

Noncausal AR-ARCH model and its applications to financial time series

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Zhan, Y., Ling, S., Liu, Z. and Wang, S. ORCID:
<https://orcid.org/0000-0003-2113-5521> (2025) Noncausal AR-ARCH model and its applications to financial time series. International Journal of Finance & Economics. ISSN 1099-1158 doi: 10.1002/ijfe.3171 Available at <https://centaur.reading.ac.uk/122241/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/ijfe.3171>

Publisher: Wiley-Blackwell

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur


CentAUR

Central Archive at the University of Reading

Reading's research outputs online

SPECIAL ISSUE ARTICLE OPEN ACCESS

Noncausal AR-ARCH Model and Its Applications to Financial Time Series

Yaosong Zhan^{1,2} | Shiqing Ling³ | Zhenya Liu⁴ | Shixuan Wang⁵ 

¹Business School, Sun Yat-sen University, Shenzhen, China | ²Research Center for Innovation, Entrepreneurship, and Technology Finance, Sun Yat-sen University, Shenzhen, China | ³Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, China | ⁴EM Normandie Business School, Métis Lab, Caen, France | ⁵Department of Economics, University of Reading, Reading, UK

Correspondence: Shixuan Wang (shixuan.wang@reading.ac.uk)

Received: 29 September 2024 | **Revised:** 28 February 2025 | **Accepted:** 5 April 2025

Funding: Yaosong Zhan's research was supported by the National Natural Science Foundation of China (grant number 72403260). Shiqing Ling's research was partially supported by Hong Kong Research Grants Commission Grants (grant numbers 16303118, 16301620, 16300621, 16500522, and SRF522236S02).

Keywords: financial time series | noncausal AR-ARCH | noncausal variance

ABSTRACT

We extend the noncausal autoregressive models by introducing noncausality into the variance component, allowing the volatility to depend on future prices as well. We refer to this model as the noncausal AR-ARCH model, and it enables us to account for shocks arising from market agents who possess more information and engage in forward-looking trading behaviours, leading to a better fit for financial time series. In terms of parameter estimation, we develop a quasi-maximum likelihood estimation method and establish its asymptotic properties. Building on this, we propose three hypothesis testing statistics to determine whether the data exhibits a noncausal AR structure and whether the innovation term follows a noncausal ARCH model. The simulation results demonstrate the consistency of the parameter estimation as well as the good size control and high power of the hypothesis tests in detecting noncausal structures. In our empirical applications, we employ the proposed model in both stock markets and crude oil futures markets. Our empirical findings indicate that the variance is causal in the US stock market but noncausal in the Chinese stock market. Furthermore, we observe a noticeable distinction between Brent and WTI crude oil futures, as Brent exhibits noncausality in both its mean and variance, whereas WTI follows a purely causal process.

JEL Classification: C22, C51, G12

1 | Introduction

The noncausal autoregressive (NAR) model has gained attention in time series modelling for its ability to capture nonlinear patterns within a stationary framework. Unlike causal models, which rely solely on lags, noncausal models incorporate both lags and leads, enabling dependence on future values. This makes them well suited to model the nonlinear and asymmetric features often observed in financial time series, such as asset price bubbles (Nyberg et al. 2012). While causal models struggle to capture such phenomena, the NAR model can effectively accommodate these explosive dynamics (Gouriéroux and Zakoian 2013).

The noncausality in asset price data primarily stems from differences in the information sets available to agents. Firstly, when agents possess a greater information set (than an econometrician who is estimating a univariate model) and engage in forward-looking trading activities, a causal model may lead to a non-fundamental solution based on rational expectations (Lanne and Saikkonen 2011). Secondly, agents in the market have heterogeneous expectations, yet the model may assume homogeneous beliefs, which can also lead to non-fundamental solutions in equilibrium (Lof 2013). Due to the limitations of causal models, they often cannot accommodate different information sets and heterogeneous agents. NAR models, however,

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *International Journal of Finance & Economics* published by John Wiley & Sons Ltd.

can account for the true information sets of market agents in a non-explicit manner, allowing for a better fit for the financial time series data.

Noncausality implies that asset prices are the result of agents making predictions about the future based on their information and then engaging in forward-looking trading behaviours. The NAR model has been proposed to analyse the mean of the data, which are approximated through a noninvertible moving average process (Lanne and Saikkonen 2011). In a NAR model, the current value is influenced not only by past and present values but also by future values. Additionally, the error term in such a model is predictable and can be considered a non-fundamental shock (Alessi et al. 2011).

Importantly, noncausality might also be presented in the volatility as well, which has not been considered in the literature. Volatility reflects the magnitude of shocks to asset prices, and these shocks are often attributed to the trading activities of noise traders, who are typically considered irrational and trade randomly. As a result, the shocks are often considered to be white noise. However, shocks can also stem from some market participants who have more information about future prices. On one hand, some participants react to factors like rumours or the actions of other traders, which can impact volatility (Brown 1999). This type of information is usually excluded from NAR models. On the other hand, market agents may possess exceptional professional abilities, enabling them to gather more information from the market and gain insights that influence their trading behaviour, which can lead to forward-looking shocks in asset prices. Kyle (1985) proposes that traders with private information about an asset's forward-looking value can impact prices through varying trade sizes, with price volatility reflecting the incorporation of this information. Aase et al. (2012) suggest that noise traders, who are partially informed, can correlate their trades with the true asset price, making their actions non-random and introducing forward-looking patterns into price volatility. Bottazzi et al. (2011) provides experimental evidence showing that market agents' forward-looking expectations influence price confidence intervals, which affect price fluctuations. Thus, their forward-looking trading activities introduce noncausality in volatility.

Therefore, we introduce the noncausality into the variance component of the noncausal AR-ARCH model for financial time series. Noticeably, the variance of the innovation term comprises two components: (1) a fixed variance that represents the shocks to asset prices from the irrational behaviour of noise traders, and (2) a variance that is related to future prices, reflecting the shocks that arise from the forward-looking trading activities of market agents who possess a larger set of information. For parameter estimation, we develop a quasi-maximum likelihood estimation (QMLE) method and establish its asymptotic properties. Building on this, we propose three hypothesis tests to determine whether the data exhibit a noncausal AR structure and whether the innovation term follows a noncausal ARCH model.

We conduct a simulation analysis of our model and compare its performance with existing causal and noncausal processes. Our

findings show that processes with noncausality in both the mean and variance demonstrate increased volatility and pronounced volatility clustering, characteristics frequently encountered in financial markets. Regarding the performance of QMLE and hypothesis testing, we find that the parameter estimates from the developed QMLE converge to the true parameter values with the growth of sample sizes. We also achieve good size control and high power for the three hypothesis tests under different data generation processes. Additionally, when comparing the proposed AR-ARCH model with other benchmark models (including causal and mixed causal-noncausal AR models), it consistently outperforms them in terms of higher likelihood, and lower Bayesian information criterion (BIC) and mean squared error (MSE) values.

In our empirical applications, we employ the noncausal AR-ARCH model to stock markets and crude oil markets. For the stock markets, we perform a detailed analysis of Price-to-Earnings (PE) ratio processes of the S&P 500 index in the US and the CSI 300 index in China. We find that a NAR model provides a better fit for both PE processes. However, there are several differences in our findings. The S&P 500 PE exhibits a two-order lead, while the CSI 300 PE has only a one-order lead. Additionally, the volatility of the CSI 300 PE demonstrates noncausality, whereas the volatility of the S&P 500 is causal. The differences in the noncausal structures of the two PE processes reflect that in the US stock market, market participants have a longer term forward-looking perspective, resulting in more lead terms, while the shocks are only generated by noise traders. In contrast, in the Chinese stock market, shocks originate from both noise traders and market participants who possess more information and engage in forward-looking trades. Consequently, the shocks on the asset price are not purely random noise. For the crude oil markets, we focus on the West Texas Intermediate (WTI) and Brent futures and examine their basis (the difference between the futures price and spot price). Although these two types of crude oil futures are highly similar, our empirical findings indicate that the WTI crude oil futures basis does not exhibit noncausality and can be adequately modelled by an AR(4) model. In contrast, the Brent crude oil futures demonstrate noncausality in both mean and variance, indicating that using a noncausal AR-ARCH model can uncover subtle differences in similar asset price dynamics.

Our contributions to the existing literature primarily lie in three aspects. First, we extend the existing NAR model by incorporating noncausality into the volatility component to achieve a better fit to financial data. Under the rational expectations setting, Hansen and Sargent (1991) find that if investors have access to a larger information set compared to economists, the solution of the linear rational expectation models becomes noncausal. Lanne and Saikkonen (2011) argue that the larger information set of investors is manifested in two aspects. Firstly, investors have access to more information than economists, leading to a problem of missing variables in the modelling. Secondly, there exists heterogeneous information or belief among investors. Based on this, they propose the NAR model to describe data with such characteristics. Lof (2013) also supports this finding and further proposes that a larger information set also reflects that investors

have knowledge of nonlinear relationships among financial variables. They validated this proposition through simulation analysis. Gouriéroux et al. (2020) show that relaxing the assumption of finite variance still results in a noncausal stationary equilibrium. Additionally, performing seasonal adjustments on economic data also introduces noncausality (Hecq et al. 2017). Existing literature has primarily focused on introducing noncausality in the mean of autoregressive models. However, understanding the volatility of economic and financial time series is crucial, and current research on the noncausality of the innovation term remains limited. Our study bridges the gap by linking the ARCH model with noncausality, investigating the noncausality of innovation terms, and finding that it can effectively capture empirical volatility features.

Secondly, we develop a QMLE method for the parameter estimation of the proposed noncausal AR-ARCH model and establish its asymptotic properties. Parameter estimation in noncausal models is primarily based on approximate maximum likelihood estimation (Breid et al. 1991). Lii and Rosenblatt (1996) study the maximum likelihood estimation method under non-Gaussian noise conditions. Under certain conditions, the estimated parameters asymptotically approach the results under normal noise. Wu and Davis (2010) study the performance of an estimator based on least absolute deviation (LAD) for noncausal models, and the estimator exhibits consistency and asymptotic normality. When the probability distribution function of the volatility is unknown, the Quasi-Likelihood Estimation method proposed by Huang and Pawitan (2000) can estimate parameters consistently. Unlike the approximate maximum likelihood estimation, our QLME method does not necessitate the specification of the full probability distribution and also allows for a more flexible variance. Additionally, our QMLE can provide estimates when both the mean and variance of the model are contingent upon future values. To determine the noncausal order for estimations, Gouriéroux and Jasiak (2018) examine the properties of order misspecification in noncausal processes and propose a robust approach to identify the noncausal order. In order to address the practical difficulties of existing estimation methods, Cavaliere et al. (2020) propose a bootstrapping method, which does not require any assumptions about the distribution of volatility, yet yields consistent estimators.

Our QMLE method provides estimates on the coefficients of the AR component as well as the coefficients of the ARCH component. It captures the distribution of the innovation term, which is essential for probabilistic forecasting using noncausal models. Existing forecasting methods are based on assumptions about the probability distribution of the innovation term. Lanne et al. (2012) propose a simulations-based forecasting approach. The main idea is to simulate the noncausal part of the error term to obtain the probability distribution of the time series in the future, enabling point prediction and density prediction. Building upon this, Gouriéroux and Jasiak (2016) provide a closed-form function prediction method. It leverages the counterpart of the sample-based innovation term to represent the theoretical expectation, thus allowing for the prediction of future density. Regardless of whether it is the sample-based approach or simulation-based approach, the prediction results converge to

the same distribution related to the innovation term (Hecq and Voisin 2021). Fries (2021) also proposes a closed-form formulation of the distribution during explosive bubble episodes under the varying tail assumption.

Thirdly, we also make an empirical contribution by examining the noncausal structure in the financial time series, providing further insights into the stock markets between the US and China as well as crude oil futures markets. Lanne and Saikkonen (2011) consider a close relationship between inflation rates and people's expectations, and apply noncausal models to fit the inflation dynamics in the United States. The results show that the root mean square error (RMSE) obtained using the noncausal model is smaller than that of other causal AR models. In fitting the behaviour of financial market bubbles, Gouriéroux and Zakoian (2013), Hencic and Gouriéroux (2015), Gouriéroux and Zakoian (2017), and Fries and Zakoian (2019) use data such as stock prices, the NASDAQ composite price index, and the exchange rate of Bitcoin to the US dollar, noncausal models are able to effectively capture the local intense changes in the data, identifying bubble characteristics. Lof and Nyberg (2017) find that noncausal models fit commodity prices well and can explain the predictive power of exchange rates for commodity prices. Our study considers the noncausal structure in the volatility and reveals differences between the Chinese and US stock markets. Specifically, the volatility structure of the CSI 300 index exhibits noncausality, while the S&P 500 index's volatility remains causal. Additionally, our findings show that even within highly similar commodity financial time series, such as crude oil futures, there are notable differences in volatility.

The paper is structured as follows: Section 2 outlines the construction of the noncausal AR-ARCH model, including the method for estimating parameters and conducting hypothesis tests. Section 3 presents the results of simulation analyses. Section 4 provides the empirical studies on the stock markets and crude oil prices. Finally, the paper is concluded in Section 5.

