infty} \int\left\{\zeta_{j}^{0}(-v)+\beta_{j}^{0}(-v)\right\}\left\{\zeta_{j}^{0}(v)+\beta_{j}^{0}(v)\right\}^{\prime} d W(v)
$$

and for $A_{2}=\left(\begin{array}{ll}1 & 1\end{array}\right)$,

$$
\tilde{\phi}_{T}-\phi_{0}=-\left\{A_{2} H_{0}^{m d s} A_{2}^{\prime}\right\}^{-1} A_{2} \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right)+o_{p}\left(T^{-1 / 2}\right)
$$

Similarly, for $A_{1}=(-1,0) \otimes I_{p}$,

$$
\begin{aligned}
& \frac{\partial}{\partial \vartheta_{1}^{*}} Q_{T}^{m d s}\left(\tilde{\vartheta}_{T}^{*}\right)=\frac{\partial}{\partial \gamma} Q_{T}^{m d s}\left(\tilde{\vartheta}_{T}^{*}\right) \\
= & \frac{\partial}{\partial \vartheta_{1}^{*}} Q_{T}^{m d s}\left(\vartheta_{0}^{*}\right)+\frac{2}{\pi} \sum_{j=1}^{T-1} k^{2}\left(\frac{j}{m}\right)\left(1-\frac{|j|}{T}\right) \int 2 \operatorname{Re}\left\{\dot{\sigma}_{\gamma_{0}, j}^{(1,0)}(0, v) \overline{\dot{\sigma}_{\phi_{0}, j}^{(1,0)}(0, v)}\right\} d W(v)\left(\tilde{\phi}_{T}-\phi_{0}\right)+o_{p}\left(T^{-1 / 2}\right) \\
= & \frac{\partial}{\partial \vartheta_{1}^{*}} Q_{T}^{m d s}\left(\vartheta_{0}^{*}\right)+H_{12,0}^{*}\left(\tilde{\phi}_{T}-\phi_{0}\right)+o_{p}\left(T^{-1 / 2}\right) \\
= & A_{1} \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right)+A_{1} H_{0}^{m d s} A_{2}^{\prime}\left(\tilde{\phi}_{T}-\phi_{0}\right)+o_{p}\left(T^{-1 / 2}\right) \\
= & A_{1} \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right)-A_{1} H_{0}^{m d s} A_{2}^{\prime}\left\{A_{2} H_{0}^{m d s} A_{2}^{\prime}\right\}^{-1} A_{2} \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right)+o_{p}\left(T^{-1 / 2}\right)
\end{aligned}
$$

or, up to $o_{p}\left(T^{-1 / 2}\right)$ terms,

$$
\begin{aligned}
\frac{\partial}{\partial \gamma} Q_{T}^{m d s}\left(\tilde{\vartheta}_{T}^{*}\right) & =\left[A_{1}-A_{1} H_{0}^{m d s} A_{2}^{\prime}\left\{A_{2} H_{0}^{m d s} A_{2}^{\prime}\right\}^{-1} A_{2}\right] \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right) \\
& =\left[I_{p} \quad \vdots-A_{1} H_{0}^{m d s} A_{2}^{\prime}\left\{A_{2} H_{0}^{m d s} A_{2}^{\prime}\right\}^{-1}\right]\left[\begin{array}{c}
A_{1} \\
A_{2}
\end{array}\right] \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right) \\
& =\Xi_{1} A \frac{\partial}{\partial \vartheta} Q_{T}^{m d s}\left(\vartheta_{0}\right)
\end{aligned}
$$

where $\Xi_{1}:=\left[\begin{array}{lll}I_{p} & \vdots & -A_{1} H_{0}^{m d s} A_{2}^{\prime}\left\{A_{2} H_{0}^{m d s} A_{2}^{\prime}\right\}^{-1}\end{array}\right]$, so that

$$
T^{1 / 2} \frac{\partial}{\partial \gamma} Q_{T}^{m d s}\left(\tilde{\vartheta}_{T}^{*}\right) \rightarrow_{d} N\left(0, S_{0}\right)
$$

where

$$
S_{0}:=\Xi_{1} A V_{0}^{m d s} A^{\prime} \Xi_{1}^{\prime}
$$

and $V_{0}^{m d s}:=\mathbb{V}\left[\varepsilon_{t}\left(R_{t-1}^{(0)}+S_{t-1}^{(0)}\right)\right]$ because $\varepsilon_{t}\left(R_{t-1}^{(0)}+S_{t-1}^{(0)}\right)$ is a $m d s$,

$$
\begin{aligned}
V_{0}^{m d s}=\sum_{j=1}^{\infty} \sum_{\ell=1}^{\infty} \iint & \mathbb{E}\left[\varepsilon_{t}^{2} z_{t-j}^{0}(v) z_{t-\ell}^{0}(-u)\right] \\
& \times\left\{\zeta_{j}^{0}(-v)+\beta_{j}^{0}(-v)\right\}\left\{\zeta_{\ell}^{0}(u)+\beta_{\ell}^{0}(u)\right\}^{\prime} d W(v) d W(u),
\end{aligned}
$$

where

$$
\begin{aligned}
R_{t-1}^{(0)} & =i \sum_{j=1}^{\infty} \int z_{t-j}^{0}(v) \zeta_{j}^{0}(-v) d W(v), \quad \zeta_{j}^{0}(v):=-\sum_{n=j}^{\infty} \delta_{n}\left(\theta_{0}\right) \varphi_{j-n}^{(1,0)}(0, v) \\
S_{t-1}^{(0)} & =i \sum_{j=1}^{\infty} \int z_{t-j}^{0}(v) \beta_{j}^{0}(-v) d W(v), \quad \beta_{j}^{0}(v)=-\delta_{-j}\left(\vartheta_{0}\right) v \varphi_{j}^{(2,0)}(0, v) .
\end{aligned}
$$

