Empirical Likelihood Test For The Application of SWQMELE In Fitting An ARMA-GARCH Model

Mo Zhou*, Liang Peng† and Rongmao Zhang‡

Abstract

Fitting an ARMA-GARCH model has become a common practice in financial econometrics. Because the asymptotic normality of the quasi maximum likelihood estimation (QMLE) requires finite fourth moment for both errors and the sequence itself, self-weighted quasi maximum exponential likelihood estimation (SWQMELE) has been proposed to reduce the moment constraints but requires the errors to have zero median instead of zero mean. Because changing zero mean to zero median destroys the ARMA-GARCH structure and has a serious effect on skewed data, this paper proposes an efficient empirical likelihood test for zero mean of errors in the application of SWQMELE to ensure that the model still concerns conditional mean. A simulation study confirms the good finite sample performance before applying the test to the US housing price indexes and financial returns for the study of comovement.

Keywords: ARMA-GARCH model, empirical likelihood, quasi-maximum likelihood estimation, self-weighted quasi maximum exponential likelihood estimation.

*College of Mathematics, Zhejiang University
†Department of Risk Management and Insurance, Georgia State University
‡College of Mathematics, Zhejiang University. Corresponding author: rmzhang@zju.edu.cn

1 Introduction

It has become a practical technique to model heteroscedasticity of a financial/economic variable by an ARMA-GARCH sequence since Engle (1982) and Bollerslev (1986). A standard ARMA(p,q)-GARCH(r,s) model is defined as

$$\begin{cases} y_t = \varphi + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} + \varepsilon_t, \ \varepsilon_t = \eta_t \sqrt{h_t}, \\ h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j h_{t-j}, \end{cases} \tag{1.1}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$ $(i = 1, ..., r)$, $\beta_j \geq 0$ $(j = 1, ..., s)$, $\varphi \in \mathbb{R}$, $\phi_i \in \mathbb{R}$ $(i = 1, ..., p)$, $\psi_j \in \mathbb{R}$ $(j = 1, ..., q)$, and $\{\eta_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance one.

A commonly employed statistical inference for fitting model (1.1) is the so-called quasi-maximum likelihood estimation (QMLE); see Ling and Li (1997), Jeantheau (1998), Berkes et al. (2003), Ling and McAleer (2003), Hall and Yao (2003), and Francq and Zakoïan (2004). It is known that the QMLE has a normal limit when both $E\varepsilon_t^4 < \infty$ and $E\eta_t^4 < \infty$. In practice, $\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j$ is often close to one, suggesting that the assumption of $E\varepsilon_t^4 < \infty$ may be problematic. To reduce this moment constraint, Ling (2007) proposed a self-weighted local quasi maximum likelihood estimation, which has a normal limit when $E\eta_t^4 < \infty$ and $E|\varepsilon_t|^\tau < \infty$ for some $\tau > 0$. When the model (1.1) becomes a pure GARCH process, Hall and Yao (2003) showed that the QMLE has a stable law limit when $E\eta_t^4 = \infty$. Without surprising, one has to use a different estimation technique based on some condition different from $E\eta_t^2 = 1$ to ensure a normal limit in the case of $E\eta_t^4 = \infty$. For example, Peng and Yao (2003) showed that the least absolute deviation estimation for a GARCH sequence has a normal limit when $E\eta_t^4 = \infty$, but the median of η_t^2 is one. Zhu and Ling (2012) proposed the so-called self-weighted quasi maximum exponential likelihood estimation (SWQMELE) by assuming $E|\eta_t| = 1$ and zero median of η_t instead of $E\eta_t = 0$ and $E\eta_t^2 = 1$ in (1.1), which ensures a normal limit regardless of $E\eta_t^4 = \infty$ and/or $E\eta_t^4 = \infty$. When $E|\eta_t|= d > 0$ in the standard ARMA-GARCH model (1.1) is unknown, after defining

$$\eta_t^* = \eta_t/d,\ h_t^* = d^2 h_t,\ \alpha_i^* = d^2\alpha_i \text{ for } i = 0, 1, \cdots, r, \text{ and } \beta_j^* = \beta_j \text{ for } j = 1, \cdots, s, \tag{1.2}$$

model (1.1) is equivalent to

$$y_t = \varphi + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} + \varepsilon_t,\ \varepsilon_t = \eta_t^* \sqrt{h_t^*},\ h_t^* = \alpha_0^* + \sum_{i=1}^{r} \alpha_i^* \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j^* h_{t-j}^*. \tag{1.3}$$

Therefore, changing $E\eta_t^2 = 1$ to $E|\eta_t|= 1$ is a simple scale transformation of parameters α_i's. More importantly, this scale transformation generally does not change risk measure inference such as conditional Value-at-Risk and conditional Expected Shortfall of y_t given $\mathcal{F}_{t-1}$, Value-at-Risk and Expected Shortfall of η_t, and comovement of two sequences. Here, $\mathcal{F}_t$ is the σ-field generated by $\{y_s : s \le t\}$. For example, it follows from (1.1) and (1.3) that

$$\begin{aligned} P(y_t \le y|\mathcal{F}_{t-1}) &= P(\eta_t \le \tfrac{y-\varphi-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j}}{\sqrt{h_t}}|\mathcal{F}_{t-1}) \\ &= P(\eta_t^* \le \tfrac{y-\varphi-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{j=1}^{q}\psi_j\varepsilon_{t-j}}{\sqrt{h_t^*}}|\mathcal{F}_{t-1}). \end{aligned}$$

That is, the conditional Value-at-Risk of y_t given $\mathcal{F}_{t-1}$ computed from (1.1) is the same as that computed from (1.3). However, changing zero mean of η_t in (1.1) to zero median requires a complicated shift transformation, which can not maintain the ARMA-GARCH structure. For example, if the median of η_t is m, then a simple transformation of (1.1) to have zero median is

$$\varepsilon_t = m\sqrt{h_t} + (\eta_t - m)\sqrt{h_t},$$

which does not have the GARCH structure due to the additional term $m\sqrt{h_t}$. That is, changing zero mean of η_t to zero median destroys the ARMA-GARCH structure and makes the SWQMELE not applicable. We refer to Fan, Qi and Xiu (2014) for more details on the transformation effect for skewed data. Therefore, it is of importance to test $H_0 : E\eta_t = 0$ when SWQMELE is applied to (1.1) under the assumptions of $E|\eta_t|= 1$ and zero median of η_t. If this null hypothesis is rejected, then the ARMA part in (1.1) no longer models the conditional mean of y_t.

This paper investigates the possibility of using the empirical likelihood method to test the hypothesis above. The empirical likelihood method is introduced by Owen (1988, 1990) and is a data-driven method combining the advantages of parametric and nonparametric methods. Under some regularity conditions, the associated empirical likelihood ratio statistic asymptotically follows a chi-squared distribution function, and the shape of the obtained confidence interval/region is determined automatically by the data. Because of its effectiveness, the empirical likelihood method has been applied in various fields to provide powerful tests and accurate interval estimation. These include Owen (1991) and Kolaczyk (1994) in general regression problems, Chuang and Chan (2002) in unstable autoregressive (AR) models, Chan and Ling (2006) in GARCH models, Liu et al. (2008) and Ciuperca and Salloum (2015) in the detection of change point, Fan and Huang (2005) in the varying-coefficient partially linear model, Chen et al. (2012) in the threshold AR models, and Zhang et al. (2019) in the tail index inference of GARCH-type models. We refer to Owen (2001) for an overview of the empirical likelihood method.

Our empirical motivation is the study of the comovement of housing price indexes in the states, where an AR-GARCH model is fitted to the house price index for each state, and QMLE is often employed; see Zimmer (2012, 2015). However, Huang, Peng and Yao (2019) confirmed that both $E\eta_t^4$ and $E\varepsilon_t^4$ may be infinite by using the Hill estimate (Hill (1975)) and found that the estimates for the mean and median of η_t are close to zero by applying the SWQMELE. This paper provides a formal test for $E\eta_t = 0$ in using the SWQMELE to fit an ARMA-GARCH model.

We organize this paper as follows. Section 2 provides the empirical likelihood test and its asymptotic result. Sections 3 and 4 present a simulation study and real data analyses on US housing price indexes and financial returns, respectively. Section 5 summarizes our conclusions. We put all proofs in the Appendix and the supplementary file.