2 | The Model and Estimation Method

In this section, we first present the theoretical framework of the model. We detail the underlying assumptions and the equilibrium system that governs the asset price dynamics. Following this, we develop a QMLE method to estimate parameters of the model, along with presenting its consistency and asymptotic normality. This includes deriving the quasi-maximum likelihood estimator and discussing its properties and asymptotic behaviour.

2.1 | The Model

At time t , we assume that the price P_t of an asset depends on its demand d_t and supply q_t . Following Gouriéroux et al. (2020), the equilibrium system is formulated as follows:

$$\begin{cases} d_t = \alpha_1 P_t + \alpha_2 \mathbb{E}_t(P_{t+1}) + z_{1,t}, & \text{(demand equation),} \\ q_t = \beta P_t + z_{2,t}, & \text{(supply equation)} \end{cases} \quad (1)$$

where the variables $z_{1,t}$ and $z_{2,t}$ denote the demand and supply shocks. We assume that $z_{1,t}$ and $z_{2,t}$ are mean-zero random variables. The variances of $z_{1,t}$ and $z_{2,t}$ are composed of two terms:

$$\sigma^2(z_{i,t}) = \rho_i^2 + k_i \sigma_t^2(P_{t+1}), \quad i = 1, 2 \quad (2)$$

where ρ_1 and ρ_2 represent the fixed variance associated with shocks from noise traders; k_1 and k_2 are distinct parameters reflecting the impact of these shocks on the supply and demand components; $\sigma_t^2(P_{t+1})$ denotes the variance that depends on future prices, originating from market participants who engage in forward-looking trading activities, defined as follows:

$$\sigma_t^2(P_{t+1}) = \mathbb{E}_t(P_{t+1})^2. \quad (3)$$

We set $d_t = q_t$ to obtain the equilibrium price P_t ,

$$P_t = \phi \mathbb{E}_t(P_{t+1}) + e_t. \quad (4)$$

with $\phi = \alpha_2 / (\beta - \alpha_1)$, $e_t = (z_{1,t} - z_{2,t}) / (\beta - \alpha_1)$, and

$$e_t \sim \mathcal{N}\left(0, \frac{\rho_1^2 + \rho_2^2 + (k_1 + k_2) \sigma_t(P_{t+1})^2}{(\beta - \alpha_1)^2}\right). \quad (5)$$

We observe that both equilibrium price P_t and its volatility are driven by the market expectation on the $(t+1)^{\text{th}}$ period as shown in Equation (4). Let $S_{t+1} = \mathbb{E}_t(P_{t+1})$, and we can simplify Equation (4) into the following process:

$$\begin{cases} P_t = \phi S_{t+1} + e_t, \\ e_t = \eta_t (\omega + \alpha S_{t+1}), \\ S_{t+1} = P_{t+1} + \xi_{t+1}, \end{cases} \quad (6)$$

where ξ_t is the expected errors, which is assumed to have zero mean and σ_ξ^2 variance, the variable e_t represents the shock on the price P_t , ω describes the fixed variance, αS_{t+1} captures the variance related with forward-looking trading, and η_t is the standard normal random variable.

We can rewrite Equation (6) as follows:

$$\begin{cases} P_t = \phi P_{t+1} + u_t, \\ u_t = \eta_t (\omega + \alpha P_{t+1}) + \xi_{t+1} (\phi + \alpha \eta_t). \end{cases} \quad (7)$$

By setting $E(u_t | P_{t+1}) = 0$, the variance of u_t is

$$\begin{aligned} h_{t+1} &\equiv \mathbb{E}(u_t^2 | P_{t+1}) \\ &= \mathbb{E}\left[(\omega \eta_t + \alpha \eta_t P_{t+1})^2 | P_{t+1}\right] + \mathbb{E}(\phi \xi_{t+1} + \alpha \eta_t \xi_{t+1})^2 \\ &= (\omega + \alpha P_{t+1})^2 + \mathbb{E}(\phi + \alpha \eta_t)^2 \sigma_\xi^2 \\ &= (\omega + \alpha P_{t+1})^2 + (\phi^2 + \alpha^2) \sigma_\xi^2 \\ &= \sigma_u^2 + (\omega + \alpha P_{t+1})^2, \end{aligned} \quad (8)$$

where $\sigma_u^2 = (\phi^2 + \alpha^2) \sigma_\xi^2$. Thus, we define Equation (7) as a non-causal AR-ARCH (NAR-NARCH) model, and it should be

highlighted that the variance depends on future values. The following theorem provides the solution for the NAR-NARCH model.

Theorem 2.1.

i. The process $\{P_t\}$ from Equation (7) has the representation:

$$P_t = \sum_{i=1}^{\infty} \prod_{j=0}^i (\phi + \alpha \eta_{t+j}) \bar{\xi}_{t+i} + \bar{\xi}_t \quad (9)$$

if and only if $v \equiv \mathbb{E} \log |\phi + \alpha \eta_t| < 0$ where $\bar{\xi}_t = \omega \eta_t + \phi \xi_{t+1} + \alpha \eta_t \xi_{t+1}$, and $\mathbb{E}|P_t|^\delta < \infty$ for some $\delta \in (0, 1)$.

ii. Define $\bar{\xi}_t^* = \eta_t + \phi \xi_{t-1} + \alpha \eta_t \xi_{t-1}$ and

$$P_t^* = \sum_{i=1}^{\infty} \prod_{j=0}^i (\phi + \alpha \eta_{t-j}) \bar{\xi}_{t-i}^* + \bar{\xi}_t^*.$$

Then, for $\forall f: \mathbb{R}^m \rightarrow \mathbb{R}$,

$$f(P_t, P_{t+1}, \dots, P_{t+m}) = f(P_t^*, P_{t+1}^*, \dots, P_{t+m}^*),$$

and $\{P_t^*\}$ is stationary and ergodic with $\mathbb{E}|P_t^*|^\delta < \infty$.

The proof is given in Appendix A: Proof of Theorem 2.1. To ensure the condition that $v = \mathbb{E} \log |\phi + \alpha \eta_t| < 0$, the region of (ϕ, α) is given in Figure 1.

2.2 | Estimation Method

We supposing that P_1, P_2, \dots, P_n are generated by Equation (7). Since u_t is not necessary to be normal, the conditional quasi log-likelihood function can be written as follows:

$$\mathcal{L}(\theta) = - \sum_{t=1}^n \left\{ \log \left[\sigma_u^2 + (\omega + \alpha P_{t+1})^2 \right] + \frac{(P_t - \phi P_{t+1})^2}{\sigma_u^2 + (\omega + \alpha P_{t+1})^2} \right\},$$

where $\theta = (\sigma_u^2, \phi, \omega, \alpha)$. The maximizer, $\hat{\theta}_n$, of $\mathcal{L}(\theta)$ is called the quasi-maximum likelihood estimator (QLME). We assume the parameter θ satisfies the following condition:

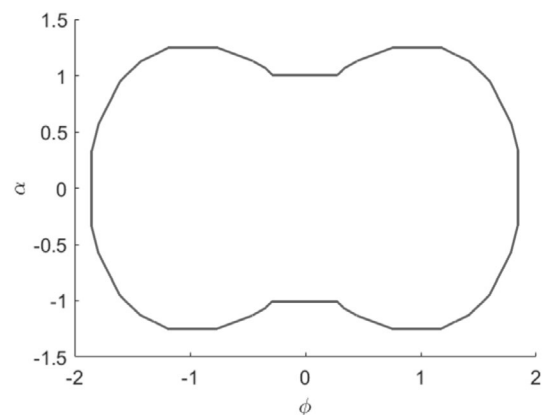


FIGURE 1 | The region of (ϕ, α) with $v < 0$.

Assumption 2.1. The parameter space is

$$\Theta = \{\theta = (\phi, \sigma_u^2, \omega, \alpha) : \mathbb{E} \log |\phi + \alpha \eta_t| < 0\},$$

where ϕ , ω , and α being bounded by a constant B , and $\sigma_u \leq \sigma_u^2 \leq \bar{\sigma}_u$, B , σ_u , and $\bar{\sigma}_u$ are some finite positive constants, and θ_0 is an interior point in Θ .

The t -th term in the log-likelihood function is as follows:

$$\mathcal{M}_t(\theta) = -\log \left[\sigma_u^2 + (\omega + \alpha P_{t+1})^2 \right] - \frac{(P_t - \phi P_{t+1})^2}{\sigma_u^2 + (\omega + \alpha P_{t+1})^2}. \quad (10)$$

Its partial derivatives can be represented as follows:

$$\frac{\partial \mathcal{M}_t(\theta)}{\partial \phi} = \frac{2P_{t+1}(P_t - \phi P_{t+1})}{\sigma_u^2 + (\omega + \alpha P_{t+1})^2}. \quad (11)$$

Let $\lambda = (\sigma_u^2, \omega, \alpha)'$, then

$$\frac{\partial \mathcal{M}_t(\theta)}{\partial \lambda} = -\frac{\partial h_{t+1}(\lambda)}{\partial \lambda} \left[1 - \frac{(P_t - \phi P_{t+1})^2}{h_{t+1}(\lambda)} \right] \frac{1}{h_{t+1}(\lambda)}, \quad (12)$$

where

$$h_{t+1}(\lambda) = \sigma_u^2 + (\omega + \alpha P_{t+1})^2 \quad (13)$$

and

$$\frac{\partial h_{t+1}(\lambda)}{\partial \lambda} = [1, 2(\omega + \alpha P_{t+1}), 2(\omega + \alpha P_{t+1})P_{t+1}]^T, \quad (14)$$

$$\frac{\partial^2 \mathcal{M}_t(\theta)}{\partial \phi^2} = -\frac{2P_{t+1}^2}{h_{t+1}(\lambda)}, \quad (15)$$

$$\frac{\partial^2 \mathcal{M}_t(\theta)}{\partial \lambda \partial \lambda'} = \frac{1}{h_{t+1}(\lambda)^2} \left[\frac{\partial h_{t+1}(\lambda)}{\partial \lambda} \frac{\partial h_{t+1}(\lambda)}{\partial \lambda'} \right] \left[1 - \frac{2(P_t - \phi P_{t+1})^2}{h_{t+1}(\lambda)} \right], \quad (16)$$

$$\frac{\partial^2 \mathcal{M}_t(\theta)}{\partial \lambda \partial \phi} = -\frac{\partial h_{t+1}(\lambda)}{\partial \lambda} \frac{2P_{t+1}(P_t - \phi P_{t+1})}{h_{t+1}(\lambda)^2}. \quad (17)$$

Furthermore, we can derive the asymptotic distribution of the estimators.

Theorem 2.2. Suppose that $\{P_1, P_2, \dots, P_n\}$ is a solution of Equation (7) with Equation (9) and Assumption 2.1 holding, then as $n \rightarrow \infty$, it follows that.

i. $\hat{\theta}_n \rightarrow \theta_0$ in probability, where θ_0 is the true parameter.

ii. Furthermore, if $\mathbb{E} \eta_t^4 < \infty$ and $\mathbb{E} \xi_t^4 < \infty$, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathcal{N}(0, \Sigma^{-1} \Omega \Sigma^{-1}). \quad (18)$$

Because $E(u_t | P_{t+1}) = 0$, Σ can be represented as

$$\Sigma = -\mathbb{E} \begin{bmatrix} \frac{2P_{t+1}^2}{h_{t+1}} & 0 \\ 0 & \frac{1}{h_{t+1}(\lambda)^2} \left[\frac{\partial h_{t+1}(\lambda)}{\partial \lambda} \frac{\partial h_{t+1}(\lambda)}{\partial \lambda'} \right] \end{bmatrix}. \quad (19)$$

And we have

$$\Omega = \mathbb{E} \left[\frac{\partial \mathcal{M}_t(\theta)}{\partial \theta} \frac{\partial \mathcal{M}_t(\theta)}{\partial \theta} \right]_{\theta=\theta_0} = \mathbb{E} \begin{bmatrix} \frac{4P_{t+1}^2 u_t^2}{h_{t+1}(\lambda)^2} & -\frac{2P_{t+1} u_t}{h_{t+1}(\lambda)^2} \left(1 - \frac{u_t^2}{h_{t+1}(\lambda)} \right) \frac{\partial h_{t+1}(\lambda)}{\partial \lambda'} \\ -\frac{2P_{t+1} u_t}{h_{t+1}(\lambda)^2} \left(1 - \frac{u_t^2}{h_{t+1}(\lambda)} \right) \frac{\partial h_{t+1}(\lambda)}{\partial \lambda} & \frac{1}{h_{t+1}(\lambda)^2} \frac{\partial h_{t+1}(\lambda)}{\partial \lambda} \frac{\partial h_{t+1}(\lambda)}{\partial \lambda'} \left(1 - \frac{u_t^2}{h_{t+1}(\lambda)} \right)^2 \end{bmatrix}. \quad (20)$$