For calculations we can use Gaussian $W$, for which we obtain

$$
\begin{aligned}
& i \int \hat{\varepsilon}_{t-j}(v) \hat{\beta}_{j}(-v) d W(v) \\
= & \frac{\delta_{-j}\left(\tilde{\vartheta}_{T}\right)}{T-j} \sum_{r=1+j}^{T} \hat{\varepsilon}_{r}^{2}\left[\begin{array}{c}
\left(\hat{\varepsilon}_{t-j}-\hat{\varepsilon}_{r-j}\right) \exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{t-j}-\hat{\varepsilon}_{r-j}\right)^{2}\right\} \\
-\frac{1}{T-j} \sum_{t=1+j}^{T}\left(\hat{\varepsilon}_{t-j}-\hat{\varepsilon}_{r-j}\right) \exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{t-j}-\hat{\varepsilon}_{r-j}\right)^{2}\right\}
\end{array}\right]
\end{aligned}
$$

and

$$
i \int \hat{z}_{t-j}(v) \hat{\zeta}_{j, m}(-v) d W(v)=\sum_{n=|j|}^{T+j-1} k((n-j) / m) \delta_{n}(\hat{\theta}) \frac{1}{i} \int \hat{z}_{t-j}(v) \hat{\varphi}_{\hat{\theta},|j|-n}^{(1,0)}(0,-v) d W(v)
$$

where, $j=1,2, \ldots, t-1, n=j, j+1, \ldots, T-1$,

$$
\begin{aligned}
& \frac{1}{i} \int \hat{z}_{t-j}(v) \hat{\varphi}_{\hat{\theta},|j|-n}^{(1,0)}(0,-v) d W(v) \\
= & \frac{1}{T-n+j} \sum_{r=1}^{T-n+j} \hat{\varepsilon}_{r}\left[\exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{t-j}-\hat{\varepsilon}_{r+n-j}\right)^{2}\right\}-\frac{1}{T-j} \sum_{t=1+j}^{T} \exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{t-j}-\hat{\varepsilon}_{r+n-j}\right)^{2}\right\}\right] .
\end{aligned}
$$

Then, to compute $\hat{H}_{0, m}^{m d s}$ we need, $j=1,2, \ldots, T-1$,

$$
\begin{aligned}
\int \hat{\beta}_{j}(-v) \hat{\beta}_{j}(v)^{\prime} d W(v)= & -\delta_{-j}\left(\tilde{\vartheta}_{T}\right) \delta_{-j}\left(\tilde{\vartheta}_{T}\right)^{\prime} \int v^{2}\left|\hat{\varphi}_{\theta_{T}, j}^{(2,0)}(0, v)\right|^{2} d W(v), \\
\int \hat{\zeta}_{j, m}(-v) \hat{\zeta}_{j, m}(v)^{\prime} d W(v)= & \sum_{n=j}^{T+j-1} \sum_{m=j}^{T+j-1} k\left(\frac{n-j}{m}\right) k\left(\frac{m-j}{m}\right) \delta_{n}\left(\tilde{\vartheta}_{T}\right) \delta_{m}\left(\tilde{\vartheta}_{T}\right)^{\prime} \\
& \times \int \hat{\varphi}_{\theta_{T}, j-n}^{(1,0)}(0,-v) \hat{\varphi}_{\theta_{T}, j-m}^{(1,0)}(0, v) d W(v)
\end{aligned}
$$

and $\int \hat{\zeta}_{j, m}(-v) \hat{\beta}_{j}(v)^{\prime} d W(v)$ is

$$
\sum_{n=j}^{T+j-1} k\left(\frac{n-j}{m}\right) \delta_{n}\left(\tilde{\vartheta}_{T}\right) \delta_{-j}\left(\tilde{\vartheta}_{T}\right)^{\prime} \int v \hat{\varphi}_{\theta_{T}, j-n}^{(1,0)}(0,-v) \hat{\varphi}_{\theta_{T}, j}^{(2,0)}(0, v) d W(v),
$$

where

$$
\int v^{2}\left|\hat{\varphi}_{\theta_{T}, j}^{(2,0)}(0, v)\right|^{2} d W(v)=\frac{1}{(T-j)^{2}} \sum_{t=1+j}^{T} \sum_{r=1+j}^{T} \hat{\varepsilon}_{t}^{2} \hat{\varepsilon}_{r}^{2}\left(1-\left(\hat{\varepsilon}_{t-j}-\hat{\varepsilon}_{r-j}\right)^{2}\right) \exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{t-j}-\hat{\varepsilon}_{r-j}\right)^{2}\right\}
$$