2 Methodology and Asymptotic Result

Consider the ARMA(p,q)-GARCH(r,s) model (1.1), and let $\boldsymbol{\theta} = (\boldsymbol{\gamma}', \boldsymbol{\delta}')'$ denote the unknown parameters with true value $\boldsymbol{\theta}_0$, where $\boldsymbol{\gamma} = (\varphi, \phi_1, ..., \phi_p, \psi_1, ..., \psi_q)'$ and $\boldsymbol{\delta} = (\alpha_0, \alpha_1, ..., \alpha_r, \beta_1, ..., \beta_s)'$. Let $\Theta = \Theta_\gamma \times \Theta_\delta$ be the parameter space with $\Theta_\gamma \subset \mathbb{R}^{p+q+1}$, $\Theta_\delta \subset \mathbb{R}_0^{r+s+1}$, $\mathbb{R} = (-\infty, \infty)$ and $\mathbb{R}_0 = [0, \infty)$. Given the observations $\{y_n, ..., y_1\}$ and the initial values $\{y_0, y_{-1}, ...\}$, we write the parametric form of (1.1) as

$$\begin{cases} \varepsilon_t(\boldsymbol{\gamma}) = y_t - \varphi - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{j=1}^q \psi_j \varepsilon_{t-j}(\boldsymbol{\gamma}), \ \eta_t(\boldsymbol{\theta}) = \dfrac{\varepsilon_t(\boldsymbol{\gamma})}{\sqrt{h_t(\boldsymbol{\theta})}}, \\ h_t(\boldsymbol{\theta}) = \alpha_0 + \sum_{i=1}^r \alpha_i \varepsilon_{t-i}^2(\boldsymbol{\gamma}) + \sum_{j=1}^s \beta_j h_{t-j}(\boldsymbol{\theta}). \end{cases} \tag{2.4}$$

Write $\eta_t(\boldsymbol{\theta}_0) = \eta_t$, $\varepsilon_t(\boldsymbol{\gamma}_0) = \varepsilon_t$, $h_t(\boldsymbol{\theta}_0) = h_t$, and denote $\phi(z) = 1 - \sum_{i=1}^p \phi_i z^i$, $\psi(z) = 1 + \sum_{i=1}^q \psi_i z^i$, $\alpha(z) = \sum_{i=1}^r \alpha_i z^i$, and $\beta(z) = 1 - \sum_{i=1}^s \beta_i z^i$. Like Ling (2007) and Zhu and Ling (2012), we take $y_i = 0$ for $i \leq 0$ and impose the following regularity conditions.

Assumption 1. *Assume that Θ_γ and Θ_δ are compact, and $\boldsymbol{\theta}_0$ is an interior point in Θ. For each $\boldsymbol{\theta} \in \Theta$, $\phi(z) \neq 0$ when $|z| \leq 1$, and $\phi(z)$ and $\psi(z)$ have no common root with $\phi_p \neq 0$ or $\psi_q \neq 0$.*

Assumption 2. *For each $\boldsymbol{\theta} \in \Theta$, $\alpha(z)$ and $\beta(z)$ have no common root, $\alpha(1) \neq 1$, $\alpha_r + \beta_s \neq 0$, and $\sum_{i=1}^s \beta_i < 1$.*

Assumption 3. *$\{\eta_t\}$ is a sequence of independent and identically distributed random variables.*

Assumption 4. *$E[(w_t + w_t^2)\xi_{\rho,t-1}^4] < \infty$ for any $\rho \in (0,1)$, where $\xi_{\rho,t} = 1 + \sum_{i=0}^\infty \rho^i |y_{t-i}|$.*

Assumption 5. *η_t has zero median with $E|\eta_t|=1$, $E\eta_t^2<\infty$, and a continuous density function $g(x)$ satisfying $g(0)>0$ and $\sup_{x\in\mathbb{R}} g(x)<\infty$.*

Assumption 6. *$E|\varepsilon_t|^{\iota}<\infty$ for some $\iota>0$.*

Assumption 1 implies the stationarity, invertibility, and identifiability of the ARMA part of model (2.4), under which it follows that

$$\psi^{-1}(z)=\sum_{i=0}^{\infty}a_{\psi}(i)z^i \quad \text{and} \quad \phi(z)\psi^{-1}(z)=\sum_{i=0}^{\infty}a_{\gamma}(i)z^i,$$

where $\sup_{\Theta_\gamma} a_\psi(i)=O(\rho^i)$ and $\sup_{\Theta_\gamma} a_\gamma(i)=O(\rho^i)$ for some $0<\rho<1$.

Assumption 2 is the identification condition for the GARCH part of the model (2.4). Under this condition, we have

$$\beta^{-1}(z)=\sum_{i=0}^{\infty}a_{\beta}(i)z^i \quad \text{and} \quad \alpha(z)\beta^{-1}(z)=\sum_{i=1}^{\infty}a_{\delta}(i)z^i,$$

where $\sup_{\Theta_\delta} a_\beta(i)=O(\rho^i)$ and $\sup_{\Theta_\delta} a_\delta(i)=O(\rho^i)$ for some $0<\rho<1$; see Lemma 2.1 in Ling (1999). Assumption 6 ensures the stationarity of $\{\varepsilon_t\}$.

The weight w_t in Assumption 4 is used to reduce the moment condition on ε_t as Zhu and Ling (2012). A particular choice suggested by Zhu and Ling (2012) and Pan, Wang and Yao (2007) is employed in the simulation study and empirical analysis below.

Assumption 5 allows Zhu and Ling (2012) to study the SWQMELE, which reduces the moment condition on η_t for having a normal limit. More specifically, by temporarily assuming that η_t follows the standard double exponential distribution, one can minimize the following weighted negative log-likelihood function to obtain the SWQMELE:

$$L_{sn}(\boldsymbol{\theta})=\frac{1}{n}\sum_{t=1}^{n}w_t l_t(\boldsymbol{\theta}) \quad \text{and} \quad l_t(\boldsymbol{\theta})=\log\sqrt{h_t(\boldsymbol{\theta})}+\frac{|\varepsilon_t(\boldsymbol{\gamma})|}{\sqrt{h_t(\boldsymbol{\theta})}}.$$

As argued in the introduction, changing zero mean of η_t in the standard ARMA-GARCH model to zero median destroys the ARMA-GARCH structure and has a serious effect on skewed data. Therefore, it becomes important to test

$$H_0 : E\eta_t = 0 \quad \text{against} \quad H_1 : E\eta_t \neq 0 \tag{2.5}$$

under the above Assumptions 1–6. When this null hypothesis can not be rejected, the ARMA part still models the conditional mean, and conditional mean and conditional median are equal.

To formulate an empirical likelihood test, we follow the idea of using estimating equations in Qin and Lawless (1994). First, we calculate the score functions of the SWQMELE, i.e.,

$$\begin{aligned}\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{1}{2h_t(\boldsymbol{\theta})}\frac{\partial h_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{1}{\sqrt{h_t(\boldsymbol{\theta})}}\frac{\partial \varepsilon_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\theta}} sgn(\eta_t(\boldsymbol{\theta})) - \frac{|\varepsilon_t(\boldsymbol{\gamma})|}{2\sqrt{h_t(\boldsymbol{\theta})}}\frac{1}{h_t(\boldsymbol{\theta})}\frac{\partial h_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \frac{1}{2h_t(\boldsymbol{\theta})}\frac{\partial h_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}(1 - |\eta_t(\boldsymbol{\theta})|) + \frac{1}{\sqrt{h_t(\boldsymbol{\theta})}}\frac{\partial \varepsilon_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\theta}} sgn(\eta_t(\boldsymbol{\theta})),\end{aligned}$$

where sgn denotes the sign function. Put $\boldsymbol{D}_{t,1}(\boldsymbol{\theta}) = w_t \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, $D_{t,2}(\boldsymbol{\theta}, \mu) = w_t(\eta_t(\boldsymbol{\theta}) - \mu)$, and $\mu = \mu(\boldsymbol{\theta}) = E\eta_t(\boldsymbol{\theta})$ with true value μ_0. Then $\boldsymbol{\theta}$ and μ can be estimated simultaneously by solving the following estimating equations:

$$\sum_{t=1}^{n} \boldsymbol{D}_{t,1}(\boldsymbol{\theta}) = 0 \quad \textit{and} \quad \sum_{t=1}^{n} D_{t,2}(\boldsymbol{\theta}, \mu) = 0.$$

This defines the following empirical likelihood function

$$L(\boldsymbol{\theta}, \mu) = \sup\{\prod_{t=1}^{n}(np_t) : p_1 \geq 0, ..., p_n \geq 0, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t \boldsymbol{D}_t(\boldsymbol{\theta}, \mu) = 0\},$$

where $\boldsymbol{D}_t(\boldsymbol{\theta}, \mu) = \big(\boldsymbol{D}'_{t,1}(\boldsymbol{\theta}), D_{t,2}(\boldsymbol{\theta}, \mu)\big)'$. Using the Lagrange multiplier technique, the log empirical likelihood function becomes

$$l(\boldsymbol{\theta}, \mu) = -2\log(L(\boldsymbol{\theta}, \mu)) = 2\sum_{t=1}^{n} \log\{1 + \boldsymbol{\lambda}'\boldsymbol{D}_t(\boldsymbol{\theta}, \mu)\}, \tag{2.6}$$

where $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\theta}, \mu)$ satisfies

$$\sum_{t=1}^{n} \frac{\boldsymbol{D}_t(\boldsymbol{\theta}, \mu)}{1 + \boldsymbol{\lambda}'\boldsymbol{D}_t(\boldsymbol{\theta}, \mu)} = 0.$$

Because we are only interested in μ, we consider the profile empirical likelihood ratio $l_p(\mu) = \min_{\boldsymbol{\theta}\in\Theta} l(\boldsymbol{\theta}, \mu)$. The following theorem shows that the Wilks theorem holds for the proposed empirical likelihood method.