When η_t and ξ_t are systemics, Σ and Ω can be estimated by their sample version:

$$\hat{\Sigma} = -\frac{1}{n} \sum_{t=1}^n \left[\frac{\partial^2 \mathcal{M}_t(\theta)}{\partial \theta \partial \theta'} \right]_{\theta=\hat{\theta}_n}, \quad (21)$$

$$\hat{\Omega} = \frac{1}{n} \sum_{t=1}^n \left[\frac{\partial \mathcal{M}_t(\theta)}{\partial \theta} \frac{\partial \mathcal{M}_t(\theta)}{\partial \theta'} \right]_{\theta=\hat{\theta}_n}, \quad (22)$$

where $\hat{h}_{t+1}(\hat{\lambda}_n) = \hat{\sigma}_{nu} + (\hat{\omega}_n + \hat{\alpha}_n P_{t+1})^2$ and

$$\frac{\partial \hat{h}_{t+1}(\hat{\lambda}_n)}{\partial \lambda} = [1, 2(\hat{\omega}_n + \hat{\alpha}_n P_{t+1}), 2(\hat{\omega}_n + \hat{\alpha}_n P_{t+1})P_{t+1}]'. \quad (23)$$

The proof is given in Appendix A: Proof of Theorem 2.2. Then, we can demonstrate that $\hat{\Sigma}$ and $\hat{\Omega}$ serve as consistent estimators for Σ and Ω , respectively. Through Theorem 2.2, we can formulate test statistics for the null hypotheses that $\phi = 0$ and $\alpha = 0$. For any non-zero constant vector $CC' \neq 0$, we have

$$\sqrt{n}C'(\hat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}(0, C'\Sigma^{-1}\Omega\Sigma^{-1}C).$$

Let $C = (1, 0, 0, 0)'$. Then

$$T_1 = \frac{[\sqrt{n}C'(\hat{\theta} - \theta_0)]^2}{C'\hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1}C} = \frac{[\sqrt{n}(\hat{\phi}_n - 0)]^2}{C'\hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1}C} \sim \chi^2(1), \quad (24)$$

which can be used to test the hypothesis:

$$H_0: \phi = 0 \text{ vs. } H_A: \phi \neq 0.$$

This test determines whether a noncausal structure exists in the mean function of the AR process. Let $C = (0, 0, 0, 1)'$. Then

$$T_2 = \frac{[\sqrt{n}C'(\hat{\theta} - \theta_0)]^2}{C'\hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1}C} = \frac{[\sqrt{n}(\hat{\alpha}_n - 0)]^2}{C'\hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1}C} \sim \chi^2(1), \quad (25)$$

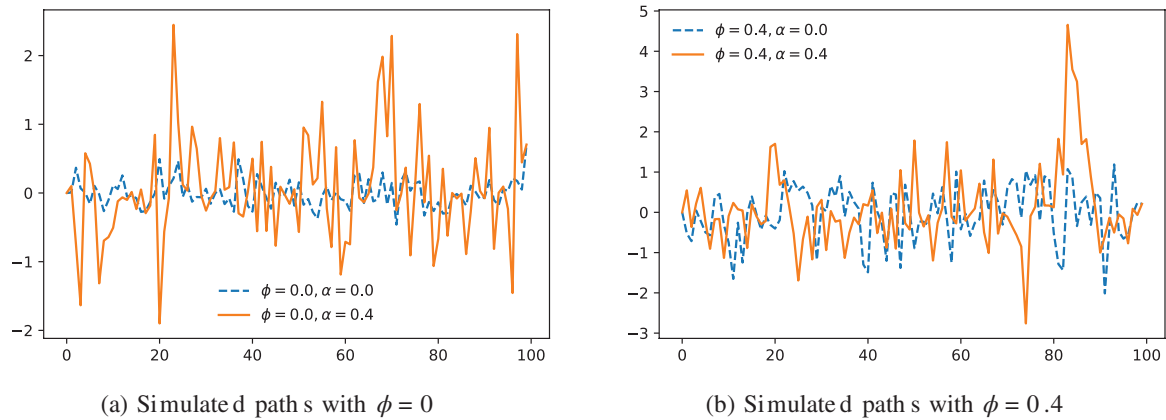


FIGURE 2 | Simulated paths of causal and noncausal processes. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

which can be used to test the hypothesis:

$$H_0: \alpha = 0 \text{ vs. } H_A: \alpha \neq 0.$$

This test determines whether a noncausal structure exists in the mean function of the ARCH process. Let $C = \begin{bmatrix} 1, 0, 0, 0 \\ 0, 0, 0, 1 \end{bmatrix}'$, then

$$T_3 = \left[\sqrt{n} C' (\hat{\theta}_n - \theta) C \right]' \left(C' \hat{\Sigma}_n^{-1} \hat{\Omega}_n \hat{\Sigma}_n^{-1} C \right)^{-1} \left[\sqrt{n} C' (\hat{\theta}_n - \theta) C \right] \sim \chi^2(2), \quad (26)$$

which can be used to test the hypothesis:

$$H_0: \phi = \alpha = 0 \text{ vs. } H_A: \phi \neq 0 \text{ or } \alpha \neq 0.$$

This test examines for noncausal structures in both the mean and volatility. To test whether the process has noncausal components, T_3 can first be applied. If it is rejected, it implies that there exist noncausal components. Subsequently, T_1 and T_2 are applied to test whether the AR or ARCH model exhibits a noncausal structure.

3 | Simulation

We conduct a series of simulation study to investigate the finite sample performance of the QMLE and the three hypothesis tests. Based on Equation (7), we initially generate the terminal value P_T randomly from the interval (0, 1) and subsequently compute the preceding values P_t , where $t = T - 1, T - 2, \dots, 1$, through an inverse recursion sequence. As an illustration, Figure 2a,b compare the simulated time series with and without noncausal structure, respectively. We can observe that when the process is a noncausal AR with $\phi \neq 0$, the fluctuations are severe, and significant local explosions can be observed. While traditional models like AR-GARCH are effective in capturing conditional heteroscedasticity and the mixed causal-noncausal AR model (MAR) is useful for modelling nonlinear stationary processes, both often struggle with the noncausality observed in process variance. As shown in Figure 2, introducing noncausal variance ($\alpha = 0.4$) leads to more clustered and severe fluctuations as the process value increases, while volatility becomes milder when the process value is low. In contrast, the NAR-NARCH model offers a more suitable framework for these data due to its flexibility in modelling both present and future dependencies in mean

and variance. The noncausal structure allows for a more accurate representation of volatility dynamics and anticipates shocks from future events. By incorporating noncausal components in volatility, this model better captures the forward-looking behaviour of market participants and their trading disturbances, providing improved fitting of process volatility.

Next, we investigate the finite sample performance of our QMLE method. The parameters ϕ and α determine the noncausal structure of the process, representing the dependence of the mean and variance on forward-looking values. We consider four different combinations of ϕ and α values: $(\phi, \alpha) = \{(0.3, 0.3), (0.3, 0.7), (0.7, 0.3), (0.7, 0.7)\}$. The first combination implies that the mean and variance of the process are less influenced by the future values. Conversely, the second and third combinations imply that the process is more dependent on the future values in either the mean or the variance. The last combination represents the case where both the mean and variance of the process heavily rely on the future values. The parameter ω is set at 0.2. The expected error, ξ_t , is assumed to follow either a normal distribution or a Student's t -distribution with 4 degree of freedom, a mean of zero, and a variance of 0.5. Hence, the values of σ_u^2 equal 0.09 for $(\phi, \alpha) = (0.3, 0.3)$, 0.29 for $(\phi, \alpha) = (0.3, 0.7)$ and $(0.7, 0.3)$, and 0.49 for $(\phi, \alpha) = (0.7, 0.7)$. We consider four sample sizes, 100, 250, 500, and 1000.

Table 1 provides the simulation results. It reports the means, standard deviations, and theoretical standard deviations (TSD) of the estimators. The TSD is calculated as the square root of the i -th diagonal element of $\hat{\Sigma}^{-1} \hat{\Omega} \hat{\Sigma}^{-1}$, and it represents the true standard deviations corresponding to the estimated parameters. The closer the simulated standard deviations are to the TSD, the more accurate the simulation results are.

The results indicate that the TSD approximates the simulated results as the sample size grows to 1000 for the combination of $(\phi, \alpha) = (0.3, 0.3)$. Additionally, the TSD and standard deviations closely align for all sample sizes for the combinations $(\phi, \alpha) = (0.3, 0.7)$ and $(0.7, 0.3)$. Regarding the estimated values, even with a small sample size, the estimations for each combination are accurate and closely approximate the true values. Besides, the standard deviations decrease as the sample size increases.

TABLE 1 | Finite-sample properties of QLME estimator.

DGP										
N	Parameter	$\phi = 0.3, \alpha = 0.3$			$\phi = 0.3, \alpha = 0.7$			$\phi = 0.7, \alpha = 0.3$		
		Mean	St.dev.	TSD	Mean	St.dev.	TSD	Mean	St.dev.	TSD
Panel A: ξ_t follows normal distribution										
100	ϕ	0.284	0.132	0.149	0.262	0.161	0.162	0.660	0.104	0.090
	ω	0.201	0.229	2.291	0.249	0.298	0.284	0.308	0.313	3.501
	α	0.333	0.238	0.940	0.632	0.314	0.298	0.237	0.192	2.225
	σ_u^2	0.059	0.167	0.479	0.136	0.217	0.244	0.226	0.249	2.734
250	ϕ	0.291	0.072	0.077	0.288	0.092	0.110	0.686	0.058	0.049
	ω	0.247	0.138	1.423	0.216	0.149	0.164	0.268	0.199	2.242
	α	0.289	0.144	0.570	0.681	0.164	0.167	0.258	0.118	0.140
	σ_u^2	0.061	0.079	0.479	0.214	0.087	0.108	0.243	0.192	2.029
500	ϕ	0.295	0.050	0.054	0.294	0.063	0.061	0.693	0.039	0.032
	ω	0.241	0.101	0.962	0.204	0.097	0.103	0.230	0.124	1.492
	α	0.285	0.108	0.400	0.691	0.102	0.095	0.279	0.082	0.095
	σ_u^2	0.064	0.053	0.427	0.265	0.050	0.055	0.250	0.107	1.264
1000	ϕ	0.298	0.034	0.036	0.298	0.044	0.041	0.696	0.027	0.022
	ω	0.226	0.080	0.747	0.202	0.067	0.062	0.213	0.073	0.773
	α	0.287	0.081	0.283	0.695	0.070	0.065	0.290	0.051	0.054
	σ_u^2	0.073	0.032	0.258	0.285	0.035	0.037	0.284	0.014	0.447
Panel B: ξ_t follows Student's t -distribution										
100	ϕ	0.280	0.133	0.126	0.262	0.165	0.207	0.660	0.106	0.115
	ω	0.199	0.239	2.302	0.252	0.309	0.347	0.321	0.321	1.558
	α	0.358	0.268	0.801	0.678	0.349	0.382	0.256	0.232	0.226
	σ_u^2	0.056	0.155	0.465	0.143	0.225	0.199	0.238	0.249	0.723
								0.644	0.157	0.177
								0.238	0.347	0.513
								0.672	0.268	0.335
								0.454	0.239	0.746

(Continues)

TABLE 1 | (Continued)

DGP													
N	Parameter	$\phi = 0.3, \alpha = 0.3$			$\phi = 0.3, \alpha = 0.7$			$\phi = 0.7, \alpha = 0.3$			$\phi = 0.7, \alpha = 0.7$		
		Mean	St.dev.	TSD	Mean	St.dev.	TSD	Mean	St.dev.	TSD	Mean	St.dev.	TSD
250	ϕ	0.290	0.076	0.085	0.290	0.094	0.103	0.685	0.059	0.068	0.682	0.083	0.082
	ω	0.237	0.152	1.398	0.217	0.177	0.212	0.271	0.221	1.197	0.206	0.191	0.335
	α	0.316	0.178	0.663	0.700	0.188	0.241	0.269	0.148	0.167	0.697	0.127	0.184
	σ_u^2	0.061	0.072	0.519	0.225	0.104	0.103	0.241	0.197	0.593	0.466	0.114	0.330
500	ϕ	0.294	0.051	0.058	0.296	0.065	0.086	0.693	0.040	0.040	0.692	0.057	0.078
	ω	0.235	0.116	1.123	0.208	0.120	0.149	0.233	0.146	0.954	0.206	0.133	0.199
	α	0.305	0.130	0.415	0.700	0.121	0.173	0.284	0.099	0.123	0.704	0.084	0.104
	σ_u^2	0.066	0.052	0.432	0.255	0.070	0.075	0.249	0.110	0.315	0.478	0.077	0.232
1000	ϕ	0.297	0.036	0.032	0.298	0.046	0.049	0.696	0.028	0.027	0.696	0.040	0.039
	ω	0.227	0.089	0.743	0.203	0.089	0.085	0.213	0.098	0.467	0.201	0.097	0.133
	α	0.294	0.097	0.276	0.703	0.086	0.090	0.294	0.068	0.062	0.702	0.061	0.071
	σ_u^2	0.072	0.034	0.271	0.285	0.047	0.042	0.275	0.055	0.151	0.488	0.058	0.163

The simulation findings suggest that when the noncausal structure is more pronounced in the process, the standard deviations of ϕ and α are smaller, and the QLME performs better. The results also show that the distribution of ξ_t has limited impact on the QLME estimation, as demonstrated by the comparison between Panel A and Panel B of Table 1.