and for $n=j, j+1, \ldots$,

$$
\begin{aligned}
& \int \hat{\varphi}_{\theta_{T},|j|-n}^{(1,0)}(0,-v) \hat{\varphi}_{\theta_{T},|j|-m}^{(1,0)}(0, v) d W(v) \\
& =-\frac{1}{(T-n-j)(T-m-j)} \sum_{t=1}^{T-n+j} \sum_{r=1}^{T-m+j} \hat{\varepsilon}_{t} \hat{\varepsilon}_{r} \exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{t+n-j}-\hat{\varepsilon}_{r+m-j}\right)^{2}\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
& \int \hat{\varphi}_{\theta_{T}, j \mid-n}^{(1,0)}(0,-v) v \hat{\varphi}_{\theta_{T}, j}^{(2,0)}(0, v) d W(v) \\
& =-\frac{1}{T-n+j} \frac{1}{T-j} \sum_{t=1}^{T-n+j} \sum_{r=1+j}^{T} \hat{\varepsilon}_{t} \hat{\varepsilon}_{r}^{2}\left(\hat{\varepsilon}_{t+n-j}-\hat{\varepsilon}_{r-j}\right) \exp \left\{-\frac{1}{2}\left(\hat{\varepsilon}_{t+n-j}-\hat{\varepsilon}_{r-j}\right)^{2}\right\} .
\end{aligned}
$$

### 3.9.4. Tests based on iid dependence measures

For the LM tests of the IID null hypotheses $H_{I I D}^{(A P)}$ and $H_{I I D}$ the objective function based on the generalized spectral density function is now

$$
Q_{T}^{i i d}(\vartheta):=\frac{2}{\pi} \sum_{j=1}^{T-1} k^{2}\left(\frac{j}{m}\right)\left(1-\frac{|j|}{T}\right) \int\left|\hat{\sigma}_{\vartheta, j}(u, v)\right|^{2} d W(u, v)
$$

where $W(u, v)=W(u) W(v), \vartheta=(\phi, \theta)^{\prime}$ for $\operatorname{ARMA}(1,1)$ model and $\vartheta=\phi$ for AP model. Then

$$
\frac{\partial}{\partial \vartheta} Q_{T}^{i i d}(\vartheta)=\frac{2}{\pi} \sum_{j=1}^{T-1} k^{2}\left(\frac{j}{m}\right)\left(1-\frac{|j|}{T}\right) \int 2 \operatorname{Re}\left\{\hat{\sigma}_{\vartheta, j}(u, v) \frac{\partial}{\partial \vartheta} \overline{\hat{\sigma}_{\vartheta, j}(u, v)}\right\} d W(u, v)
$$

so that imposing the null $\vartheta=\vartheta_{0}$,

$$
\frac{\partial}{\partial \vartheta} Q_{T}^{i i d}\left(\vartheta_{0}\right)=\frac{1}{T} \sum_{t=2}^{T}\left(e_{t}^{0} X_{t-1}^{(0)}+x_{t}^{0} E_{t-1}^{(0)}\right)+o_{p}\left(T^{-1 / 2}\right)
$$

where for $a=0,1$ and $z_{t}^{0}=z_{t}\left(\vartheta_{0} ; u\right)=e^{i u \varepsilon_{t}}-\mathbb{E}\left[e^{i u \varepsilon_{t}}\right]=e^{i u \varepsilon_{t}}-\varphi(u)$,

$$
\begin{aligned}
e_{t}^{0}=e_{t}\left(\theta_{0}\right) & :=\frac{1}{i} \int z_{t}\left(\theta_{0} ; u\right) u \varphi(-u) d W(u) \\
x_{t}^{0}=x_{t}\left(\theta_{0}\right) & :=\frac{1}{i} \int z_{t}\left(\theta_{0} ; u\right) \varphi^{(1)}(-u) d W(u),
\end{aligned}
$$

and $X_{t-1}^{(0)}$ and $E_{t-1}^{(0)}$ are given by

$$
\begin{aligned}
X_{t-1}^{(0)}=X_{t-1}^{(0)}\left(\vartheta_{0}\right) & :=\frac{4}{\pi} \sum_{j=1}^{t-1} \delta_{j}\left(\vartheta_{0}\right) x_{t-j}^{0} \\
E_{t-1}^{(0)}=E_{t-1}^{(0)}\left(\vartheta_{0}\right) & :=\frac{4}{\pi} \sum_{j=1}^{t-1} \delta_{-j}\left(\vartheta_{0}\right) e_{t-j}^{0},
\end{aligned}
$$

for the same coefficients $\delta_{j}\left(\vartheta_{0}\right)$ from (3.9.3), so the score approximation for the ARMA model becomes