Theorem 1. *Suppose (1.1) satisfies Assumptions 1–6. Under $H_0 : E\eta_t = 0$, $l_p(0)$ converges in distribution to a chi-squared limit with one degree of freedom as $n \to \infty$.*

Based on the above theorem we reject $H_0 : E\eta_t = 0$ at the level $1 - a$ whenever $l_p(0) > \chi^2_{1,a}$, where $\chi^2_{1,a}$ denotes the a-th quantile of a chi-squared distribution function with one degree of freedom.

The next theorem shows that the proposed test has power.

Theorem 2. *Suppose that Assumptions 1-6 hold for the model (1.1). When $\mu_0 = M/\sqrt{n}$ for some constant M, $l_p(0)$ converges in distribution to a non-central chi-squared limit with one degree of freedom and the noncentral parameter $M^2(Ew_t)^2$ as $n \to \infty$.*

3 Simulation

This section examines the finite sample performance of the proposed empirical likelihood test by generating data from an AR(1)-GARCH(1,1) model:

$$y_t = \varphi + \phi_1 y_{t-1} + \varepsilon_t,\ \varepsilon_t = \eta_t \sqrt{h_t},\ h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}.$$

We take $\boldsymbol{\theta}_0 = (0.0797, -0.0465, 0.0347, 0.1572, 0.8057)$, $n = 200$ or 500 or 1000 or 2500 or 5000, and choose w_t as

$$w_t = \left(\max\left\{1, \frac{1}{C}\sum_{k=1}^{\infty}\frac{1}{k^9}|y_{t-k}|I\{|y_{t-k}|> C\}\right\}\right)^{-4} \tag{3.7}$$

with C being the 95% or 90% quantile of $\{|y_t|\}_{t=1}^n$. This particular choice is suggested by Ling (2007) and Pan, Wang and Yao (2007) when $E|\varepsilon_t|< \infty$. For computing the size of the proposed empirical likelihood test at levels 10% and 5%, we consider $\eta_t = \widetilde{\eta}_t / E|\widetilde{\eta}_t|$, where $\widetilde{\eta}_t \sim Laplace(0,1)$ and $\widetilde{\eta}_t \sim N(0,1)$. Because they are symmetric, the considered η_t has both zero mean and zero median. Table 1 reports the computed empirical sizes based on 10000 repetitions. We observe that i) the coverage probability is robust to the two choices of C, ii) the normal error provides a less accurate size than the Laplace error as we fit the ARMA-GARCH model based on a Laplace distribution, iii) and the accuracy improves for a large sample size.

Table 1: We report the sizes of the profile empirical likelihood test for $H_0 : E\eta_t = 0$, where C in (3.7) is chosen as the 95% (left panel) and 90% (right panel) quantile of $\{|y_t|\}_{t=1}^n$

n	Level 5% Laplace	Level 10% Laplace	Level 5% Normal	Level 10% Normal	Level 5% Laplace	Level 10% Laplace	Level 5% Normal	Level 10% Normal
200	0.0644	0.1181	0.0419	0.0797	0.0644	0.1160	0.0418	0.0801
500	0.0418	0.0843	0.0320	0.0654	0.0431	0.0849	0.0306	0.0683
1000	0.0455	0.0900	0.0328	0.0766	0.0458	0.0915	0.0333	0.0750
2500	0.0425	0.0879	0.0387	0.0803	0.0419	0.0856	0.0392	0.0811
5000	0.0456	0.0946	0.0436	0.0939	0.0455	0.0941	0.0418	0.0914

.

Next, we compute the power of the proposed empirical likelihood test at levels 10% and 5% by considering the standardized Beta(1.2, 1) and Beta(2, 2.7), i.e., $E|\eta_t|= 1$ and zero median. By drawing 100,000 random samples with sample size 1,000,000 from these two Beta distributions, we find that $E\eta_t$ is around -0.0659 for Beta(1.2,1) and 0.0660 for Beta(2,2.7), i.e., H_0 is false. Table 2 reports the computed empirical powers, which shows that i) the proposed test is powerful, ii) a larger sample size gives a better power, iii) the power is robust to the two choices of C, iv) and Beta(2, 2.7) has slightly better power than Beta (1.2, 1) as the former one departs a bit more away from the null hypothesis.

Table 2: We report the powers of the profile empirical likelihood test for $H_0 : E\eta_t = 0$, where C in (3.7) is chosen as the 95% (left panel) and 90% (right panel) quantile of $\{|y_t|\}_{t=1}^n$.

n	Level 5% Beta(1.2,1)	Level 10% Beta(1.2,1)	Level 5% Beta(2,2.7)	Level 10% Beta(2,2.7)	Level 5% Beta(1.2,1)	Level 10% Beta(1.2,1)	Level 5% Beta(2,2.7)	Level 10% Beta(2,2.7)
200	0.1249	0.1914	0.0977	0.1657	0.1192	0.1851	0.0986	0.1660
500	0.1746	0.2712	0.1783	0.2846	0.1687	0.2674	0.1755	0.2734
1000	0.3364	0.4705	0.3830	0.5090	0.3247	0.4584	0.3678	0.5030
2500	0.7408	0.8347	0.8083	0.8857	0.7274	0.8243	0.7974	0.8776

In summary, the proposed empirical likelihood test has a reasonable size and is powerful and robust to the two choices of C. It remains challenging to find an optimal weight function to improve the size.

4 Real Analysis

4.1 US Housing Price Indexes

This subsection applies the proposed test to the quarterly percentage changes of the state-level house price index (HPI) of California, Florida, Nevada, and Arizona from 1975 to 2018, which were estimated and published by the Federal Housing Finance Agency. The total number of observations for each state is 176. To study the comovement and contagion in the housing market, one often fits an AR-GARCH model to each HPI sequence. Recently, Huang, Peng and Yao (2019) found that an AR(3)-GARCH(1,1) model fits the data well by using the SWQMELE and the assumptions of $E\eta_t^4 < \infty$ and $E\varepsilon_t^4 < \infty$ are questionable after estimating the tail indexes of η_t and ε_t by the well-known Hill tail index estimate (Hill (1975)); see Table 3 for the fitted models. Although the estimates for the mean and median of η_t are close to zero (see Table 4 below), Huang, Peng and Yao (2019) didn't formally test $H_0 : E\eta_t = 0$ to ensure the AR part models both conditional mean and conditional median. An application of the proposed empirical likelihood test gives P-values 0.267, 0.480, 0.637, and 0.830 for California, Florida, Nevada, and Arizona, respectively; see Table 4. Therefore, the null hypothesis $H_0 : E\eta_t = 0$ can not be rejected, i.e., the employed AR(3)-GARCH(1,1) model still studies the conditional mean, and the application of combining copula and SWQMELE for studying comovement of house pricing indexes in the literature is methodologically sound as copula is invariant to the scale transformation.

Table 3: SWQMELE for fitting AR(3)-GARCH(1,1) models.

State	$\hat{\varphi}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\omega}$	$\hat{\alpha}_1$	$\hat{\beta}_1$
CA	0.3067	0.7274	-0.0069	0.1220	0.3327	0.3305	0.1252
FL	0.2662	0.5008	0.0344	0.3174	0.0666	0.1380	0.6706
NV	0.2703	0.4475	0.0094	0.3231	0.0821	0.2810	0.5647
AZ	0.4155	0.3253	0.1169	0.2565	0.0333	0.3037	0.5810

Table 4: Profile empirical likelihood test for $E\eta_t = 0$.

State	mean $\hat{\eta}_t$	median $\hat{\eta}_t$	mean$\|\hat{\eta}_t\|$	$\hat{l}_p(\mu_0)$	P-value
CA	-0.0944	-0.0068	1.0161	1.2332	0.267
FL	-0.0661	-0.0217	0.9888	0.4995	0.480
NV	0.0245	-0.0230	0.9951	0.2231	0.637
AZ	0.0028	-0.0454	0.9863	0.0459	0.830

4.2 Financial returns

In this subsection, we study the daily S&P500 and Microsoft Stock (MSFT) close prices from August 3, 2009 to July 29, 2019, which gives a total of 2514 observations for each sequence. We fit an ARMA-GARCH model to the 100 times log-returns. We plot the prices and log-returns in Figures 1 and 2, respectively.