The results of the hypothesis tests are presented in Table 2. T_1 , T_2 , and T_3 are the statistics computed under the null hypotheses that $\phi = 0$, $\alpha = 0$, and $\phi = \alpha = 0$, respectively. The significance level is set at 5%. As can be seen from the table, the empirical size of T_1 , T_2 , and T_3 are close to the theoretical result of 5%. Our hypothesis tests are effective in detecting noncausal structure in the process. The empirical power for the single tests of ϕ and α are close to 100%, and the test performs better for α than ϕ . The empirical power of the joint hypothesis test is also near 100% whenever the noncausal structure exists. The hypothesis testing results are consistent and perform well regardless of whether ξ_t follows a normal distribution or a Student's t -distribution. To control the experiment error rate, we further provide the T_4 statistic defined as $T_4 = \max(T_1, T_2)$, which can control the Type-I error and provide union–intersection tests (Casella and Berger 2024). The simulation results show that T_4 has a good empirical size when the sample size is around 500. In the case where $\phi \neq 0$, T_4 performs better than when $\alpha \neq 0$. Moreover, the choice of the distribution of ξ has marginal impact on the results.

The model selection results comparing the proposed NAR-NARCH model with other causal models are presented in Table 3. The comparison follows the procedure outlined in Breid et al. (1991), where we calculate the maximum likelihood functions and BIC as the criteria. The considered benchmark models include noncausal AR(1) NAR(1), ARMA(1,1), and AR(1)-GARCH(1,1).

The results indicate that for different parameter settings and sample sizes, the log-likelihood of the proposed NAR-NARCH model generally outperforms the other models. This suggests that our model is less likely to suffer from misspecification problems. Moreover, the results improve with increasing sample size, and the gap between the proposed model and the other candidates widens, particularly when α is larger. However, varying ϕ from 0.3 to 0.7 does not have a significant impact on the difference in log likelihood functions. These findings imply that α plays a more crucial role in the model selection process, as it directly influences the noncausal variance of the innovation terms. When α is small, the variance tends to converge to a causal ARCH model. Therefore, larger values of both α and ϕ lead to better model selection outcomes.

4 | Empirical Applications

In this section, we apply the model to stock markets and crude oil futures markets. By examining market data, we aim to uncover whether the noncausal structure exists in these financial time series data, providing insights into the underlying market dynamics.

4.1 | Stock Markets

We apply our proposed noncausal AR-ARCH model to examine whether the PE ratio processes of the stock market indexes have a noncausal structure. Our study is based on the fact that investors' forward-looking behaviour and the broader information sets they possess can lead to a NAR pattern in stock markets (Lanne and Saikkonen 2011; Lof 2013).

We utilise daily PE (Trailing Twelve Months, TTM) data from the CSI 300 and S&P 500 indexes for the period from 2012 to 2023, totalling over 2500 observations.¹ We choose to analyse the PE ratio because it effectively captures deviations of asset prices from their fundamentals and reveals nonlinear dynamics influenced by market sentiment (Coakley and Fuertes 2006). The PE ratios for the indexes are calculated by taking a market value-weighted average of the PE of the constituent stocks. We obtain the demeaned PE ratio process by subtracting the rolling annual average. This process highlights fluctuations and potential bubble characteristics, as illustrated in Figure 3.

We test for unit roots on the demeaned CSI 300 PE and S&P 500 PE processes by using the Phillips–Perron test (Phillips and Perron 1988), which accommodates serial correlation and heteroskedasticity in the errors.² The results show a test statistic of -2.99 (p value: 0.04) for CSI 300 PE and a statistic of -2.76 (p value: 0.07) for S&P 500 PE. For both of them, we reject the unit root null hypothesis at the 10% significance level. These results provide weak evidence for the stationarity of both processes after demeaning.

We compare the MAR model with our proposed noncausal AR-ARCH process to determine which one better fits the data. The analysis aims to provide insights into the underlying dynamics of stock market and their potential noncausal structures. To identify the appropriate order of NAR models, a two-step procedure from Lanne and Saikkonen (2011) is adopted. Firstly, AR(p) models are fitted to the PE processes to identify the order p that minimises the BIC. The results indicate that the order $p = 1$ for the PE of the CSI 300, and $p = 2$ for the PE of the S&P 500. Secondly, a grid search is performed among all possible MAR models, where $r + s = p$, to find the model with the highest likelihood function. The analysis reveals that the MAR(0, 1) model is optimal for the CSI 300, while the MAR(0, 2) model is best suited for the S&P 500.

Table 4 presents the model selection results. In the case of the CSI 300 PE process, we observe that the NAR-NARCH model yields a higher log-likelihood, a lower BIC, and a smaller MSE compared to the MAR(0, 1) model, indicating that the NAR-NARCH model is more suitable for the PE process of the CSI 300. This contrasts with the S&P 500 PE, where the MAR(0,2) model exhibits a higher log-likelihood and a smaller BIC. Thus, although the PE processes for both the CSI 300 and S&P 500 are not stationary, the origins of their noncausal structures differ.

The estimated parameters and their significance, as shown in Table 5, highlight the differences. For the CSI 300 PE, all

TABLE 2 | Rejection rates of the null hypothesis.

N	Parameter	DGP			
		$\phi = 0.0, \alpha = 0.0$	$\phi = 0.0, \alpha = 0.5$	$\phi = 0.5, \alpha = 0.0$	$\phi = 0.5, \alpha = 0.5$
Panel A: ξ_t follows normal distribution					
100	T1	0.086	0.210	0.829	0.638
	T2	0.052	0.529	0.076	0.613
	T3	0.102	0.538	0.942	0.809
	T4	0.125	0.514	0.847	0.627
250	T1	0.048	0.041	0.967	0.941
	T2	0.015	0.814	0.081	0.817
	T3	0.036	0.818	0.994	0.966
	T4	0.063	0.806	0.969	0.812
500	T1	0.052	0.019	0.959	0.960
	T2	0.009	0.921	0.073	0.923
	T3	0.023	0.972	0.995	0.993
	T4	0.060	0.912	0.959	0.923
1000	T1	0.052	0.015	0.949	0.953
	T2	0.007	0.942	0.081	0.947
	T3	0.023	0.993	0.992	0.997
	T4	0.058	0.941	0.949	0.943
Panel B: ξ_t follows Student's t -distribution					
100	T1	0.080	0.240	0.805	0.641
	T2	0.052	0.531	0.118	0.619
	T3	0.096	0.543	0.933	0.820
	T4	0.116	0.520	0.834	0.648
250	T1	0.053	0.043	0.935	0.938
	T2	0.017	0.815	0.114	0.813
	T3	0.034	0.820	0.988	0.963
	T4	0.069	0.808	0.944	0.807
500	T1	0.049	0.019	0.968	0.959
	T2	0.007	0.914	0.119	0.921
	T3	0.024	0.918	0.992	0.992
	T4	0.056	0.912	0.971	0.929
1000	T1	0.048	0.015	0.984	0.941
	T2	0.006	0.942	0.130	0.939
	T3	0.021	0.993	0.994	0.997
	T4	0.054	0.940	0.985	0.943

parameters of the NAR-NARCH model are significant. In contrast, for the S&P 500 PE, the *p* values for the T1 and T3 tests are significant, while the *p* value for the T2 test is not. This suggests that the parameter α , which represents the coefficient of the forward-looking component in the innovation term, is not

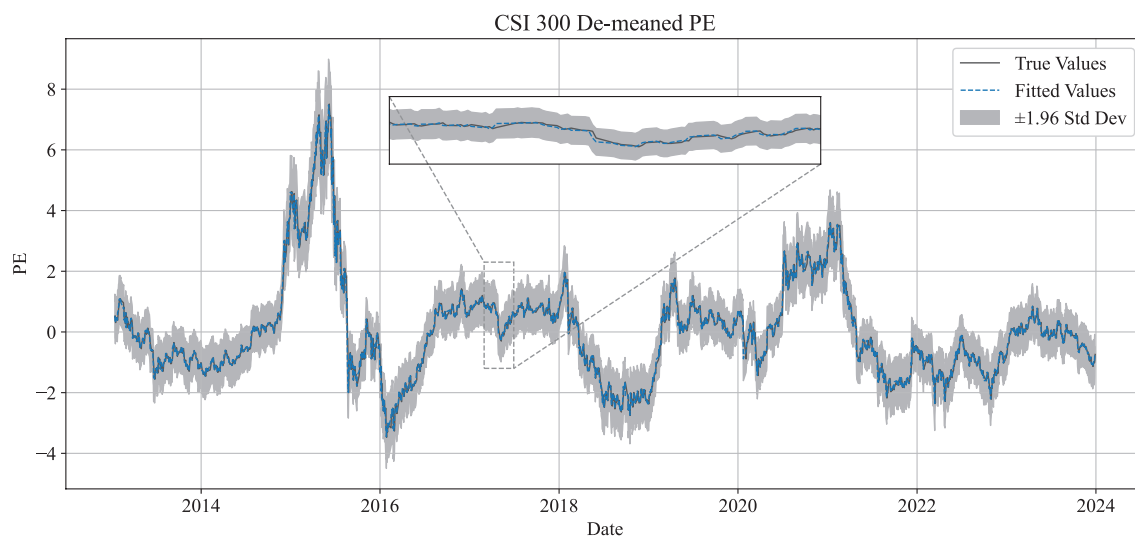
significantly different from zero, indicating the absence of a noncausal structure. Therefore, the noncausality in the CSI 300 PE is present in both the mean and variance of the process. In contrast, for the S&P 500, noncausality is limited to the mean, with the forward-looking component leading by two periods.

TABLE 3 | Simulation results on model selection by maximising the likelihood function.

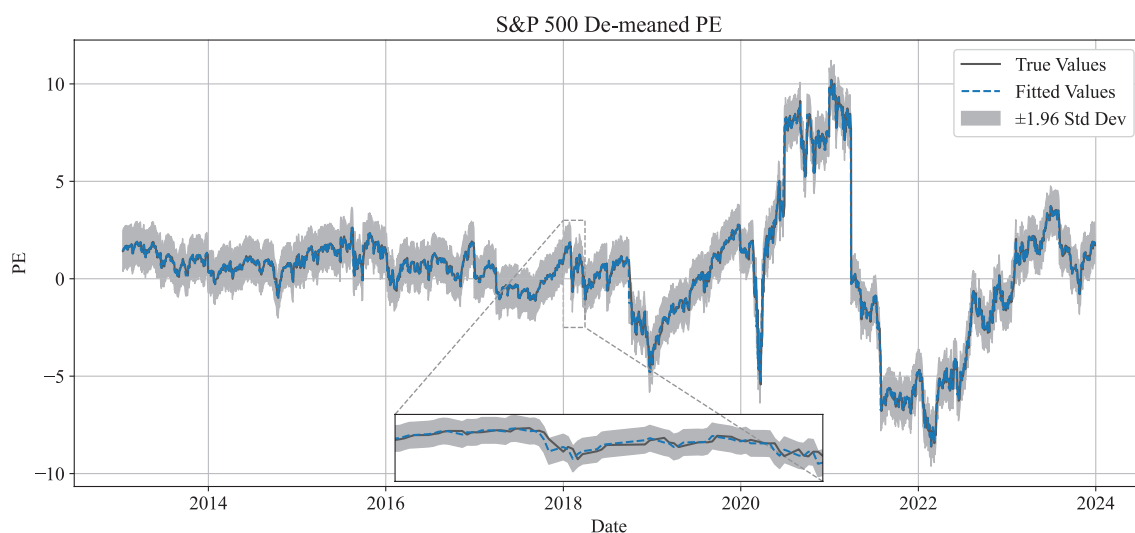
		DGP							
		$\phi = 0.3, \alpha = 0.3$		$\phi = 0.3, \alpha = 0.7$		$\phi = 0.7, \alpha = 0.3$		$\phi = 0.7, \alpha = 0.7$	
N	Model	Log-likelihood	BIC	Log-likelihood	BIC	Log-likelihood	BIC	Log-likelihood	BIC
Panel A: ξ_t follows normal distribution									
100	NAR-NARCH	-24	66	-80	177	-95	207	-141	318
	NAR	-110	229	-118	246	-117	244	-157	324
	ARMA	-118	254	-192	403	-197	412	-314	647
	AR-GARCH	-110	244	-173	370	-184	391	-256	534
250	NAR-NARCH	-87	195	-247	517	-310	641	-454	930
	NAR	-331	673	-394	799	-398	808	-494	999
	ARMA	-326	674	-473	968	-554	1130	-1091	2204
	AR-GARCH	-304	636	-406	840	-477	981	-921	1870
500	NAR-NARCH	-220	464	-669	1362	-575	1174	-1169	2362
	NAR	-706	1424	-870	1752	-822	1656	-1265	2543
	ARMA	-666	1357	-1152	2328	-1056	2137	-2220	4464
	AR-GARCH	-630	1292	-945	1920	-922	1876	-1461	2954
1000	NAR-NARCH	-386	799	-1398	2824	-1224	2476	-2261	4450
	NAR	-1432	2877	-1727	3468	-1752	3518	-2301	4615
	ARMA	-1338	2703	-2501	5029	-2455	4938	-4584	9195
	AR-GARCH	-1226	2487	-1943	3921	-1886	3806	-2900	5834
Panel B: ξ_t follows Student's t -distribution									
100	NAR-NARCH	-10	37	-78	174	-89	195	-152	301
	NAR	-105	219	-124	256	-129	267	-166	340
	ARMA	-113	244	-202	423	-181	381	-306	629
	AR-GARCH	-103	229	-164	351	-165	353	-245	512
250	NAR-NARCH	-93	207	-306	634	-272	566	-491	1003
	NAR	-325	661	-386	782	-395	801	-512	1035
	ARMA	-328	679	-586	1195	-539	1099	-1017	2057
	AR-GARCH	-304	636	-498	1023	-459	947	-740	1507
500	NAR-NARCH	-148	320	-599	1223	-624	1272	-1001	2047
	NAR	-690	1392	-829	1671	-869	1750	-1049	2110
	ARMA	-643	1310	-1189	2402	-1207	2439	-2004	4032
	AR-GARCH	-585	1201	-875	1780	-997	2025	-1517	3065
1000	NAR-NARCH	-384	795	-1295	2618	-1261	2549	-2296	4620
	NAR	-1422	2858	-1728	3469	-1738	3490	-2356	4725
	ARMA	-1345	2718	-2455	4937	-2313	4654	-4647	9322
	AR-GARCH	-1182	2398	-1822	3678	-1930	3894	-3003	6040