$$
\frac{1}{T} \sum_{t=2}^{T}\left(e_{t}^{0} X_{t-1}^{(0)}+x_{t}^{0} E_{t-1}^{(0)}\right)=\frac{1}{T} \frac{4}{\pi} \sum_{t=2}^{T} \sum_{j=1}^{t-1}\left(e_{t}^{0} \delta_{j}\left(\vartheta_{0}\right) x_{t-j}^{0}+x_{t}^{0} \delta_{-j}\left(\vartheta_{0}\right) e_{t-j}^{0}\right)=\frac{1}{T} \frac{4}{\pi} \sum_{t=2}^{T}\binom{x_{t}^{0} e_{t-1}^{0}}{-e_{t}^{0} x_{t-1}^{0}}
$$

which has full rank limit variance.
For the AP process we obtain in the same way the score approximation by multiplying by $(1,1)$,

$$
\frac{1}{T} \sum_{t=2}^{T}(1,1)\left(e_{t}^{0} X_{t-1}^{(0)}+x_{t}^{0} E_{t-1}^{(0)}\right)=\frac{1}{T} \frac{4}{\pi} \sum_{t=2}^{T} \sum_{j=1}^{t-1}\left(x_{t}^{0} e_{t-1}^{0}-e_{t}^{0} x_{t-1}^{0}\right)
$$

and inference could be conducted in the usual way as before since there are no nuisance parameters under either null hypothesis using

$$
L M_{T}=T \frac{\partial}{\partial \vartheta} Q_{T}^{i i d}\left(\vartheta_{0}\right)^{\prime} \widehat{A V a r}\left(T^{1 / 2} \frac{\partial}{\partial \vartheta} Q_{T}^{i i d}\left(\vartheta_{0}\right)\right)^{-1} \frac{\partial}{\partial \vartheta} Q_{T}^{i i d}\left(\vartheta_{0}\right)
$$

for testing both hypothesis approximating the null distributions by $\chi_{2}^{2}$ and $\chi_{1}^{2}$ variables by estimating for the ARMA model

$$
\operatorname{AVar}\left(T^{1 / 2} \frac{\partial}{\partial \vartheta} Q_{T}^{i i d}\left(\vartheta_{0}\right)\right)=\mathbb{E}\left(\begin{array}{cc}
\left(x_{t}^{0} e_{t-1}^{0}\right)^{2} & -x_{t}^{0} e_{t}^{0} x_{t-1}^{0} e_{t-1}^{0} \\
-x_{t}^{0} e_{t}^{0} x_{t-1}^{0} e_{t-1}^{0} & \left(e_{t}^{0} x_{t-1}^{0}\right)^{2}
\end{array}\right)=\left(\begin{array}{cc}
\sigma_{e}^{2} \sigma_{x}^{2} & -\sigma_{x e}^{2} \\
-\sigma_{x e}^{2} & \sigma_{e}^{2} \sigma_{x}^{2}
\end{array}\right)
$$

by the iid property of $\varepsilon_{t}$, or for the AP model

$$
A \operatorname{Var}\left(T^{1 / 2} \frac{\partial}{\partial \vartheta} Q_{T}^{i i d}\left(\vartheta_{0}\right)\right)=2\left(\sigma_{e}^{2} \sigma_{x}^{2}-\sigma_{x e}^{2}\right)
$$

where $\left(\sigma_{e}^{2}, \sigma_{x}^{2}, \sigma_{x e}^{2}\right)$ can be estimated by sample moments using

$$
\begin{aligned}
& \hat{e}_{t}:=e_{t}(\hat{\theta})=\frac{1}{i} \int \hat{z}_{t}(u) u \hat{\varphi}_{\hat{\theta}}(-u) d W(u) \\
& \hat{x}_{t}:=x_{t}(\hat{\theta})=\frac{1}{i} \int \hat{z}_{t}(u) \hat{\varphi}_{\hat{\theta}}^{(1)}(-u) d W(u),
\end{aligned}
$$

in

$$
\widehat{\operatorname{AVar}}\left(T^{1 / 2} \frac{\partial}{\partial \vartheta} Q_{T}^{i i d}\left(\vartheta_{0}\right)\right)=\left(\begin{array}{cc}
\hat{\sigma}_{e}^{2} \hat{\sigma}_{x}^{2} & -\hat{\sigma}_{x e}^{2} \\
-\hat{\sigma}_{x e}^{2} & \hat{\sigma}_{e}^{2} \hat{\sigma}_{x}^{2}
\end{array}\right), \quad 2\left(\hat{\sigma}_{e}^{2} \hat{\sigma}_{x}^{2}-\hat{\sigma}_{x e}^{2}\right) .
$$

The LM testing of the AP model against ARMA under iid, $H_{A P / i d d}$ : $\phi_{0}=\theta_{0}$ in model (3.3.1) would follow the same lines as under $m d s$, just using now

$$
\begin{aligned}
& \Xi_{1}:=\left[I_{p} \vdots-A_{1} H_{0}^{i i d} A_{2}^{\prime}\left\{A_{2} H_{0}^{i i d} A_{2}^{\prime}\right\}^{-1}\right] \text { with } \\
& \qquad H_{0}^{i i d}:=\rho_{1} \rho_{2}\left(\boldsymbol{\Sigma}_{0}+\boldsymbol{\Sigma}_{0}^{-}\right)-\rho_{0}^{2}\left(\boldsymbol{\Sigma}_{0}^{\mp}+\boldsymbol{\Sigma}_{0}^{\mp \prime}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\Sigma}_{0} & :=\sum_{j=1}^{\infty} \delta_{j}\left(\theta_{0}\right) \delta_{j}^{\prime}\left(\theta_{0}\right) \\
\boldsymbol{\Sigma}_{0}^{-} & :=\sum_{j=1}^{\infty} \delta_{-j}\left(\theta_{0}\right) \delta_{-j}^{\prime}\left(\theta_{0}\right) \\
\boldsymbol{\Sigma}_{0}^{\mp} & :=\sum_{j=1}^{\infty} \delta_{-j}\left(\theta_{0}\right) \delta_{j}^{\prime}\left(\theta_{0}\right)
\end{aligned}
$$