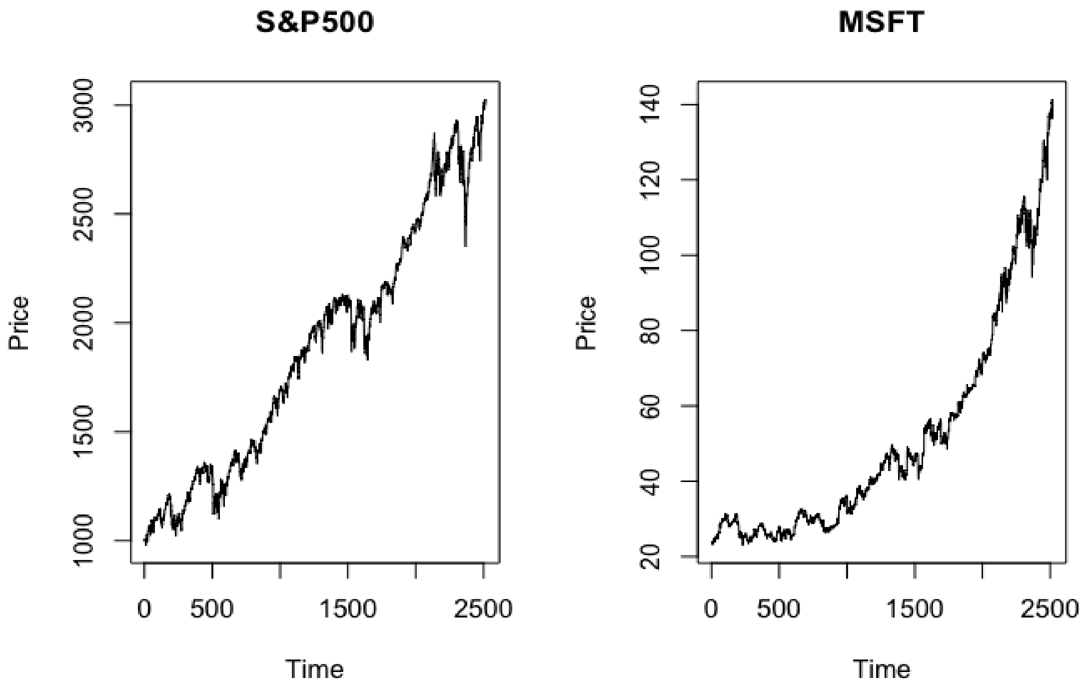

Figure 1: The daily S&P500 and MSFT close prices from August 3, 2009 to July 9, 2019.

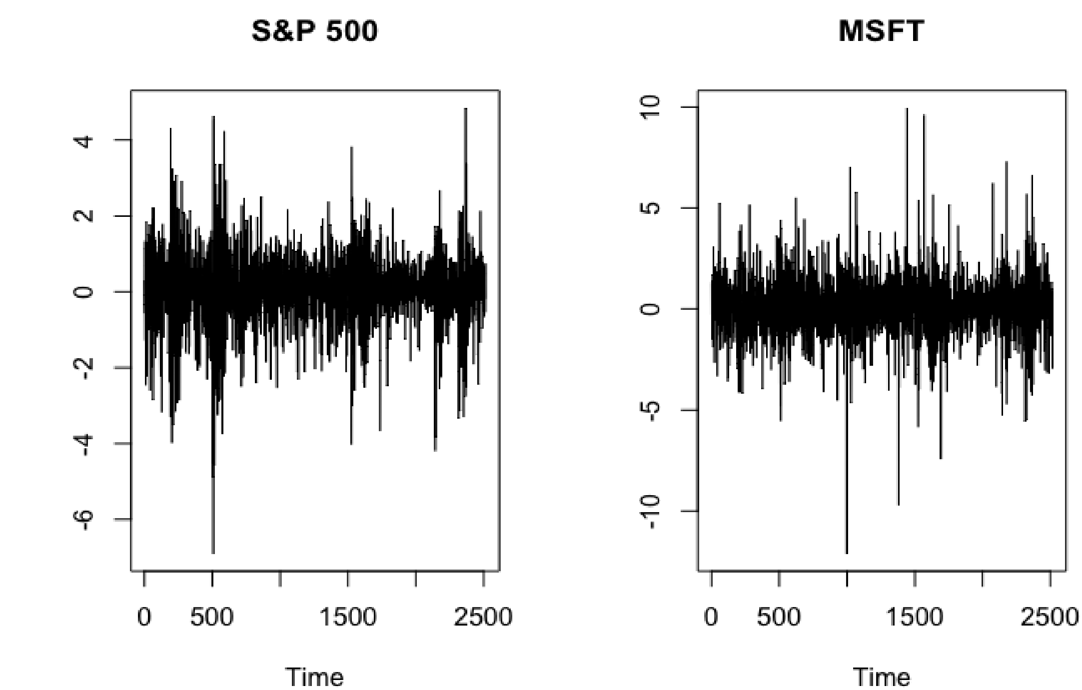

Figure 2: The 100 times log-return of daily S&P500 and MSFT close prices from August 3, 2009 to July 9, 2019.

To apply the proposed test, we first use the function "auto.arima" in the R package "forecast" with AIC to obtain the appropriate order of the ARMA part, which suggests an ARMA(2,2)-GARCH(1,1) model for S&P500 returns and an ARMA(1,1)-GARCH(1,1) model for MSFT returns. An application of the quasi maximum likelihood inference in the R package "fGarch" gives the fitted ARMA(2,2)-GARCH(1,1) model for S&P 500 as

$$\begin{cases} y_t = \underset{(0.0015)}{0.0054} - \underset{(0.0118)}{0.0287} y_{t-1} + \underset{(0.0115)}{0.9469} y_{t-2} + \underset{(0.0071)}{0.0052} \varepsilon_{t-1} - \underset{(0.0072)}{0.9784} \varepsilon_{t-2} + \varepsilon_t, \\ \varepsilon_t = \eta_t \sqrt{h_t}, \; h_t = \underset{(0.0084)}{0.0345} + \underset{(0.0244)}{0.1588} \varepsilon_{t-1}^2 + \underset{(0.0247)}{0.8042} h_{t-1}, \end{cases}$$

and the fitted ARMA(1,1)-GARCH(1,1) model for MSFT as

$$\begin{cases} y_t = \underset{(0.0105)}{0.0217} + \underset{(0.0974)}{0.7716} y_{t-1} - \underset{(0.0952)}{0.8083} \varepsilon_{t-1} + \varepsilon_t, \\ \varepsilon_t = \eta_t \sqrt{h_t}, \; h_t = \underset{(0.1353)}{0.3391} + \underset{(0.0372)}{0.1206} \varepsilon_{t-1}^2 + \underset{(0.0846)}{0.7216} h_{t-1}. \end{cases}$$

Here, the numbers in brackets are standard errors.

The autocorrelation functions of the estimated η_t and $|\eta_t|$ in Figure 3 show that the fitted models are reasonable for both log-returns. Using the R package 'evmix', we plot the Hill estimates for the estimated η_t and the 90% confidence intervals in Figure 4, which suggests that the quasi maximum likelihood estimation for S&P500 may have a normal limit, but it is not for MSFT. Next, we examine whether SWQMELE is applicable.

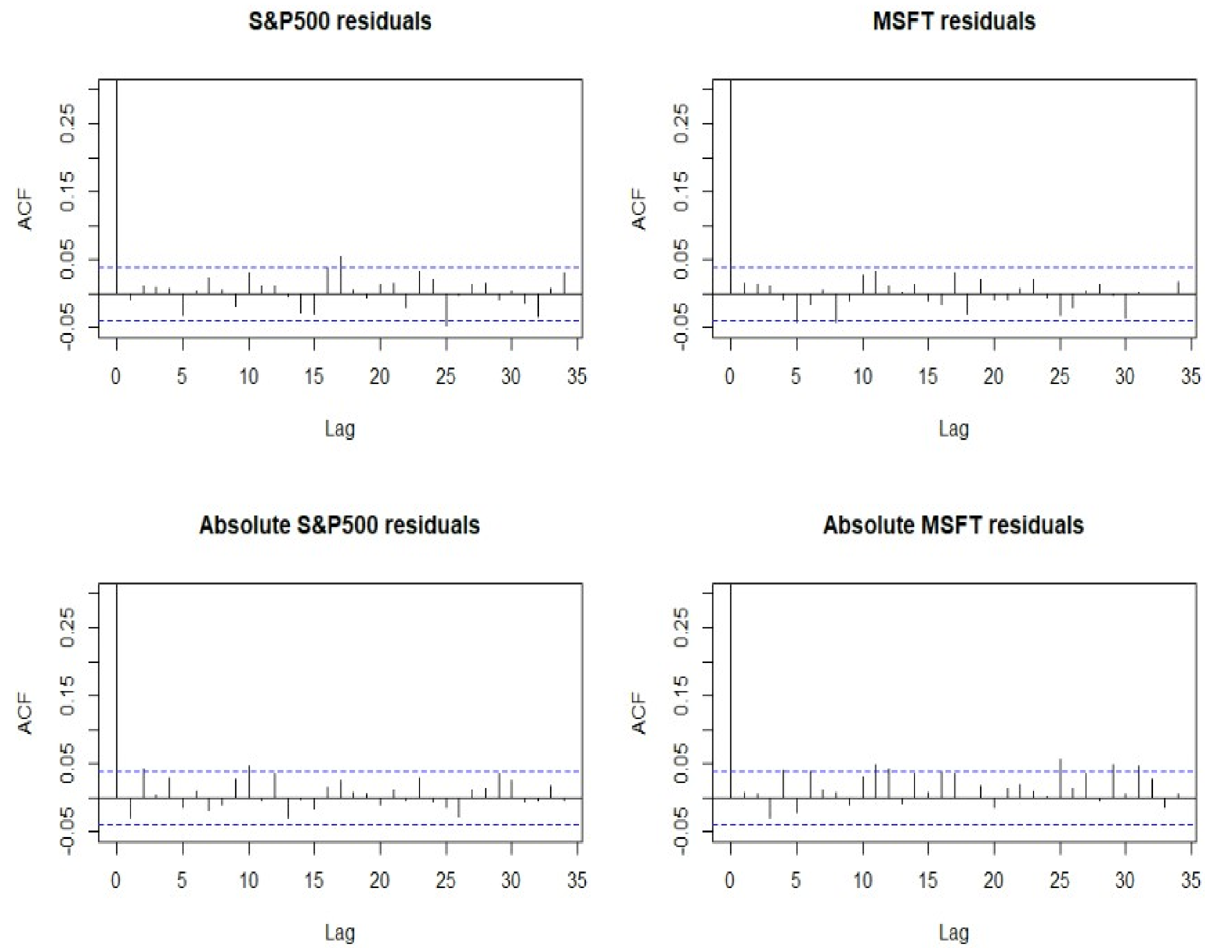

Figure 3: We plot the autocorrelation functions computed from the estimated η_t's and $|\eta_t|$'s for S&P 500 and MSFT.