We further use the residuals for these two PE processes to obtain the Quantile–Quantile (Q–Q) plots as shown by Figure 4a,b. We can observe that the residuals of the CSI 300 PE deviate significantly from a normal distribution,

particularly in the tail regions. In contrast, the residuals of the S&P 500 PE are closer to a normal distribution, with smaller deviations in the tails. The distributional characteristics of the residuals also indirectly suggest differences in



(a) Demeaned CSI 300 PE



(b) Demeaned S&P 500 PE

FIGURE 3 | Demeaned PE and its fitted values. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

TABLE 4 | Model selection results for stock markets.

Index	Model	Log-likelihood	BIC	MSE
CSI 300 PE	NAR-NARCH	-7288.57	14608.70	0.03
	MAR(0,1)	-8652.50	17312.89	0.05
S&P 500 PE	NAR-NARCH	-3934.43	7900.41	0.10
	MAR(0,2)	-3876.26	7768.30	0.10

the variance of the innovations between the CSI 300 and the S&P 500.

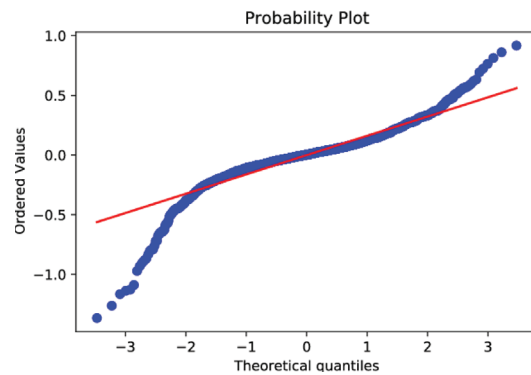
These findings suggest that while investors in both the Chinese and US stock markets are forward-looking and possess more

private information, the shocks in the Chinese market are not purely white noise; they also exhibit forward-looking behaviour. Conversely, shocks in the US market are primarily disturbances from noise traders. The differences in the noncausality of the PE in the two markets mainly stem from the investor structure. The Chinese stock market is predominantly composed of retail investors. However, some of these investors possess professional knowledge and extensive experience, allowing them to have a forward-looking perspective and access to more information. Meanwhile, retail investors are often influenced by sentiment and other market agents, which is not accounted for by the MAR model. In contrast, the US stock market is primarily made up of institutional investors with longer investment horizons, which is why there is an additional leading term in the noncausal structure.

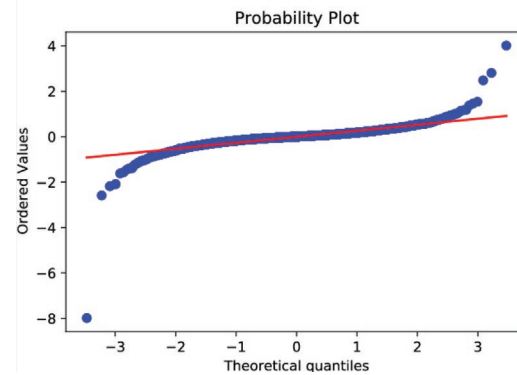
We now provide a more detailed analysis of the parameters of the NAR-NARCH models for the two PE processes. The parameter ϕ represents the relationship between the asset price and its future expectations. A higher ϕ implies a stronger relationship.

TABLE 5 | Parameter estimation for PE.

Index	ϕ	ω	α	σ_u^2	$T1\ p$	$T2\ p$	$T3\ p$	$T4\ p$
CSI 300	0.995	0.011	0.075	0.015	0.00	0.00	0.00	0.00
S&P 500	0.992	0.195	-0.079	0.043	0.00	0.23	0.00	0.00



(a) CSI 300



(b) S&P 500

FIGURE 4 | Quantile–Quantile plot of residuals of the PE. The red line represents the standard normal distribution. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

The parameter measures the correlation between the variance and the future price, reflecting how future price expectations influence present volatility. This also indicates the nature of the market structure, where trading may either be random or forward-looking. The ϕ coefficients of the two models are quite similar, yet there is a significant disparity in the ω values. The ω parameter represents the magnitude of non-fundamental shocks in the noncausal process that are unrelated to forward-looking behaviour. For the CSI 300 PE, the ω value is 0.011, whereas for the S&P 500 PE, it is 10 times larger at 0.195. This suggests that the magnitude of non-fundamental shocks in the S&P 500 PE is less associated with the forward-looking activities of market participants, which is consistent with the insignificant result for the α coefficient. The estimated α coefficients for the two PE processes are of opposite signs, indicating a different relationship between the variance of the innovation term and P_{t+1} . Based on the partial derivatives (Equation (8)),

$$\frac{\partial h_{t+1}}{\partial P_{t+1}} = 2\alpha^2 P_{t+1} + 2\omega\alpha, \quad (27)$$

and by substituting the estimated values, it can be observed that for the CSI 300 index, the variance h_{t+1} increases with PE when the PE is greater than -0.15. Thus, for the Chinese stock market, if market participants anticipate a rise in future prices, their trading will cause greater disturbances in the price, making it less stable. This can be attributed to heightened market sentiment as asset prices increase, which leads to higher price volatility. In contrast, for the US stock market, α is negative. Based on the partial derivatives, when the PE ratio is less than 2.46, higher PE implies lower volatility. However, when the PE ratio exceeds 2.46, volatility increases with higher PE. This can be explained by the fact that when prices are low, investors tend to be more cautious as prices increase. However, as prices rise further, they become more emotional, which induces higher volatility.

Figure 3 shows the fitted values from the NAR-NARCH model under the normal distribution assumption and their corresponding confidence intervals for the stock markets. The fitted values are close to the true values, and the confidence intervals encompass the volatility of the true values, even during rapid price drops and recoveries in the CSI 300 PE process. For the S&P 500 PE, the fitted values show a slight deviation from the true values during periods of severe fluctuation, but the true values remain close to the confidence intervals. This can be attributed to the fact that the most suitable model for the S&P 500 is MAR(0, 2).

We apply a rolling window estimation method with a 250-day window width to handle the parameters that might be time-varying. We focus on the noncausal parameters ϕ and α , and Figure 5 presents the estimation results for the CSI 300 Index. ϕ remains stable from 2014 to 2019. In 2020, its value exhibits large volatility and drops significantly, before recovering again. As for α , it also experiences strong volatility in 2020, which can be attributed to the pandemic shock to the stock market. In the whole sample, α is consistently higher than zero, indicating a strong noncausal variance, with the lowest value around 0.10. The mean of the estimated parameters for the entire time period shows that the mean of ϕ is approximately 0.995, and α is around 0.075, which is consistent with our estimation based on the whole sample.

4.2 | Crude Oil Futures Markets

Crude oil is an important industry resource, and it also serves as a financial asset offering valuable risk management tools. Its price is determined not only by industry supply and demand but also by investor trading, which is shaped by their expectations of future market trends. Therefore, investors in the crude oil futures market can more effectively discern changes in market supply

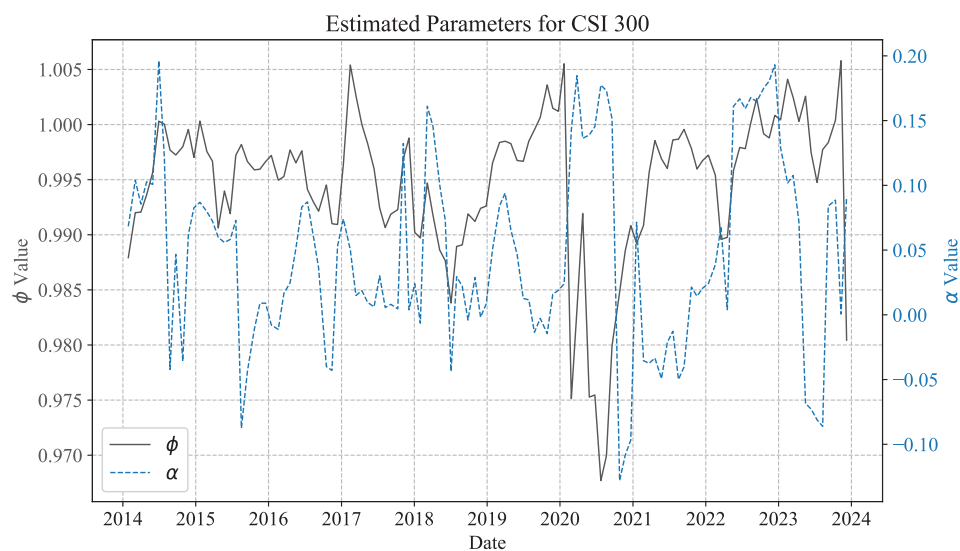
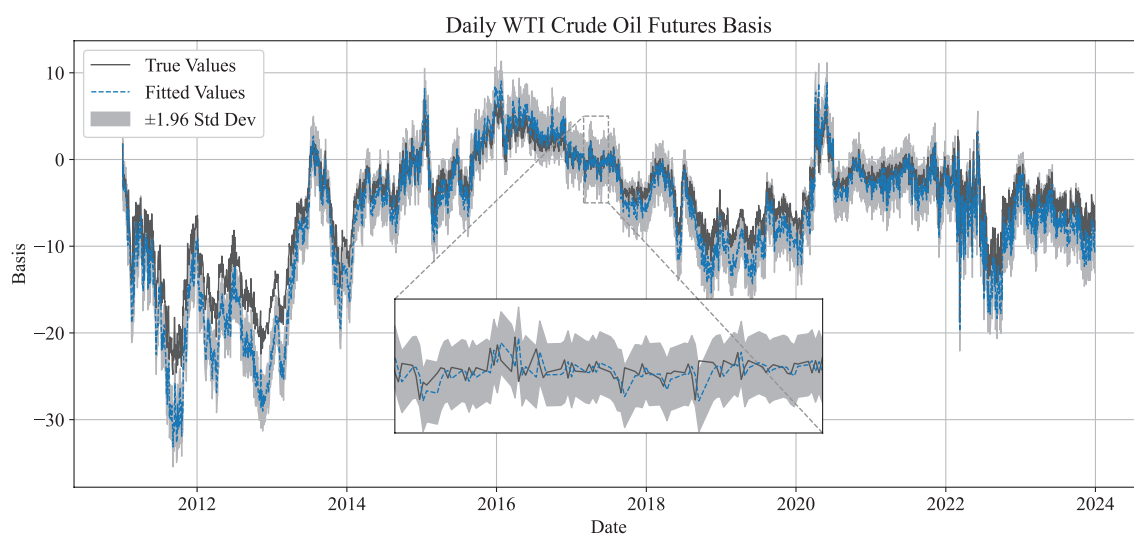
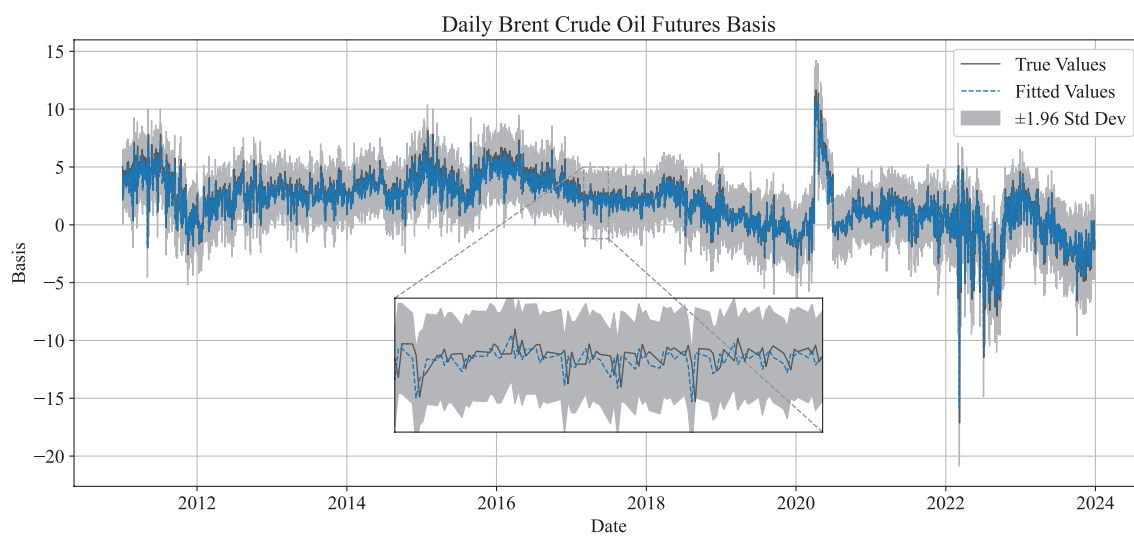


FIGURE 5 | Rolling Window estimation for CSI 300 PE. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]



(a) Daily WTI Crude Oil Futures Basis



(b) Daily Brent Crude Oil Futures Basis

FIGURE 6 | Detrended crude oil prices and its fitted values. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

and demand, gain access to broader industry information, and engage in more forward-looking trading behaviour. This is particularly important given the long cycle time associated with the extraction, transportation, and sale of crude oil. We apply our model to the crude oil futures market and conduct an empirical analysis using daily data on the basis of WTI and Brent crude oil futures from 2012 to 2023. The basis is calculated by subtracting the OPEC basket crude oil price from the settlement price of crude oil futures. Figure 6 shows the crude futures basis.