$\rho_{0}:=-\int \varphi^{(1)}(u) u \varphi(-u) d W(u), \rho_{1}:=\int\left|\varphi^{(1)}(u)\right|^{2} d W(u)$ and $\rho_{2}:=\int u^{2}|\varphi(u)|^{2} d W(u)$, which could be estimated easily from residuals, and

$$
V_{0}^{i i d}:=A \operatorname{Var}\left(T^{1 / 2} \frac{\partial}{\partial \vartheta} Q_{T}^{i i d}\left(\vartheta_{0}\right)\right)=\sigma_{e}^{2} \sigma_{x}^{2}\left(\boldsymbol{\Sigma}_{2 a}+\boldsymbol{\Sigma}_{2 a}^{-}\right)+\sigma_{x e}^{2}\left(\boldsymbol{\Sigma}_{2 a}^{\mp}+\boldsymbol{\Sigma}_{2 a}^{\mp \prime}\right) .
$$

For instance,

$$
\begin{aligned}
& \hat{e}_{t}=\frac{1}{T} \sum_{r=1}^{T}\left\{e^{-\frac{1}{2}\left(\hat{\varepsilon}_{t}-\hat{\varepsilon}_{r}\right)^{2}}\right\}\left(\hat{\varepsilon}_{t}-\hat{\varepsilon}_{r}\right)-\frac{1}{T^{2}} \sum_{r=1}^{T} \sum_{s=1}^{T}\left\{e^{-\frac{1}{2}\left(\hat{\varepsilon}_{s}-\hat{\varepsilon}_{r}\right)^{2}}\right\}\left(\hat{\varepsilon}_{s}-\hat{\varepsilon}_{r}\right) \\
& \hat{x}_{t}=\frac{1}{T} \sum_{r=1}^{T} \hat{\varepsilon}_{r}\left\{e^{-\frac{1}{2}\left(\hat{\varepsilon}_{t}-\hat{\varepsilon}_{r}\right)^{2}}-\frac{1}{T} \sum_{s=1}^{T} e^{-\frac{1}{2}\left(\hat{\varepsilon}_{s}-\hat{\varepsilon}_{r}\right)^{2}}\right\}
\end{aligned}
$$

while

$$
\begin{aligned}
& \hat{\rho}_{0}:=-\int \hat{\varphi}_{\hat{\theta}}^{(1)}(u) u \hat{\varphi}_{\hat{\theta}}(-u) d W(u)=-\frac{1}{T^{2}} \sum_{r=1}^{T} \sum_{s=1}^{T}\left\{e^{-\frac{1}{2}\left(\hat{e}_{s}-\hat{\varepsilon}_{r}\right)^{2}}\right\} \hat{\varepsilon}_{r}\left(\hat{\varepsilon}_{r}-\hat{\varepsilon}_{s}\right) \\
& \hat{\rho}_{1}:=\int\left|\hat{\varphi}_{\hat{\theta}}^{(1)}(u)\right|^{2} d W(u)=\frac{1}{T^{2}} \sum_{r=1}^{T} \sum_{s=1}^{T}\left\{e^{-\frac{1}{2}\left(\hat{\varepsilon}_{s}-\hat{\varepsilon}_{r}\right)^{2}}\right\} \hat{\varepsilon}_{s} \hat{\varepsilon}_{r} \\
& \hat{\rho}_{2}:=\int u^{2}\left|\hat{\varphi}_{\hat{\theta}}(u)\right|^{2} d W(u)=\frac{1}{T^{2}} \sum_{r=1}^{T} \sum_{s=1}^{T}\left\{e^{-\frac{1}{2}\left(\hat{\varepsilon}_{s}-\hat{\varepsilon}_{r}\right)^{2}}\right\}\left\{1-\left(\hat{\varepsilon}_{s}-\hat{\varepsilon}_{r}\right)^{2}\right\} .
\end{aligned}
$$

## BIBLIOGRAPHY

Alessi, L., Barigozzi, M. and Capasso, M. (2011) Non-fundamentalness in structural econometric models: A review. International Statistical Review, 79, 16-47.

Andrews, B., Davis, R. A. and Breidt, F. J. (2006) Maximum likelihood estimation for all-pass time series models. Journal of Multivariate Analysis, 97, 1638-1659.

Andrews, B., Davis, R. A., Breidt, F. J. et al. (2007) Rank-based estimation for all-pass time series models. The Annals of Statistics, 35, 844-869.

Andrews, D. W. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica: Journal of the Econometric Society, 817-858.