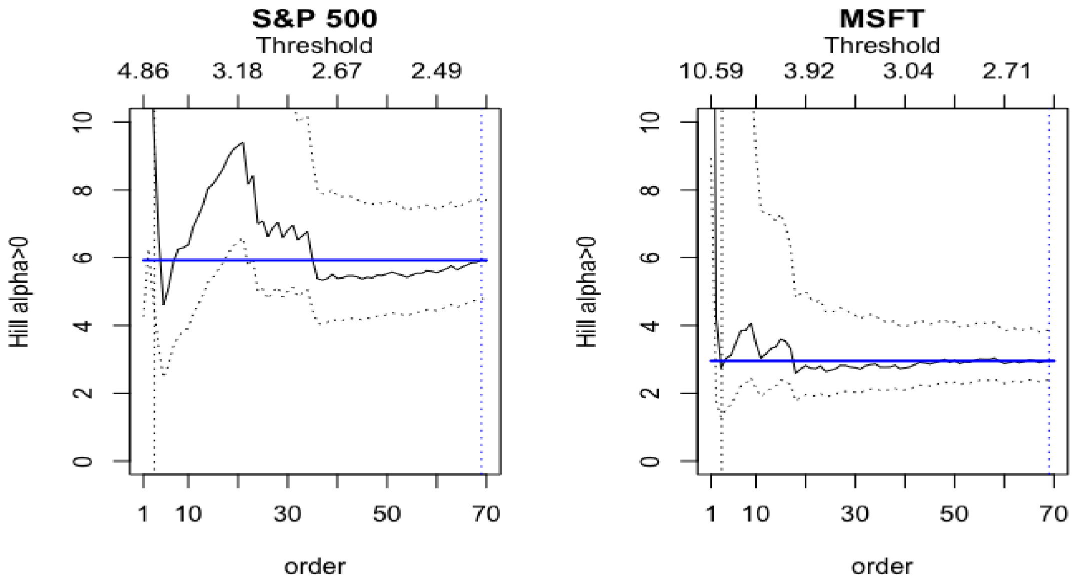

Figure 4: Using the R package 'evmix', we compute the Hill estimates from the estimated η_t's and the 90% confidence intervals (dotted lines) for S&P 500 and MSFT.

An application of the SWQMELE shows that the mean of η_t is -0.0201 and 0.0029, the median of η_t is 0.0335 and -0.0083, and the mean of $|\eta_t|$ is 1.0048 and 1.0014, respectively, for S&P 500 and MSFT. The proposed empirical likelihood test gives $\hat{l}_p(0) = 16.1283$ and 0.6071, respectively, for S&P 500 and MSFT with the corresponding P-values 5.919×10^{-5} and 0.436. Therefore, the null hypothesis of $H_0 : E\eta_t = 0$ is rejected for S&P 500 but is not rejected for MSFT, although the estimated mean and median of η_t are close to each other for both series. This analysis suggests that it is better and sound to apply the SWQMELE to MSFT log-returns because the sequence does not have enough moments and has both zero mean and zero median. In contrast, the application of the SWQMELE to the S&P500 log-returns means that we no longer model the conditional mean as the proposed test does not support both zero mean and zero median. Because the sequence has enough moments shown by the Hill estimate, it may be good to use the QMLE for the S&P500 log-returns.

In summary, it is practically vital to test whether the error in an ARMA-GARCH model has both zero mean and zero median in applying the SWQMELE because changing zero mean to zero median destroys the ARMA-GARCH structure.

5 Conclusions

QMLE is often employed to fit an ARMA-GARCH model, which requires finite fourth moments for both errors and the sequence itself to ensure a normal limit. Having a normal limit is necessary to ensure that the standard residual-based bootstrap method can be employed to quantify uncertainty. The self-weighted quasi maximum exponential likelihood estimation (SWQMELE) significantly reduces the moment conditions. Still, it requires errors to have zero median instead of zero mean in the original ARMA-GARCH models. Generally, changing zero mean to zero median destroys the ARMA-GARCH structure. This paper proposes an efficient empirical likelihood method for testing zero mean of errors in using the SWQMELE to fit an ARMA-GARCH model for returns. Applications to real datasets show that the SWQMELE does not apply to S&P500 log-returns, but is useful for US housing price indexes and MFST log-returns.

Acknowledgments

We thank two reviewers for helpful comments. Peng's research was partly supported by the Simons Foundation and the NSF grant of DMS-2012448. Zhang's research was supported by grants from NSFC (No.11771390/11371318), the USyd-ZJU Partnership Collaboration Awards, and the Fundamental Research Funds for the Central Universities.

References

- Berkes, I., Horváth, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli* **9**: 201–227.

- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307–327.
- Chan, N. H. and Ling, S. (2006). Empirical likelihood for models. *Econometric Theory* **22**: 403–428.
- Chen, H. Q., Chong, T. T. L and Bai, J. S. (2012). Theory and applications of TAR model with two threshold variables. *Econometric Reviews* **31**: 142–170.
- Chuang, C. S. and Chan, N. H. (2002). Empirical likelihood for autoregressive models with applications to unstable time series. *Statistica Sinica* **12**: 387–407.
- Ciuperca, G. and Salloum, Z. (2015). Empirical likelihood test in a posteriori change-point nonlinear model. *Metrika: International Journal for Theoretical and Applied Statistics* **78**: 919–952.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1007.
- Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11**: 1031-1057.
- Fan, J., Qi, L. and Xiu, D. (2014). Quasi Maximum Likelihood Estimation of GARCH Models with Heavy-Tailed Likelihoods. *Journal of Business and Economics Statistics* **32**: 178-205.
- Francq, C. and Zakoïan, J. M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processe. *Bernoulli* **10**: 605–637.
- Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application.* New York: Academic Press.
- Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**: 285–317.
- Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *Annals of Statistics* **3**: 1163–1174.
- Huang, H., Peng, L. and Yao, V.W. (2019). Comovements and asymmetric tail dependence in state housing prices in the USA: A nonparametric approach. *Journal of Applied Econo-*

metrics **34**: 843–849.

- Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* **14**: 70-86.
- Kolaczyk, E. D. (1994). Empirical Likelihood for generalized linear models. *Statistica Sinica* **4**: 2007–2020.
- Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *Journal of Applied probability* **36**: 688-705.
- Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* **140**: 849–873.
- Ling, S. and Li, W.K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association* **92**: 1184–1194.
- Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH Model. *Econometric Theory* **19**: 280–310.
- Liu, Y., Zou, L. and Zhang, R. (2008). Empirical likelihood ratio test for a change-point in linear regression model. *Communications in Statistics - Theory and Methods* **37**: 2551–2563.
- Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**: 237–249.
- Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics* **18**: 90–120.
- Owen, A. B. (1991). Empirical likelihood for linear models. *Annals of Statistics* **19**: 1725–1747.
- Owen, A.B. (2001). Empirical Likelihood. *Chapman Hall.*
- Pan, J., Wang, H. and Yao, Q. (2007). Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econometric Theory* **23**: 852–879.
- Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**: 967-975.

- Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics* **22:** 300-325.
- Zhang, R., Li, C. and Peng, L. (2019). Inference for the tail index of a GARCH(1,1) model and an AR(1) model with ARCH(1) errors. *Econometric Reviews* **38**: 151–169.
- Zhu, K. and Ling, S. (2012). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA–GARCH/IGARCH models. *Annals of Statistics* **39**: 2131–2163.
- Zimmer, D. M. (2012). The role of copulas in the housing crisis. *Review of Economics and Statistics* **94**: 607-620.
- Zimmer, D. M. (2015). Asymmetric dependence in house prices: evidence from USA and international data. *Empirical Economics* **49**: 161-183.

Appendix: Proofs of Theorems

First we need some lemmas, where the lengthy proofs of Lemmas 2 and 3 are given in the supplementary file. Recall μ_0 denotes the true value of $\mu = E\eta_t$ and is zero under the null hypothesis $H_0 : E\eta_t = 0$.