We also test unit roots for the basis processes of the WTI and Brent crude oil futures by the Phillips–Perron test. The test statistics are -6.31 (p value: 0) for WTI basis and -26.55 (p value: 0) for Brent basis.³ At the 1% significant, we reject the unit root null hypothesis. Thus, there is strong evidence of stationarity for the two processes.

The results of model selection for crude oil futures basis are presented in Table 6. For WTI, the MAR(4,0) model is found to be more suitable for the data, exhibiting a larger log likelihood, a smaller BIC, and a lower MSE. Additionally, the absence of a lead term in the MAR model implies that the data does not

TABLE 6 | Model selection results for crude oil futures.

Futures	Model	Log-likelihood	BIC	MSE
WTI	NAR-NARCH	-5958.30	11949.03	2.25
	MAR(4,0)	-5364.45	10761.33	2.13
Brent	NAR-NARCH	-5147.81	10328.08	1.98
	MAR(6,1)	-5233.66	10524.11	2.03

TABLE 7 | Parameter estimation for crude oil futures.

Price	ϕ	ω	α	σ_u^2	T1 p	T2 p	T3 p	T4 p
WTI	0.979	0.160	-0.064	1.897	0.00	0.00	0.00	0.00
Brent	0.926	0.871	-0.322	1.269	0.00	0.00	0.00	0.00

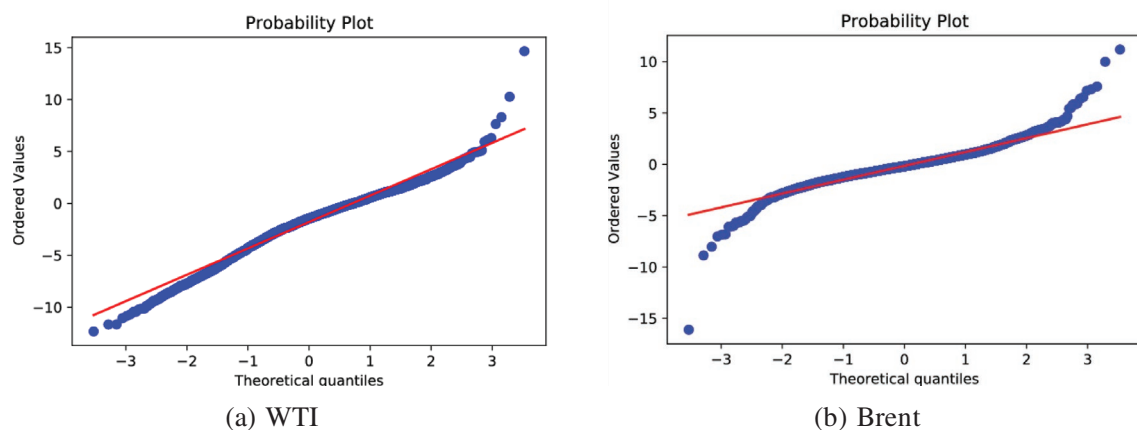


FIGURE 7 | Quantile–Quantile plot of residuals of future basis. The red line represents the standard normal distribution. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

exhibit a noncausal process. In contrast, for Brent, the NAR-NARCH model is more appropriate, and the most fitting MAR model is MAR(6,1), which includes one lead term. This indicates that the Brent basis data exhibits a noncausal structure in both its mean and variance.

We also obtain the estimated parameters for the WTI and Brent basis on the NAR-NARCH model as shown by Table 7. The ϕ parameter is larger for WTI, indicating a stronger correlation between present prices and future expectations. This implies that mainstream investors are more inclined to trade based on forward-looking information. Additionally, the α parameters for both WTI and Brent are negative. This suggests that when the basis is lower than 2.5 for WTI and 2.7 for Brent, a higher basis is associated with lower volatility, as investors trade more cautiously. However, when the basis is much higher, market participants become more emotional, leading to higher volatility.

Figure 6 shows the NAR-NARCH model fitted values under the normal distribution assumption and their corresponding confidence intervals for the crude oil markets. We observe that the fitted values exhibit smaller deviations from the true values for Brent than for WTI when using the NAR-NARCH model. This suggests that the NAR-NARCH model is better suited for fitting Brent data, while the MAR(4,0) model is more appropriate for WTI. However, it is evident that our fitted values capture the dynamic patterns well, and the confidence intervals encompass the spikes in the true values.

Figure 7 presents the Q–Q plots of the residuals for the WTI's MAR(4,0) model and the Brent's NAR-NARCH model. It can be observed that the residuals of the WTI closely approximate a normal distribution, with the exception of outliers in the tail region. However, the residuals for Brent deviate significantly from the normal distribution in the tail areas. The characteristics of

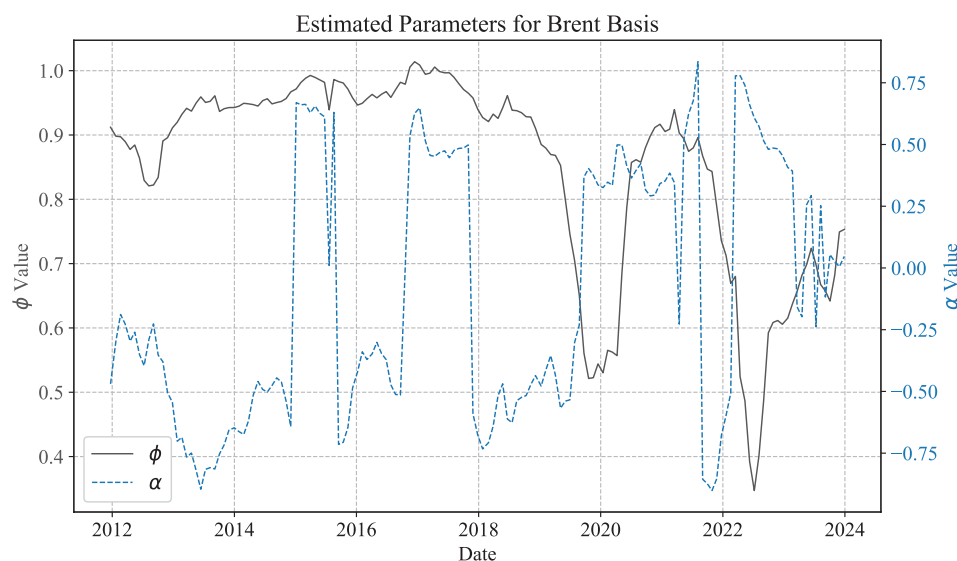


FIGURE 8 | Rolling Window estimation for Brent Basis. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/ijfe.3171)]

the residuals suggest that the WTI futures basis should follow an AR process, while the Brent futures basis exhibits noncausal features. These noncausal features imply that investor traders in the Brent crude oil futures market have access to a larger information set, likely due to the Brent crude oil futures market's closer association with the Organization of the Petroleum Exporting Countries (OPEC).

Figure 8 presents the rolling window estimation of parameters ϕ and α for the crude oil markets. Before 2020, ϕ is stable, and after that, its value fluctuates. It drops significantly from 0.9 to 0.5 in 2020, then recovers in 2021, and drops again in 2022, reaching 0.4, indicating weaker autocorrelations. Additionally, α is more stable before 2020, with a few periods of increase, turning positive in 2015 and 2017, while remaining negative during other periods. After 2020, α becomes positive and exhibits more volatility. This change reflects shifts in investor trading behaviour in the crude oil markets. The future price of oil used to have a negative influence on variance before 2020, but after 2020, it positively influences variance. This shift can be attributed to the COVID-19 pandemic shock, which caused oil prices to decrease, even becoming negative. Furthermore, the work-from-home trend reshaped oil demand and supply, which further influenced crude oil investor behaviour. As for the stable levels of the parameters, ϕ is around 0.92 and α is around -0.3 , which aligns with the estimation results based on the whole sample.

5 | Conclusion

We introduce the noncausality to the variance component and propose the noncausal AR-ARCH model. In this way, we extend the NAR model by considering the shocks related to future prices. This extension is particularly important for capturing the anticipatory patterns of financial time series data. Additionally, we provide the asymptotic properties of estimators and hypothesis testing methods for this model. Through comprehensive simulation analyses, we validate the proposed QMLE method

and show good control size and high power for the three hypothesis tests.

We present two empirical applications of our proposed model. Firstly, we apply the NAR-NARCH model to the US and Chinese stock markets, revealing distinct behavioural patterns. Notably, our analysis indicates that the presence of shocks related to future prices is more pronounced in the Chinese market, leading to a noncausal structure in the variance. In contrast, the US market, while lacking noncausality in variance, exhibits higher lead orders, suggesting stronger forward-looking behaviour by rational participants. Secondly, we utilise our model to analyse the subtleties in the behaviour of two closely related assets: WTI and Brent crude oil futures basis. Our findings offer valuable insights into the nuanced differences between these assets, which may not be immediately evident through traditional analysis. Specifically, we observe that the WTI market tends to display causal characteristics, whereas the Brent market exhibits a higher degree of noncausality in both the mean and variance, indicating a more complex interplay of factors influencing its price dynamics.

Acknowledgements

The authors would like to thank the two anonymous referees for their constructive suggestions, which have significantly contributed to improving the quality of the paper. Yaosong Zhan's research was supported by the National Natural Science Foundation of China (grant number 72403260). Shiqing Ling's research was partially supported by Hong Kong Research Grants Commission Grants (grant numbers 16303118, 16301620, 16300621, 16500522 and SRF522236S02).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

- ¹ PE (T/M) is the trailing twelve months Price-to-Earnings ratio, which is calculated by dividing the current market price of a stock by its average earnings per share over the trailing twelve months.
- ² The bandwidth of the variance estimator of Newey and West (1987) is set to 28, based on the criterion $\lfloor (12(T/100)^{1/4}) \rfloor$ proposed by Schwert (1989).
- ³ The bandwidth of the variance estimator of Newey and West (1987) is set to 29, based on the criterion $\lfloor (12(T/100)^{1/4}) \rfloor$ proposed by Schwert (1989).