Andrews, D. W. and Ploberger, W. (1996) Testing for serial correlation against an arma (1, 1) process. Journal of the American Statistical Association, 91, 13311342.

Angrist, J., Chernozhukov, V. and Fernández-Val, I. (2006) Quantile regression under misspecification, with an application to the us wage structure. Econometrica, 74, 539-563.

Baruník, J. and Kley, T. (2019) Quantile coherency: A general measure for dependence between cyclical economic variables. The Econometrics Journal, 22, 131-152.

Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961) Distribution free tests of independence based on the sample distribution function. Sandia Corporation.

Boldin, M. V., Simonova, G. I. and Ti_urin, I. N. (1997) Sign-based methods in linear statistical models, vol. 162. American Mathematical Soc.

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, 1509-1526.

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

Breidt, F. J., Davis, R. A., Trindade, A. A. et al. (2001) Least absolute deviation estimation for all-pass time series models. The Annals of Statistics, 29, 919-946.

Breusch, T. S. (1978) Testing for autocorrelation in dynamic linear models. Australian economic papers, 17, 334-355.

Brockwell, P. J. and Davis, R. A. (2009) Time series: theory and methods. Springer Science \& Business Media.

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

Cavaliere, G., Nielsen, H. B. and Rahbek, A. (2020) Bootstrapping noncausal autoregressions: with applications to explosive bubble modeling. Journal of Business $\mathcal{E}$ Economic Statistics, 38, 55-67.

Chen, B., Choi, J. and Escanciano, J. C. (2017) Testing for fundamental vector moving average representations. Quantitative Economics, 8, 149-180.

Davies, R. B. (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 74, 33-43.

Delgado, M. A. (1996) Testing serial independence using the sample distribution function. Journal of Time Series Analysis, 17, 271-285.

Doukhan, P. and Louhichi, S. (1999) A new weak dependence condition and applications to moment inequalities. Stochastic processes and their applications, 84, 313-342.

Du, Z. and Escanciano, J. C. (2015) A nonparametric distribution-free test for serial independence of errors. Econometric Reviews, 34, 1011-1034.

Escanciano, J. C. (2006) Goodness-of-fit tests for linear and nonlinear time series models. Journal of the American Statistical Association, 101, 531-541.

Escanciano, J. C. and Goh, S.-C. (2012) Conditional density estimation in linear quantile regression. (unpublished).

- (2014) Specification analysis of linear quantile models. Journal of Econometrics, 178, 495-507.

Escanciano, J. C. and Velasco, C. (2006a) Generalized spectral tests for the martingale difference hypothesis. Journal of Econometrics, 134, 151-185.

- (2006b) Testing the martingale difference hypothesis using integrated regression functions. Computational Statistics \& Data Analysis, 51, 2278-2294.
- (2010) Specification tests of parametric dynamic conditional quantiles. Journal of Econometrics, 159, 209-221.

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

Gaeta, M., Briolle, F. and Esparcieux, P. (1997) Blind separation of sources applied to convolutive mixtures in shallow water. In Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics, 340-343. IEEE.

Godfrey, L. G. (1978) Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. Econometrica: Journal of the Econometric Society, 1293-1301.

Gospodinov, N. and Ng, S. (2015) Minimum distance estimation of possibly noninvertible moving average models. Journal of Business $\mathfrak{E}^{6}$ Economic Statistics, 33, 403-417.

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, 737-756.

Hagemann, A. (2011) Robust spectral analysis. arXiv preprint arXiv:1111.1965.
Hall, P. and Sheather, S. J. (1988) On the distribution of a studentized quantile. Journal of the Royal Statistical Society: Series B (Methodological), 50, 381-391.

Hansen, L. P. and Sargent, T. J. (1991) Two difficulties in interpreting vector autoregressions. Rational expectations econometrics, 1, 77-119.

Hecq, A., Issler, J. V. and Telg, S. (2020) Mixed causal-noncausal autoregressions with exogenous regressors. Journal of Applied Econometrics, 35, 328-343.

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, 307-331.

- (2017) Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The marx package. Available at SSRN $301579 \%$.

Hecq, A. and Sun, L. (2021) Selecting between causal and noncausal models with quantile autoregressions. Studies in Nonlinear Dynamics \& Econometrics, 25, 393-416.

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

Hinich, M. J. (1982) Testing for gaussianity and linearity of a stationary time series. Journal of time series analysis, 3, 169-176.

Hoeffding, W. (1948) A non-parametric test of independence. The annals of mathematical statistics, 546-557.

Hong, Y. (1998) Testing for pairwise serial independence via the empirical distribution function. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 60, 429-453.

- (1999) Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. Journal of the American Statistical Association, 94, 1201-1220.
- (2000) Generalized spectral tests for serial dependence. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62, 557-574.

Hong, Y. and Lee, Y.-J. (2005) Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. The Review of Economic Studies, 72, 499-541.

Huang, J. and Pawitan, Y. (2000) Quasi-likelihood estimation of non-invertible moving average processes. Scandinavian Journal of Statistics, 27, 689-702.

Jin, W. (2022) Estimation of time series models using the empirical distribution of residuals. Preprint, Department of Economics, Universidad Carlos III de Madrid.