Lemma 1. *Under conditions of Theorem 1, there exist a constant $\rho \in (0,1)$, a constant $C > 0$, and a neighborhood Θ_0 of $\boldsymbol{\theta}_0$ such that for any constant $\iota_1 \in (0,1)$, we have*

$$
\begin{aligned}
&(a)\quad \sup_{\Theta_0} |\varepsilon_t(\boldsymbol{\gamma})| \le C\xi_{\rho t-1},\\
&(b)\quad \sup_{\Theta_0} \left\| \frac{\partial \varepsilon_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \right\| \le C\xi_{\rho t-1},\\
&(c)\quad \sup_{\Theta_0} \left\| \frac{\partial^2 \varepsilon_t(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}\partial \boldsymbol{\gamma}'} \right\| \le C\xi_{\rho t-1},\\
&(d)\quad \sup_{\Theta_0} \left| \frac{\partial^3 \varepsilon_t(\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j \partial \gamma_j} \right| \le C\xi_{\rho t-1} \text{ where } i,j,k = 1,...,p+q+1,\\
&(e)\quad 1 \le \sup_{\Theta_0} \frac{h_t(\boldsymbol{\theta})}{\underline{\alpha_0}} \le C\xi_{\rho t-1}^2 \text{ where } \underline{\alpha_0} = \inf_{\Theta_0} \alpha_0,\\
&(f)\quad \sup_{\Theta_0} \left\| \frac{1}{h_t(\boldsymbol{\theta})} \frac{\partial h_t(\boldsymbol{\theta})}{\partial \boldsymbol{\delta}} \right\| \le C\xi_{\rho t-1}^{\iota_1},
\end{aligned}
$$

$$(g)\quad \sup_{\Theta_0}\left\|\frac{1}{h_t(\boldsymbol{\theta})}\frac{\partial^2 h_t(\boldsymbol{\theta})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\right\| \le C\xi_{\rho t-1}^{\iota_1},$$

$$(h)\quad \sup_{\Theta_0}\left|\frac{1}{h_t(\boldsymbol{\theta})}\frac{\partial^3 h_t(\boldsymbol{\theta})}{\partial\delta_i\partial\delta_j\partial\delta_k}\right| \le C\xi_{\rho t-1}^{\iota_1} \text{ where } i,j,k=1,...,r+s+1,$$

$$(i)\quad \sup_{\Theta_0}\left\|\frac{1}{\sqrt{h_t(\boldsymbol{\theta})}}\frac{\partial h_t(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}}\right\| \le C\xi_{\rho t-1},$$

$$(j)\quad \sup_{\Theta_0}\left\|\frac{1}{\sqrt{h_t(\boldsymbol{\theta})}}\frac{\partial^2 h_t(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right\| \le C\xi_{\rho t-1},$$

$$(k)\quad \sup_{\Theta_0}\left|\frac{1}{\sqrt{h_t(\boldsymbol{\theta})}}\frac{\partial^3 h_t(\boldsymbol{\theta})}{\partial\gamma_i\partial\gamma_j\partial\gamma_k}\right| \le C\xi_{\rho t-1} \text{ where } i,j,k=1,...,p+q+1,$$

$$(l)\quad \sup_{\Theta_0}\left\|\frac{1}{\sqrt{h_t(\boldsymbol{\theta})}}\frac{\partial^2 h_t(\boldsymbol{\theta})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\gamma}'}\right\| \le C\xi_{\rho t-1},$$

$$(m)\quad \sup_{\Theta_0}\left|\frac{1}{\sqrt{h_t(\boldsymbol{\theta})}}\frac{\partial^3 h_t(\boldsymbol{\theta})}{\partial\delta_i\partial\delta_j\partial\gamma_k}\right| \le C\xi_{\rho t-1} \text{ where } i,j=1,...,p+q+1, k=1,...,r+s+1,$$

$$(n)\quad \sup_{\Theta_0}\left|\frac{1}{\sqrt{h_t(\boldsymbol{\theta})}}\frac{\partial^3 h_t(\boldsymbol{\theta})}{\partial\delta_i\partial\gamma_j\partial\gamma_k}\right| \le C\xi_{\rho t-1} \text{ where } i=1,...,p+q+1, j,k=1,...,r+s+1.$$

Proof. See Ling (2007) or Chan and Ling (2006). □

Lemma 2. *Under conditions of Lemma 1, we have*

$$(a)\quad E\sup_{\Theta_0}||\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)||^2 < \infty,$$

$$(b)\quad E\sup_{\Theta_0}||\boldsymbol{P}_t(\boldsymbol{\theta},\mu_0)|| < \infty, \text{ where } \boldsymbol{P}_t = \frac{\partial\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)}{\partial\boldsymbol{\omega}'} \text{ and } \boldsymbol{\omega} = (\boldsymbol{\theta}',\mu)'.$$

Lemma 3. *Let* $V_0 = \left\{\boldsymbol{\theta} : \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \le \frac{M}{\sqrt{n}}\right\}$ *for some constant* $M>0$*. Under conditions of Lemma 1, we have*

$$(a)\quad \max_{1\le t\le n}\sup_{V_0}\|\boldsymbol{P}_t(\boldsymbol{\theta},\mu_0)\| = o_p(n),$$

$$(b)\quad \max_{1\le t\le n}\sup_{V_0}\|\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\| = o_p(\sqrt{n}),$$

(c) $\frac{1}{n}\sum_{t=1}^n \boldsymbol{P}_t(\boldsymbol{\theta},\mu_0) = \boldsymbol{\Psi} + o_p(1)$ *holds uniformly for all* $\boldsymbol{\theta}\in V_0$,

(d) $\frac{1}{n}\sum_{t=1}^n [\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\boldsymbol{D}_t'(\boldsymbol{\theta},\mu_0)] = \boldsymbol{\Omega} + o_p(1)$ *holds uniformly for all* $\boldsymbol{\theta}\in V_0$,

where $$\boldsymbol{\Psi} = \begin{pmatrix} \frac{1}{4}E\{\frac{w_t}{h_t^2(\boldsymbol{\theta}_0)}\frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\} & \mathbf{0} \\ E\{\frac{w_t}{\sqrt{h_t(\boldsymbol{\theta}_0)}}\frac{\partial\varepsilon_t(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\theta}'}\} & -Ew_t \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \Omega_{22} \end{pmatrix},$$

$$\boldsymbol{\Omega}_{11} = E\{\frac{w_t^2}{h_t(\boldsymbol{\theta}_0)}\frac{\partial\varepsilon_t(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\varepsilon_t(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\theta}'}\} + \frac{E\eta_t^2-1}{4}E\{\frac{w_t^2}{h^2(\boldsymbol{\theta}_0)}\frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\},$$

$$\boldsymbol{\Omega}_{12} = -\frac{E(\eta_t^2 sgn(\eta_t))}{2}E\{\frac{w_t^2}{h_t(\boldsymbol{\theta}_0)}\frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\},$$

$$\boldsymbol{\Omega}_{21} = -\frac{E(\eta_t^2 sgn(\eta_t))}{2}E\{\frac{w_t^2}{h_t(\boldsymbol{\theta}_0)}\frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\}, \qquad \Omega_{22} = E(w_t^2\eta_t^2).$$

Lemma 4. *Under conditions of Lemma 1, we have*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta}_0,\mu_0)\xrightarrow{d}N(0,\boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega}$ *is defined in Lemma 3.*

Proof. Recall $\mathcal{F}_t$ is the σ-field generated by $\{y_m, m\le t\}$. Then,

$$\begin{aligned}
E(\boldsymbol{D}_{t,1}(\boldsymbol{\theta}_0)|\mathcal{F}_{t-1}) &= E\left(\frac{w_t}{2h_t(\boldsymbol{\theta}_0)}\frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}(1-|\eta_t(\boldsymbol{\theta}_0)|)+\frac{w_t}{\sqrt{h_t(\boldsymbol{\theta}_0)}}\frac{\partial\varepsilon_t(\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\theta}}sgn(\eta_t(\boldsymbol{\theta}_0))\bigg|\mathcal{F}_{t-1}\right)\\
&= 0,\quad \text{and}\\
E(D_{t,2}(\boldsymbol{\theta}_0,\mu_0)|\mathcal{F}_{t-1}) &= E(w_t\eta_t|\mathcal{F}_{t-1})=w_tE\eta_t=0.
\end{aligned}$$

Hence, $\boldsymbol{D}_t(\boldsymbol{\theta}_0,\mu_0)$ is a sequence of martingale differences, and the theorem follows from the Central Limit Theorem of Martingales (see Page 58 of Hall and Heyde (1980)). □

Proof of Theorem 1. Let $\boldsymbol{\theta}=\boldsymbol{\theta}_0+\frac{\boldsymbol{\nu}}{\sqrt{n}}$, where $\boldsymbol{\nu}$ is a $(p+q+r+s+2)$-dimensonal vector. Denote $\boldsymbol{g}(\boldsymbol{\theta},\mu_0,\boldsymbol{\lambda})=\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)}{1+\boldsymbol{\lambda}'\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)}$ and $\gamma_t(\boldsymbol{\theta},\mu_0)=\boldsymbol{\lambda}'\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)$, where $\boldsymbol{\lambda}$ is a solution of equation $\boldsymbol{g}(\boldsymbol{\theta},\mu_0,\boldsymbol{\lambda})=\boldsymbol{0}$. We first show that $\max_{1\le t\le n}|\gamma_t(\boldsymbol{\theta},\mu_0)|=o_p(1)$.