References

- Aase, K. K., T. Bjuland, and B. Øksendal. 2012. "Partially Informed Noise Traders." *Mathematics and Financial Economics* 6: 93–104.
- Alessi, L., M. Barigozzi, and M. Capasso. 2011. "Non-Fundamentalness in Structural Econometric Models: A Review." *International Statistical Review* 79, no. 1: 16–47.
- Amemiya, T. 1985. *Advanced Econometrics*. Harvard University Press.
- Berkes, I., L. Horváth, and S. Ling. 2009. "Estimation in Nonstationary Random Coefficient Autoregressive Models." *Journal of Time Series Analysis* 30, no. 4: 395–416.
- Bottazzi, G., G. Devetag, and F. Pancotto. 2011. "Does Volatility Matter? Expectations of Price Return and Variability in an Asset Pricing Experiment." *Journal of Economic Behavior & Organization* 77, no. 2: 124–146.
- Breid, F. J., R. A. Davis, K.-S. Lh, and M. Rosenblatt. 1991. "Maximum Likelihood Estimation for Noncausal Autoregressive Processes." *Journal of Multivariate Analysis* 36, no. 2: 175–198.
- Brown, G. W. 1999. "Volatility, Sentiment, and Noise Traders." *Financial Analysts Journal* 55, no. 2: 82–90.
- Casella, G., and R. Berger. 2024. *Statistical Inference*. CRC Press.
- Cavaliere, G., H. B. Nielsen, and A. Rahbek. 2020. "Bootstrapping Noncausal Autoregressions: With Applications to Explosive Bubble Modeling." *Journal of Business & Economic Statistics* 38, no. 1: 55–67.
- Coakley, J., and A.-M. Fuertes. 2006. "Valuation Ratios and Price Deviations From Fundamentals." *Journal of Banking & Finance* 30, no. 8: 2325–2346.
- Fries, S. 2021. "Conditional Moments of Noncausal Alpha-Stable Processes and the Prediction of Bubble Crash Odds." *Journal of Business & Economic Statistics* 40, no. 4: 1–21.
- Fries, S., and J.-M. Zakoian. 2019. "Mixed Causal-Noncausal AR Processes and the Modelling of Explosive Bubbles." *Econometric Theory* 35, no. 6: 1234–1270.
- Gouriéroux, C., and J. Jasiak. 2016. "Filtering, Prediction and Simulation Methods for Noncausal Processes." *Journal of Time Series Analysis* 37, no. 3: 405–430.
- Gouriéroux, C., and J. Jasiak. 2018. "Misspecification of Noncausal Order in Autoregressive Processes." *Journal of Econometrics* 205, no. 1: 226–248.
- Gouriéroux, C., J. Jasiak, and A. Monfort. 2020. "Stationary Bubble Equilibria in Rational Expectation Models." *Journal of Econometrics* 218, no. 2: 714–735.
- Gouriéroux, C., and J.-M. Zakoian. 2017. "Local Explosion Modelling by Non-Causal Process." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 79, no. 3: 737–756.
- Gouriéroux, C., and J.-M. Zakoian. 2013. *Explosive Bubble Modelling by Noncausal Process*. CREST.
- Hansen, L. P., and T. J. Sargent. 1991. "Two Difficulties in Interpreting Vector Autoregressions." *Rational Expectations Econometrics* 1: 77–119.
- Hecq, A., S. Telg, and L. Lieb. 2017. "Do Seasonal Adjustments Induce Noncausal Dynamics in Inflation Rates?" *Econometrics* 5, no. 4: 48.
- Hecq, A., and E. Voisin. 2021. "Forecasting Bubbles With Mixed Causal-Noncausal Autoregressive Models." *Econometrics and Statistics* 20: 29–45.
- Hencic, A., and C. Gouriéroux. 2015. "Noncausal Autoregressive Model in Application to Bitcoin/Usd Exchange Rates." In *Econometrics of Risk*, 17–40. Springer.
- Huang, J., and Y. Pawitan. 2000. "Quasi-Likelihood Estimation of Non-Invertible Moving Average Processes." *Scandinavian Journal of Statistics* 27, no. 4: 689–702.
- Kyle, A. S. 1985. "Continuous Auctions and Insider Trading." *Econometrica* 53, no. 6: 1315–1335.
- Lanne, M., J. Luoto, and P. Saikkonen. 2012. "Optimal Forecasting of Noncausal Autoregressive Time Series." *International Journal of Forecasting* 28, no. 3: 623–631.
- Lanne, M., and P. Saikkonen. 2011. "Noncausal Autoregressions for Economic Time Series." *Journal of Time Series Econometrics* 3, no. 3: 1–32.
- Lii, K.-S., and M. Rosenblatt. 1996. "Maximum Likelihood Estimation for Nongaussian Nonminimum Phase ARMA Sequences." *Statistica Sinica* 6, no. 1: 1–22.
- Ling, S. 2004. "Estimation and Testing Stationarity for Double-Autoregressive Models." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 66, no. 1: 63–78.
- Ling, S. 2005. "Self-Weighted Least Absolute Deviation Estimation for Infinite Variance Autoregressive Models." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 67, no. 3: 381–393.
- Ling, S. 2007. "Self-Weighted and Local Quasi-Maximum Likelihood Estimators for Arma-Garch/Igarch Models." *Journal of Econometrics* 140, no. 2: 849–873.
- Lof, M. 2013. "Noncausality and Asset Pricing." *Studies in Nonlinear Dynamics and Econometrics* 17, no. 2: 211–220. <https://doi.org/10.1515/snde-2012-0035>.
- Lof, M., and H. Nyberg. 2017. "Noncausality and the Commodity Currency Hypothesis." *Energy Economics* 65: 424–433.
- Newey, W. K., and K. D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelationconsistent Covariance Matrix." *Econometrica* 55: 703–708.
- Nyberg, H., M. Lanne, and E. Saarinen. 2012. "Does Noncausality Help in Forecasting Economic Time Series?" *Economics Bulletin* 32, no. 4: 2849–2859.
- Phillips, P. C., and P. Perron. 1988. "Testing for a Unit Root in Time Series Regression." *Biometrika* 75, no. 2: 335–346.
- Schwert, G. W. 1989. "Tests for Unit Roots: A Monte Carlo Investigation." *Journal of Business & Economic Statistics* 7, no. 2: 5–17.
- Wu, R., and R. A. Davis. 2010. "Least Absolute Deviation Estimation for General Autoregressive Moving Average Time-Series Models." *Journal of Time Series Analysis* 31, no. 2: 98–112.

Appendix A

Proof of Theorem 2.1 and 2.2

Proof of Theorem 2.1

Proof. For any given m , we can rewrite Equation (7) as follows:

$$P_t = (\phi + \alpha \eta_t) P_{t+1} + \bar{\xi}_t$$

$$= \sum_{i=0}^m \prod_{j=0}^i (\phi + \alpha \eta_{t+j}) \bar{\xi}_{t+i} + \sum_{j=0}^{m+1} (\phi + \alpha \eta_{t+j}) P_{t+m+2}. \quad (A1)$$

Let

$$S_{m,t} = \sum_{i=0}^m \prod_{j=0}^i (\phi + \alpha \eta_{t+j}) \bar{\xi}_{t+j+1}. \quad (\text{A2})$$

As the Proof of Lemma A.1 in Ling (2005), if $\gamma = \mathbb{E} \log |\phi + \alpha \eta_t| < 0$, there exists a constant $u \in (0, 2)$, such that

$$P \equiv \mathbb{E} |\phi + \alpha \eta_t|^u < 1. \quad (\text{A3})$$

For any $m, n > 0$, we have

$$\begin{aligned} \mathbb{E} |S_{m+n,t} - S_{m,t}|^u &= \mathbb{E} \left| \sum_{i=m+1}^{m+n} \prod_{j=0}^i (\phi + \alpha \eta_{t+j}) \bar{\xi}_{t+j+1} \right|^u \\ &\leq \mathbb{E} |\bar{\xi}_{t+1}|^u \sum_{i=m+1}^{m+n} \prod_{j=0}^i \mathbb{E} |\phi + \alpha \eta_{t+j}|^u \\ &= \mathcal{O}(P^n). \end{aligned} \quad (\text{A4})$$

Thus, by the Cauchy criterion, we can show that

$$S_{m,t} \rightarrow S_{\infty,t} \equiv P_t, \text{ a.s. as } m \rightarrow \infty. \quad (\text{A5})$$

Thus, P_t has the following representation:

$$P_t = \sum_{i=0}^{\infty} \prod_{j=0}^i (\phi + \alpha \eta_{t+j}) \bar{\xi}_{t+j+1}. \quad (\text{A6})$$

It is easy to verify that P_t in Equation (A6) is a solution of Equation (7).

Note that P_t is a nonlinear function of $\{\bar{\xi}_t, \bar{\xi}_{t+1}, \dots\}$. It is not hard to see that P_t and P_t^* have the same distribution and thus, (ii) holds obviously.

Suppose that \tilde{P}_t is another solution of Equation (7) with respect to $\{\tilde{\xi}_t, \tilde{\xi}_{t+1}, \dots\}$. Then

$$\tilde{P}_t = G(\tilde{\xi}_t, \tilde{\xi}_{t+1}, \dots). \quad (\text{A7})$$

Denote $\tilde{P}_t^* = G(\tilde{\xi}_t, \tilde{\xi}_{t-1}, \tilde{\xi}_{t-2}, \dots)$. Then

$$\begin{aligned} \Delta P_t &\stackrel{\text{def}}{=} P_t - \tilde{P}_t \\ &= (\phi + \alpha \eta_{t+1}) \Delta P_{t-1} = \prod_{i=1}^m (\phi + \alpha \eta_{t+i}) \Delta P_{t-m}. \end{aligned} \quad (\text{A8})$$

Furthermore,

$$\mathbb{E} |\Delta P_t|^u \leq \prod_{i=1}^m \mathbb{E} |\phi + \alpha \eta_{t+i}|^u \mathbb{E} |\Delta P_{t-m}|^u = \rho^m \mathbb{E} |\Delta P_{t-m}^*|^u \rightarrow 0 \quad (\text{A9})$$

as $m \rightarrow \infty$, where $\mathbb{E} |\Delta P_{t-m}^*|^u$ is a constant since P_t^* and \tilde{P}_t^* are stationary.

Thus, $\Delta P_t = 0$ a.s. That is, $P_t = P_t^*$ a.s. When $\gamma = \mathbb{E} \log |\phi + \alpha \eta_t| \geq 0$, we conduct recursion on Equation (7) with an initial value P_0 . Similar to Lemma 4.1 in Berkes et al. (2009), we can show that

$$|P_t| \xrightarrow{P} \infty \text{ as } t \rightarrow \infty. \quad (\text{A10})$$

Thus, there is not a solution to Equation (7) in this case. \square

Proof of Theorem 2.2

Lemma A1–A6

Before proving Theorem 2.2, we need to provide Lemma A1–A6. Define

$$\ell_t^*(\theta) = -\log \left[\sigma_u^2 + (\omega + \alpha P_{t-1}^*)^2 \right] + \frac{(P_t^* - \phi P_{t-1}^*)^2}{\sigma_u^2 + (\omega + \alpha P_{t-1}^*)^2}, \quad (\text{A11})$$

and similarly define $\frac{\partial \ell_t^*(\theta)}{\partial \phi}$, $\frac{\partial \ell_t^*(\theta)}{\partial \lambda}$, $\frac{\partial^2 \ell_t^*(\theta)}{\partial \phi^2}$, $\frac{\partial^2 \ell_t^*(\theta)}{\partial \lambda \partial \lambda'}$, and $\frac{\partial^2 \ell_t^*(\theta)}{\partial \lambda \partial \phi}$ with P_t replaced by P_t^* and P_{t+1} replaced by P_{t+1}^* . Denote

$$\mathcal{T}_t(m) = \sigma \{ (\eta_t, \xi_{t-1}), \dots, (\eta_{t-m}, \xi_{t-m-1}) \} \quad (\text{A12})$$

We can have Lemma A1–A6.

Lemma A1. Under the assumptions of Theorem 2.2

- $\mathbb{E} [\ell_t^*(\theta) - \mathbb{E} [\ell_t^*(\theta) | \mathcal{T}_t(m)]]^2 = \mathcal{O}(p^m)$;
- $\mathbb{E} \left\| \frac{\partial \ell_t^*(\theta)}{\partial \theta} - \mathbb{E} \left[\frac{\partial \ell_t^*(\theta)}{\partial \theta} | \mathcal{T}_t(m) \right] \right\|^2 = \mathcal{O}(p^m)$;
- $\mathbb{E} \left\| \frac{\partial^2 \ell_t^*(\theta)}{\partial \theta \partial \theta'} - \mathbb{E} \left[\frac{\partial^2 \ell_t^*(\theta)}{\partial \theta \partial \theta'} | \mathcal{T}_t(m) \right] \right\|^2 = \mathcal{O}(p^m)$.

Lemma A2. Under the conditions of Theorem 2.2 and $\mathbb{E} \sup_{\theta} |\ell_t^*(\theta)| < \infty$

$$\sup_{\theta} \left| \frac{1}{n} \mathcal{L}_n \theta - \mathbb{E} \ell_t^*(\theta) \right| = o_p(1), \quad (\text{A13})$$

where $o_p(1)$ converges to zero in probability as $n \rightarrow \infty$.

Lemma A3. Under the condition of Theorem 2.2, $\mathbb{E} \ell_t^*(\theta)$ has a unique maximum at θ_0 .

Lemma A4. Under the condition of Theorem 2.2, it follows that

- $\mathbb{E} \sup_{\theta} \left\| \frac{\partial \ell_t^*(\theta)}{\partial \theta} \frac{\partial \ell_t^*(\theta)}{\partial \theta'} \right\| < \infty$;
- $\sup_{\theta} \left\| \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \ell_t^*(\theta)}{\partial \theta} \frac{\partial \ell_t^*(\theta)}{\partial \theta'} \right] - \mathbb{E} \left[\frac{\partial \ell_t^*(\theta)}{\partial \theta} \frac{\partial \ell_t^*(\theta)}{\partial \theta'} \right] \right\| = o_p(1)$.

Lemma A5. Under the condition of Theorem 2.2, it follows that

- $\mathbb{E} \left[\frac{\partial \ell_t^*(\theta_0)}{\partial \theta} \frac{\partial \ell_t^*(\theta)}{\partial \theta'} \right]_{\theta=\theta_0} = \Omega$;
- $\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \ell_t^*(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \xrightarrow{D} \mathcal{N}(0, \Omega)$.