- (2023) Quantile autoregression-based non-causality testing. arXiv preprint arXiv:2301.0293\%.

Kasa, K., Walker, T. B. and Whiteman, C. H. (2006) Asset prices in a time series model with perpetually disparately informed, competitive traders. CAEPR Working Paper, Economics Department, Indiana University Bloomington.

Kim, J. and Pollard, D. (1990) Cube root asymptotics. The Annals of Statistics, 191-219.

Kley, T., Volgushev, S., Dette, H., Hallin, M. et al. (2016) Quantile spectral processes: Asymptotic analysis and inference. Bernoulli, 22, 1770-1807.

Koenker, R. and Xiao, Z. (2002) Inference on the quantile regression process. Econometrica, 70, 1583-1612.

- (2006) Quantile autoregression. Journal of the American statistical association, 101, 980-990.

Lanne, M. and Luoto, J. (2013) Autoregression-based estimation of the new keynesian phillips curve. Journal of Economic Dynamics and Control, 37, 561-570.

Lanne, M., Luoto, J. and Saikkonen, P. (2012) Optimal forecasting of noncausal autoregressive time series. International Journal of Forecasting, 28, 623-631.

Lanne, M., Meitz, M., Saikkonen, P. et al. (2013) Testing for linear and nonlinear predictability of stock returns. Journal of Financial Econometrics, 11, 682-705.

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

Lee, J. and Rao, S. S. (2011) The quantile spectral density and comparison based tests for nonlinear time series. arXiv preprint arXiv:1112.2759.

Leeper, E. M., Walker, T. B. and Yang, S.-C. S. (2013) Fiscal foresight and information flows. Econometrica, 81, 1115-1145.

Lii, K.-S. and Rosenblatt, M. (1992) An approximate maximum likelihood estimation for non-gaussian non-minimum phase moving average processes. Journal of Multivariate Analysis, 43, 272-299.

- (1996) Maximum likelihood estimation for nongaussian nonminimum phase arma sequences. Statistica Sinica, 1-22.

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

- (1994) Var analysis, nonfundamental representations, blaschke matrices. Journal of Econometrics, 63, 307-325.

Lobato, I. N., Nankervis, J. C. and Savin, N. (2002) Testing for zero autocorrelation in the presence of statistical dependence. Econometric Theory, 18, 730-743.

Lof, M. and Nyberg, H. (2017) Noncausality and the commodity currency hypothesis. Energy Economics, 65, 424-433.

Nankervis, J. C. and Savin, N. E. (2010) Testing for serial correlation: Generalized andrews-ploberger tests. Journal of Business $\mathfrak{G}$ Economic Statistics, 28, 246-255.

Newey, W. K. (1991) Uniform convergence in probability and stochastic equicontinuity. Econometrica: Journal of the Econometric Society, 1161-1167.

Newey, W. K. and McFadden, D. (1994) Large sample estimation and hypothesis testing. Handbook of econometrics, 4, 2111-2245.

Portnoy, S. and Koenker, R. (1989) Adaptive $l$-estimation for linear models. The Annals of Statistics, 17, 362-381.

Ramsey, J. B. and Montenegro, A. (1992) Identification and estimation of noninvertible non-gaussian ma (q) processes. Journal of econometrics, 54, 301-320.

Rao, T. S. and Gabr, M. (1980) A test for linearity of stationary time series. Journal of time series analysis, 1, 145-158.

Rosenblatt, M. (2000) Gaussian and non-Gaussian linear time series and random fields. Springer Science \& Business Media.

Shao, X. (2011) Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. Econometric Theory, 27, 312-343.

Sidorov, D. (2010) On mixing conditions for sequences of moving averages. Theory of Probability $\mathcal{G}$ Its Applications, 54, 339-347.

Skaug, H. J. and Tjøstheim, D. (1993) A nonparametric test of serial independence based on the empirical distribution function. Biometrika, 80, 591-602.

Stinchcombe, M. B. and White, H. (1998) Consistent specification testing with nuisance parameters present only under the alternative. Econometric theory, 14, 295-325.

Taylor, S. J. (2008) Modelling financial time series. world scientific.
Tekalp, A., Kaufman, H. and Woods, J. (1986) Identification of image and blur parameters for the restoration of noncausal blurs. IEEE Transactions on Acoustics, Speech, and Signal Processing, 34, 963-972.

Velasco, C. (2022) Estimation of time series models using residuals dependence measures. The Annals of Statistics, 50, 3039-3063.

Velasco, C. and Lobato, I. N. (2018) Frequency domain minimum distance inference for possibly noninvertible and noncausal arma models. The Annals of Statistics, 46, 555-579.

Yoshihara, K.-I. (1978) Moment inequalities for mixing sequences. Kodai Mathematical Journal, 1, 316-328.


[^0]:    ${ }^{1}$ The location model is any model of the form $Y=\mu(X)+\sigma \cdot u$, where $u$ is independent of $X$. Throughout this paper, we refer to $Y_{t}=\mu\left(Y_{t-1}, Y_{t-2}, \ldots\right)+u_{t}$, where $\mu(\cdot)$ is a linear functional form, and $u_{t}$ is a sequence of iid innovations, when it comes to the location model.