Let $\boldsymbol{\lambda}=\rho\boldsymbol{r}$ with $\|\boldsymbol{r}\|=1$. Observe that

$$\begin{aligned}
0=\|\boldsymbol{g}(\boldsymbol{\theta},\mu_0,\rho\boldsymbol{r})\| &\ge \|\boldsymbol{r}'\boldsymbol{g}(\boldsymbol{\theta},\mu_0,\rho\boldsymbol{r})\|\\
&= \left|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{r}'\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)-\frac{1}{n}\rho\sum_{t=1}^{n}\frac{\boldsymbol{r}'\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\boldsymbol{D}_t'(\boldsymbol{\theta},\mu_0)\boldsymbol{r}}{1+\rho\boldsymbol{r}'\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)}\right|\\
&\ge -\frac{1}{n}\left|\sum_{t=1}^{n}\boldsymbol{r}'\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\right|+\frac{\rho\boldsymbol{r}'\boldsymbol{S}_n(\boldsymbol{\theta},\mu_0)\boldsymbol{r}}{1+\rho Z_n(\boldsymbol{\theta},\mu_0)},
\end{aligned}$$

where $\boldsymbol{S}_n(\boldsymbol{\theta},\mu_0)=\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\boldsymbol{D}_t'(\boldsymbol{\theta},\mu_0),\ Z_n(\boldsymbol{\theta},\mu_0)=\max_{1\le t\le n}\|\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\|$. Thus,

$$\begin{aligned}\frac{\rho\boldsymbol{r}'\boldsymbol{S}_n(\boldsymbol{\theta},\mu_0)\boldsymbol{r}}{1+\rho Z_n(\boldsymbol{\theta},\mu_0)} &\le \frac{1}{n}\left|\sum_{t=1}^{n}\boldsymbol{r}'\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\right| \\ &\le \frac{1}{n}\left|\sum_{t=1}^{n}\boldsymbol{r}'\boldsymbol{D}_t(\boldsymbol{\theta}_0,\mu_0)\right| + \frac{1}{n}\left|\sum_{t=1}^{n}\boldsymbol{r}'\boldsymbol{P}_t(\boldsymbol{\theta}^*,\mu_0)\right|\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \\ &\le \frac{1}{n}\left|\sum_{t=1}^{n}\boldsymbol{r}'\boldsymbol{D}_t(\boldsymbol{\theta}_0,\mu_0)\right| + \left(\frac{1}{n}\sum_{t=1}^{n}\|\boldsymbol{P}_t(\boldsymbol{\theta}^*,\mu_0)\|\right)\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \\ &= O_p(\frac{1}{\sqrt{n}})\end{aligned}$$

holds uniformly for any $\boldsymbol{\theta}\in V_0$, where the last equation follows from Lemma 4 that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta}_0,\mu_0)=O_p(1)$ and the proof of Lemma 2 in the supplementary file. Consequently,

$$\frac{\rho\boldsymbol{r}'\boldsymbol{S}_n(\boldsymbol{\theta},\mu_0)\boldsymbol{r}}{1+\rho Z_n(\boldsymbol{\theta},\mu_0)} = O_p(\frac{1}{\sqrt{n}}).$$

By Lemma 3(d), $\boldsymbol{r}'\boldsymbol{S}_n(\boldsymbol{\theta},\mu_0)\boldsymbol{r}\ge a+o_p(1)$ holds uniformly for $\boldsymbol{\theta}\in V_0$, where a is the smallest eigenvalue of $\boldsymbol{\Omega}$. Thus,

$$\rho=\|\boldsymbol{\lambda}\|=O_p(\frac{1}{\sqrt{n}}) \tag{5.8}$$

holds uniformly in $\boldsymbol{\theta}\in V_0$. Then, by Lemma 3(b),

$$\max_{1\le t\le n}|\gamma_t(\boldsymbol{\theta},\mu_0)|\le\|\lambda\|\max_{1\le t\le n}\|\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\|=o_p(1) \tag{5.9}$$

holds uniformly in $\boldsymbol{\theta}\in V_0$.

On the other hand, Taylor expansion and Lemma 3 (c) imply that

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0) &= \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{D}_t(\boldsymbol{\theta}_0,\mu_0)+\boldsymbol{P}_t(\boldsymbol{\theta}^*,\mu_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+o_p(\frac{1}{\sqrt{n}})) \\
&= \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta}_0,\mu_0)+\frac{\boldsymbol{\nu}}{n\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{P}_t(\boldsymbol{\theta}^*,\mu_0)+o_p(\frac{1}{\sqrt{n}}) \\
&= O_p(\frac{1}{\sqrt{n}})+\frac{\boldsymbol{\nu}}{\sqrt{n}}(\boldsymbol{\Psi}+o_p(1))+o_p(\frac{1}{\sqrt{n}}) \\
&= O_p(\frac{1}{\sqrt{n}})+o_p(\frac{1}{\sqrt{n}}) \\
&= O_p(\frac{1}{\sqrt{n}})
\end{aligned}$$

holds uniformly in V_0, where $\boldsymbol{\theta}^*$ lies between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$. By (5.9), we have

$$\begin{aligned}
&\mathbf{0}=\boldsymbol{g}(\boldsymbol{\theta},\mu_0,\boldsymbol{\lambda}) \\
&=\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)-\boldsymbol{S}_n(\boldsymbol{\theta},\mu_0)\boldsymbol{\lambda}+\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\gamma_t^2(\boldsymbol{\theta},\mu_0)}{1+\gamma_t(\boldsymbol{\theta},\mu_0)} \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)-\boldsymbol{S}_n(\boldsymbol{\theta},\mu_0)\boldsymbol{\lambda}+\frac{1}{(1-\max\limits_{1\leq t\leq n}|\gamma_t(\boldsymbol{\theta},\mu_0)|)}\frac{1}{n}\sum_{t=1}^{n}\Big(\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\gamma_t^2(\boldsymbol{\theta},\mu_0)\Big) \\
&\leq \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)-\boldsymbol{S}_n(\boldsymbol{\theta},\mu_0)\boldsymbol{\lambda}+\frac{\max\limits_{1\leq t\leq n}|\gamma_t(\boldsymbol{\theta},\mu_0)|^2}{(1-\max\limits_{1\leq t\leq n}|\gamma_t(\boldsymbol{\theta},\mu_0)|)}\Big(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)\Big) \\
&=\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)-\boldsymbol{S}_n(\boldsymbol{\theta},\mu_0)\boldsymbol{\lambda}+o_p(1)O_p(\frac{1}{\sqrt{n}}) \\
&=\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0)-\boldsymbol{S}_n(\boldsymbol{\theta},\mu_0)\boldsymbol{\lambda}+o_p(\frac{1}{\sqrt{n}}).
\end{aligned}$$

Because $\boldsymbol{S}_n^{-1}(\boldsymbol{\theta},\mu_0)\geq C$ in probability by Lemma 3 (d), we have

$$\boldsymbol{\lambda}=\boldsymbol{S}_n^{-1}(\boldsymbol{\theta},\mu_0)(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta},\mu_0))+\boldsymbol{L}_n \text{ with } \|\boldsymbol{L}_n\|=o_p(\frac{1}{\sqrt{n}}).$$

By Taylor expansion and $o_p(|\gamma_t(\boldsymbol{\theta},\mu_0)|^2)=o_p(1)$, we have

$$\ln(1+\gamma_t(\boldsymbol{\theta},\mu_0))=\gamma_t(\boldsymbol{\theta},\mu_0)-\frac{\gamma_t^2(\boldsymbol{\theta},\mu_0)}{2}+R_t(\boldsymbol{\theta},\mu_0),$$

where $R_t(\boldsymbol{\theta}, \mu_0)$ is the remainder term. This gives that

$$
\begin{aligned}
l(\boldsymbol{\theta}, \mu_0) &= -2 \ln L(\boldsymbol{\theta}, \mu_0) \\
&= 2 \sum_{t=1}^{n} \ln\{1 + \gamma_t(\boldsymbol{\theta}, \mu_0)\} \\
&= 2\boldsymbol{\lambda}'(\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)) - n\boldsymbol{\lambda}'\boldsymbol{S}_n(\boldsymbol{\theta}, \mu_0)\boldsymbol{\lambda} + 2\sum_{t=1}^{n} R_t(\boldsymbol{\theta}, \mu_0) \\
&= 2n(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0))'\boldsymbol{S}_n^{-1}(\boldsymbol{\theta}, \mu_0)(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)) + 2\boldsymbol{L}_n'(\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)) \\
&\quad - n(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0))'\boldsymbol{S}_n^{-1}(\boldsymbol{\theta}, \mu_0)(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)) - \boldsymbol{L}_n'(\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)) \\
&\quad - n\boldsymbol{L}_n'\boldsymbol{S}_n(\boldsymbol{\theta}, \mu_0)\boldsymbol{L}_n - (\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0))'\boldsymbol{L}_n + 2\sum_{t=1}^{n} R_t(\boldsymbol{\theta}, \mu_0) \\
&= (\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0))'\boldsymbol{S}_n^{-1}(\boldsymbol{\theta}, \mu_0)(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)) - n\boldsymbol{L}_n'\boldsymbol{S}_n(\boldsymbol{\theta}, \mu_0)\boldsymbol{L}_n + 2\sum_{t=1}^{n} R_t(\boldsymbol{\theta}, \mu_0).
\end{aligned}
$$