Lemma A6. Under the condition of Theorem 2.2, it follows that

- $\mathbb{E} \sup_{\theta} \left\| \frac{\partial^2 \ell_t^*(\theta)}{\partial \theta \partial \theta'} \right\| < \infty$;
- $\sup_{\theta} \left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 \ell_t^*(\theta)}{\partial \theta \partial \theta'} - \mathbb{E} \frac{\partial^2 \ell_t^*(\theta)}{\partial \theta \partial \theta'} \right\| = o_p(1)$.

Proof of Lemma A1–A6 and Theorem 2.2

Proof of Lemma A1. Let

$$K_t(m) = \sum_{i=1}^m \sum_{j=0}^i (\phi + \alpha \eta_{t-j}) \bar{\xi}_{t-i}^* + \bar{\xi}_t^*. \quad (\text{A14})$$

Then

$$P_t^* = (\phi + \alpha \eta_t) P_{t-1}^* + \bar{\xi}_t^* = K_t(m) + \prod_{j=0}^m (\phi + \alpha \eta_{t-j}) P_{t-m-1}. \quad (\text{A15})$$

Thus, by Theorem 2.1, there exists a $\delta \in (0, 1)$ such that

$$\mathbb{E}|P_t^* - K_t(m)|^\delta = \left[\mathbb{E}|\phi + \alpha\eta_{t-j}|^\delta \right]^{m+1} \mathbb{E}|P_{t-m-1}|^\delta = \mathcal{O}(p^m). \quad (\text{A16})$$

By Equation (A16)

$$\begin{aligned} & \mathbb{E} \left| \frac{(P_{t-1}^*)^2}{\sigma_u^2 + (\omega + \alpha P_{t-1}^*)^2} - \frac{K_t^2(m)}{\sigma_u^2 + [\omega + \alpha K_t(m)]^2} \right|^2 \\ & \leq C \mathbb{E} \left| \frac{(P_{t-1}^*)^2}{\sigma_u^2 + (\omega + \alpha P_{t-1}^*)^2} - \frac{K_t^2(m)}{\sigma_u^2 + [\omega + \alpha K_t(m)]^2} \right|^{\delta/4} \\ & \leq C \mathbb{E} \left| (P_{t-1}^*)^2 - K_t^2(m) \right|^{\delta/4} + C \mathbb{E} \left[\frac{K_t^{\delta/2}(m)}{(\omega + \alpha P_{t-1}^*)^2 - [\omega + \alpha K_t(m)]^2} \right]^{\delta/4} \\ & = \mathcal{O}(p^m). \end{aligned} \quad (\text{A17})$$

By Equation (A17)

$$\begin{aligned} & \mathbb{E} \left| \frac{(P_{t-1}^*)^2}{\sigma_u^2 + (\omega + \alpha P_{t-1}^*)^2} - \mathbb{E} \left[\frac{(P_{t-1}^*)^2}{\sigma_u^2 + (\omega + \alpha P_{t-1}^*)^2} \middle| \mathcal{T}_t(m) \right] \right|^2 \\ & \leq 2 \mathbb{E} \left| \frac{(P_{t-1}^*)^2}{\sigma_u^2 + (\omega + \alpha P_{t-1}^*)^2} - \frac{K_t^2(m)}{\sigma_u^2 + [\omega + \alpha K_t(m)]^2} \right|^2 \\ & + 2 \mathbb{E} \left[\mathbb{E} \left| \frac{(P_{t-1}^*)^2}{\sigma_u^2 + (\omega + \alpha P_{t-1}^*)^2} - \frac{K_t^2(m)}{\sigma_u^2 + [\omega + \alpha K_t(m)]^2} \middle| \mathcal{T}_t(m) \right|^2 \right] \\ & \leq 4 \mathbb{E} \left| \frac{(P_{t-1}^*)^2}{\sigma_u^2 + (\omega + \alpha P_{t-1}^*)^2} - \frac{K_t^2(m)}{\sigma_u^2 + [\omega + \alpha K_t(m)]^2} \right|^2 \\ & = \mathcal{O}(p^m). \end{aligned} \quad (\text{A18})$$

Using Equation (A18) and note that $P_t^* = (\phi + \alpha\eta_{t-1})P_{t-1}^* + \xi_t^*$, it is straightforward to show that Lemma A1 (i) holds. Similarly, we can show that Lemma A1 (ii) and (iii) hold. \square

Proof of Lemma A2. By Theorem 2.1, there exists a $\delta \in (0, 1)$ such that $\mathbb{E}|P_t|^\delta$. Thus, we can show that

$$\mathbb{E} \sup (\omega + \alpha P_{t+1})^{\delta/2} < \infty. \quad (\text{A19})$$

Using Equation (A19) and as the proof of Equation (A.4) in Ling (2005), we can show that

$$\mathbb{E} \sup |\ell_t(\theta)| < \infty. \quad (\text{A20})$$

Then, for each n ,

$$\begin{aligned} \sum_{i=1}^n \ell_i(\theta) &= G(P_1, P_2, \dots, P_n, P_{n+1}) \\ &\stackrel{d}{=} G(P_{-1}^*, P_{-2}^*, \dots, P_{-n}^*, P_{-n-1}^*) \\ &= \sum_{i=-1}^{-n} \ell_i^*(\theta). \end{aligned} \quad (\text{A21})$$

Since P_t^* is a stationary and ergodic series, $\{\ell_t^*(\theta)\}$ is so for each θ . By Lemma A1, $\ell_t^*(\theta)$ is L^2 -NED. By the Backward Strong Law of Larger Number Theorem 2.1 in Ling (2007), we have

$$\left| \frac{1}{n} \sum_{i=-1}^{-n} \ell_i^*(\theta) - \mathbb{E} \ell_t^*(\theta) \right| = o_p(1), \quad (\text{A22})$$

and hence

$$\left| \frac{1}{n} \mathcal{L}_n(\theta) - \mathbb{E} \ell_t(\theta) \right| = o_p(1) \quad (\text{A23})$$

for each θ . Furthermore, by the point-wise argument, we can obtain Lemma A2. \square

Proof of Lemma A3.

$$\begin{aligned} \mathbb{E} \ell_t(\theta) &= -\mathbb{E} \left\{ \log \left[\sigma_u^2 + (\omega + \alpha P_{t+1})^2 \right] + \frac{(P_t - \phi P_{t+1})^2}{\sigma_u^2 + (\omega + \alpha P_{t+1})^2} \right\} \\ &= -\left\{ \mathbb{E} \log \left[\sigma_u^2 + (\omega + \alpha P_{t+1})^2 \right] + \mathbb{E} \left[\frac{(\sigma_{u0}^2 + (\omega_0 + \alpha_0 P_{t+1}^*)^2)}{\sigma_u^2 + (\omega + \alpha P_{t+1})^2} \right] \right\} \\ &\quad - (\phi - \phi_0)^2 \mathbb{E} \frac{P_{t+1}^2}{\sigma_u^2 + (\omega + \alpha P_{t+1})^2}. \end{aligned} \quad (\text{A24})$$

As the proof of Lemma A2 in Ling (2004), $\mathbb{E} \ell_t(\theta)$ achieves its maximum if and only if

$$\phi = \phi_0 \text{ and } \sigma_u^2 + (\omega + \alpha P_{t+1})^2 = \sigma_{u0}^2 + (\omega_0 + \alpha_0 P_{t+1})^2. \quad (\text{A25})$$

The latter implies that $\sigma_u^2 = \sigma_{u0}^2$, $\omega = \omega_0$, and $\alpha = \alpha_0$. \square

Proof of Lemma A4. By Lemma A3, we have

$$\begin{aligned} \mathbb{E} \sup_\theta \left[\frac{\partial \ell_t(\theta)}{\partial \phi} \right]^2 &= 4 \mathbb{E} \sup_\theta \frac{P_{t+1}^2 [u_{0,t} - (\phi - \phi_0) P_{t+1}]^2}{[\sigma_u^2 + (\omega + \alpha P_{t+1})^2]^2} \\ &\leq C \mathbb{E} \sup_\theta \frac{P_{t+1}^4 + P_{t+1}^2 u_{0,t}^2}{[\sigma_u^2 + (\omega + \alpha P_{t+1})^2]^2}, \end{aligned} \quad (\text{A26})$$

where $u_{0,t} = \omega_0 \eta_t + \alpha_0 \eta_t P_{t+1} + \phi_0 \xi_{t+1} + \alpha_0 \eta_t \xi_{t+1}$. Since (η_t, ξ_{t+1}) is independent of P_{t+1} , we have

$$\mathbb{E} \sup_\theta \left[\frac{\partial \ell_t(\theta)}{\partial \phi} \right]^2 \leq C \mathbb{E} \sup_\theta \frac{P_{t+1}^4}{[\sigma_u^2 + (\omega + \alpha P_{t+1})^2]^2}. \quad (\text{A27})$$

Note that

$$P_{t+1}^2 \leq \frac{1}{\alpha^2} [\alpha P_{t+1}]^2 \leq C \xi (\omega + \alpha P_{t+1})^2 + 1 \quad (\text{A28})$$

for some constant C . Thus,

$$\mathbb{E} \sup_\theta \left[\frac{\partial \ell_t(\theta)}{\partial \phi} \right]^2 \leq C \mathbb{E} \sup_\theta \left[\frac{(\omega + \alpha P_{t+1})^2 + 1}{\sigma_u^2 + (\omega + \alpha P_{t+1})^2} \right]^2 \leq C \left(1 + \frac{1}{\sigma_u} \right) < \infty. \quad (\text{A29})$$

Similarly, we can show that other terms in (i) hold.

As the Proof of Lemma A2

$$\sum_{i=1}^n \frac{\partial \ell_i(\theta)}{\partial \theta} \frac{\partial \ell_i(\theta)}{\partial \theta'} \stackrel{d}{=} \sum_{i=-1}^{-n} \frac{\partial \ell_i^*(\theta)}{\partial \theta} \frac{\partial \ell_i^*(\theta)}{\partial \theta'}. \quad (\text{A30})$$

By Lemma A1 (ii), we can show that $\frac{\partial \ell_t^*(\theta)}{\partial \theta} \frac{\partial \ell_t^*(\theta)}{\partial \theta'}$ is L^2 -NED with the decay rate p^m . By the Backward Strong Law of Large Number Theorem 2.1 in Ling (2007). We have

$$\left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial \ell_t^*(\theta)}{\partial \theta} \frac{\partial \ell_t^*(\theta)}{\partial \theta'} - \mathbb{E} \left[\frac{\partial \ell_t^*(\theta)}{\partial \theta} \frac{\partial \ell_t^*(\theta)}{\partial \theta'} \right] \right\| = o_p(1), \quad (\text{A31})$$

and hence

$$\left\| \frac{1}{n} \sum_{t=1}^n \frac{\partial \ell_t(\theta)}{\partial \theta} \frac{\partial \ell_t(\theta)}{\partial \theta'} - \mathbb{E} \left[\frac{\partial \ell_t(\theta)}{\partial \theta} \frac{\partial \ell_t(\theta)}{\partial \theta'} \right] \right\| = o_p(1). \quad (\text{A32})$$

Furthermore, by the point-wise argument, we can show that (ii) holds. \square

Proof of Lemma A5. (i) is obtained by a direct calculation.

To (ii), note that

$$\sum_{t=1}^n \frac{\partial \ell_t(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \stackrel{d}{=} \sum_{t=1}^n \frac{\partial \ell_t^*(\theta)}{\partial \theta} \Big|_{\theta=\theta_0}. \quad (\text{A33})$$

By Lemma A1 (iii), $\frac{\partial \ell_t^*(\theta)}{\partial \theta}$ is L^2 -NED with a decay rate (p^m) , and is a martingale different in terms of the σ -field generated by $\{(\eta_s, \xi_{s-1}) : s \leq t\}$. By the Backward Strong Invariance Principle Theorem 2.2 in Ling (2007), we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \ell_t^*(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \stackrel{D}{\rightarrow} \mathcal{N}(0, \Omega), \quad (\text{A34})$$

and hence, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \ell_t(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \stackrel{D}{\rightarrow} \mathcal{N}(0, \Omega). \quad (\text{A35})$$

\square

Proof of Lemma A6. The proof of (i) is similar to that of Lemma A4 (i). Using Lemma A1 (iii) and the Backward Strong Law of Large Number Theorem 2.1 in Ling (2007), the proof of (ii) is similar to that of Lemma A4 (ii). Thus, the details are omitted. \square

Proof of Theorem 2.2. First, the parameter space Θ is a compact set and θ_0 is an interior point in Θ . $\mathcal{L}_n(\theta)$ is a continuous function in $\theta \in \Theta$ and is measurable in terms of $\{P_1, P_2, \dots, P_n\}$. By Lemmas A2 and A3, we know that the conditions for consistency in Theorem 4.1.1 in Amemiya (1985) are satisfied and hence, (i) holds.

By Lemmas A4 and A5, we know that the conditions in Theorem 4.1.3 in Amemiya (1985) are satisfied and hence, (ii) holds.

Lemma A2 to A6 show that $\ell_t^*(\theta)$ and its first and second derivatives are near-epoch dependence (NED). \square