[^1]:    ${ }^{2}$ Mixed causal and non-causal autoregressive processes have been previously categorized into general non-causal autoregressive processes in the past literature. The notation $\operatorname{MAR}(r, s)$ was

[^2]:    first adopted by Hecq et al. (2017) to describe this class of processes.

[^3]:    ${ }^{3}$ A log-normal distribution is a heavy-tailed continuous distribution defined on the positive domain, whose density function is characterized by the location parameter $\mu$ and scale parameter $\sigma$. The mean and variance are represented by $\exp \left(\mu+\sigma^{2} / 2\right)$ and $\left(\exp \left(\sigma^{2}\right)-1\right) \exp \left(2 \mu+\sigma^{2}\right)$, respectively. In this example, we employ centered log-normal distribution to be compatible with our setup of mean zero.

[^4]:    ${ }^{4}$ Actually, $\tilde{u}_{t}$ is an all-pass time series model, where all roots of the autoregressive polynomial are reciprocals of roots in the moving average polynomial and vice versa.
    ${ }^{5} \kappa_{4}(X)$ is the fourth cumulant of the random variable $X$, defined by $\kappa_{4}(X)=$ $\mathbb{E}\left((X-\mathbb{E}(X))^{4}\right)-3 \mathbb{E}^{2}\left(\mathbb{E}(X-\mathbb{E}(X))^{2}\right)$.

[^5]:    ${ }^{6}$ For purely causal processes, the constant vector consists of the parameters in the autoregressive polynomials, i.e., $\phi_{1}, \phi_{2}, \ldots, \phi_{p}$ in (1.3.3)

[^6]:    ${ }^{7}$ Another appropriate choice is the estimator from the ordinary least square of $\phi_{j}, j=1,2, \ldots, p$ in (1.2.1).

[^7]:    ${ }^{8}$ In our case, actually only $\theta_{0}(\tau)$ is required to be a uniformly bounded function from $\Upsilon \subset(0,1)$ to $\Theta \subset \mathbb{R}$, and the rest $\theta_{j}(\tau)$ are mapped to a constant for $j=1,2, \ldots, p . \forall \tau \in(0,1)$.

[^8]:    ${ }^{9}\lfloor z\rfloor$ denotes the largest integer that does not exceed z .

[^9]:    ${ }^{10}$ Truncated at a sufficiently large value to ensure the existence of variance.

[^10]:    ${ }^{11}$ The access of the replication data can be found in Fries and Zakoïan (2019)

[^11]:    ${ }^{12}$ The coefficients of the non-causal components in the $\operatorname{MAR}(r, s)$ models are not closed to the unity.

[^12]:    ${ }^{13}$ Additive property states the sum of independent variables from the same distribution would follow the distribution from the same family. The common distributions which share this property are $\alpha$-stable distribution, exponential distribution, geometric distribution, etc.

[^13]:    ${ }^{14}$ That is, an autoregressive moving average model where all of the roots of autoregressive polynomials are the reciprocals of the roots of moving average polynomials.

[^14]:    ${ }^{15}$ Though a linear expansion on the expectation of the indicator function is possible to be derived, where the remainder is shown to converge at a slower rate than $O_{p}\left(T^{-1 / 2}\right)$, see Kim and Pollard (1990). However, whether or not the methodology can be applied to our problem is unclear and even if it does, it can be very difficult technically.

[^15]:    ${ }^{16}$ Logistic distribution can also be a good candidate and it does not make much difference compared with the Gaussian distribution.

[^16]:    ${ }^{17} \kappa_{3}$ is skewness and $\kappa_{4}$ is kurtosis.

[^17]:    ${ }^{18}$ In the first experiment of $\mathrm{AR}(1)$ processes, we report the proportion of the correct root identification and the precision of the identified roots, in terms of the bias and the mean squared error, since the root is equal to the coefficient under the circumstance. However, in the case of $\mathrm{AR}(p)$ with orders greater than 1 , the roots of autoregressive polynomials are obtained through the factorization of coefficients, which can be very sensitive to this procedure. For example, $Y_{t}-3.1 Y_{t-1}+1.5 Y_{t-2}=(1-0.6 L)(1-2.5 L) Y_{t}$; with mild changes in the coefficients, $Y_{t}-2.8 Y_{t-1}+1.81 Y_{t-2}=(1-1.03 L)(1-1.77 L) Y_{t}$, which yields a totally different conclusion on the roots. Thus, there is not much meaning to discuss the precision of the roots in magnitude in AR processes with higher orders.

[^18]:    ${ }^{19}$ For the estimates based on the indicator function, both empirical $c d f$ approach and standard normal $c d f$ approach do not involve numerical integration. However, in the case of estimates using smooth functions $\Lambda$, only empirical $c d f$ and corresponding $\Lambda$ can avoid extra computational burdens. Besides, it has been shown that no significant difference between the empirical cdf approach and the standard normal $c d f$ approach in terms of the finite sample performance of the estimates. Therefore we select the empirical $c d f$ of residuals as weighting functions for simplicity.

[^19]:    ${ }^{20}$ The existence of asymmetric information in the market activates investors not only to forecast the future economic variables but also the forecasts of other investors, which forms higher-orders of belief dynamics.