Because $n\boldsymbol{L}_n'\boldsymbol{S}_n(\boldsymbol{\theta}, \mu_0)\boldsymbol{L}_n = o_p(1)$ by noting that $\|\boldsymbol{L}_n\| = o_p(\frac{1}{\sqrt{n}})$, and the last term satisfies

$$
|2\sum_{t=1}^{n} R_t(\boldsymbol{\theta}, \mu_0)| \leq 2B\sum_{t=1}^{n} |\gamma_t(\boldsymbol{\theta}, \mu_0)|^3 = o_p(1), \tag{5.10}
$$

it follows that

$$
\begin{aligned}
l(\boldsymbol{\theta}, \mu_0) &= (\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0))'\boldsymbol{S}_n^{-1}(\boldsymbol{\theta}, \mu_0)(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)) + o_p(1) \\
&= (\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0))'\boldsymbol{\Omega}^{-1}(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)) + o_p(1)
\end{aligned}
$$

holds uniformly in $\boldsymbol{\theta} \in V_0$ by part(d) of Lemma 3. Therefore,

$$
l(\boldsymbol{\theta}_0, \mu_0) = (\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}_0, \mu_0))'\boldsymbol{\Omega}^{-1}(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}_0, \mu_0)) + o_p(1).
$$

It follows from part (d) of Lemma 3 that

$$
\begin{aligned}
& \sum_{t=1}^{n} \|\boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)\|^3 \\
\leq \;& \max_{1 \leq t \leq n} \|\boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)\| \left(\sum_{t=1}^{n} \|\boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)\|^2 \right) = o_p(\sqrt{n}) \left(\sum_{t=1}^{n} \|\boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)\|^2 \right) \\
= \;& o_p(\sqrt{n}) \cdot n \cdot (\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{D}'_t(\boldsymbol{\theta}, \mu_0) \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)) = o_p(\sqrt{n}) \cdot n \cdot tr(\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{D}'_t(\boldsymbol{\theta}, \mu_0) \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)) \\
= \;& o_p(\sqrt{n}) \cdot n \cdot tr(\frac{1}{n} \sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0) \boldsymbol{D}'_t(\boldsymbol{\theta}, \mu_0)) = o_p(\sqrt{n}) \cdot n \cdot tr(\boldsymbol{\Omega} + o_p(1)) \\
= \;& o_p(\sqrt{n}) \cdot n \cdot O_p(1) = o_p(n\sqrt{n}),
\end{aligned}
$$

where tr denotes the trace of a matrix. Thus, we have $\|\boldsymbol{\lambda}\|^3 = O_p(n^{-\frac{3}{2}})$ and

$$
\sum_{t=1}^{n} |\gamma_t(\boldsymbol{\theta}, \mu_0)|^3 \leq \sum_{t=1}^{n} \|\boldsymbol{\lambda}\|^3 \|\boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)\|^3 = O_p(n^{-\frac{3}{2}}) \sum_{t=1}^{n} \|\boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)\|^3 = O_p(n^{-\frac{3}{2}}) o_p(n\sqrt{n}) = o_p(1).
$$

Define $\boldsymbol{\Delta}_n(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{D}_t(\boldsymbol{\theta}, \mu_0)$. Then,

$$
\begin{aligned}
l(\boldsymbol{\theta}, \mu_0) - l(\boldsymbol{\theta}_0, \mu_0) = & (\boldsymbol{\Delta}_n(\boldsymbol{\theta}) - \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0))' \boldsymbol{\Omega}^{-1} \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0) + \boldsymbol{\Delta}'_n(\boldsymbol{\delta}_0) \boldsymbol{\Omega}^{-1} (\boldsymbol{\Delta}_n(\boldsymbol{\theta}) - \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0)) \\
& + (\boldsymbol{\Delta}_n(\boldsymbol{\theta}) - \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0))' \boldsymbol{\Omega}^{-1} (\boldsymbol{\Delta}_n(\boldsymbol{\theta}) - \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0)) + o_p(1)
\end{aligned}
$$

holds uniformly for all $\boldsymbol{\theta} \in V_0$. Further,

$$
\boldsymbol{\Delta}_n(\boldsymbol{\theta}) - \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Big[(\boldsymbol{D}_{t,1}(\boldsymbol{\theta}) - \boldsymbol{D}_{t,1}(\boldsymbol{\theta}_0))', D_{t,2}(\boldsymbol{\theta}, \mu_0) - D_{t,2}(\boldsymbol{\theta}_0, \mu_0) \Big]'.
$$

Let $\boldsymbol{\Psi}_{11} = \frac{1}{4} E\{ \frac{w_t}{h_t^2(\boldsymbol{\theta}_0)} \frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial h_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \}$ be the first block of $\boldsymbol{\Psi}$. Then, by part (c) of Lemma 3 and Taylor expansion, we have

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\boldsymbol{D}_{t,1}(\boldsymbol{\theta}) - \boldsymbol{D}_{t,1}(\boldsymbol{\theta}_0)) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\frac{\partial \boldsymbol{D}_{t,1}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(\frac{1}{\sqrt{n}})) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\frac{\partial \boldsymbol{D}_{t,1}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \frac{\boldsymbol{\nu}}{\sqrt{n}} + o_p(\frac{1}{\sqrt{n}})) \\
&= \boldsymbol{\nu} (\frac{1}{n} \sum_{t=1}^{n} \frac{\partial \boldsymbol{D}_{t,1}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}) + o_p(1) \\
&= \boldsymbol{\Psi}_{11} \boldsymbol{\nu} + o_p(1)
\end{aligned}
$$

holds uniformly for $\boldsymbol{\theta} \in V_0$. Similarly, put $\boldsymbol{\Psi}_{21} = E\{\frac{w_t}{\sqrt{h_t(\boldsymbol{\theta}_0)}} \frac{\partial \varepsilon_t(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\theta}'}\}$, which is the second and the first block of $\boldsymbol{\Psi}$. Then,

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (D_{t,2}(\boldsymbol{\theta}, \mu_0) - D_{t,2}(\boldsymbol{\theta}_0, \mu_0)) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\frac{\partial D_{t,2}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(\frac{1}{\sqrt{n}})) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\frac{\partial D_{t,2}(\boldsymbol{\theta}_0, \mu_0)}{\partial \boldsymbol{\theta}'} \frac{\boldsymbol{\nu}}{\sqrt{n}} + o_p(\frac{1}{\sqrt{n}})) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\frac{\partial D_{t,2}(\boldsymbol{\theta}_0, \mu_0)}{\partial \boldsymbol{\theta}'} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o_p(\frac{1}{\sqrt{n}})) \\
&= \boldsymbol{\nu}(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial D_{t,2}(\boldsymbol{\theta}_0, \mu_0)}{\partial \boldsymbol{\theta}'}) + o_p(1) \\
&= \Psi_{21} \boldsymbol{\nu} + o_p(1)
\end{aligned}$$

holds uniformly for $\boldsymbol{\theta} \in V_0$.

Put $\boldsymbol{A} = (\boldsymbol{\Psi}_{11}, \boldsymbol{\Psi}_{21})$. It follows that

$$\boldsymbol{\Delta}_n(\boldsymbol{\theta}) - \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0) = (\boldsymbol{\Psi}'_{11}, \boldsymbol{\Psi}'_{21})' \boldsymbol{\nu} + o_p(1) = \boldsymbol{A}' \boldsymbol{\nu} + o_p(1)$$

holds uniformly for $\boldsymbol{\theta} \in V_0$, implying that

$$l(\boldsymbol{\theta}, \mu_0) - l(\boldsymbol{\theta}_0, \mu_0) = \boldsymbol{\nu}' \boldsymbol{A} \boldsymbol{\Omega}^{-1} \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0) + \boldsymbol{\Delta}'_n(\boldsymbol{\theta}_0) \boldsymbol{\Omega}^{-1} \boldsymbol{A}' \boldsymbol{\nu} + \boldsymbol{\nu}' \boldsymbol{A}' \boldsymbol{\Omega}^{-1} \boldsymbol{\nu} \boldsymbol{A}' + o_p(1) \tag{5.11}$$

holds uniformly for $\boldsymbol{\theta} \in V_0$. Like the proof of Lemma 1 of Qin and Lawless (1994), we know that the minimizer $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \frac{\boldsymbol{\nu}}{\sqrt{n}}$ must lie in V_0, i.e., $\hat{\boldsymbol{\nu}} = -(\boldsymbol{A} \boldsymbol{\Omega}^{-1} \boldsymbol{A}')^{-1} \boldsymbol{A} \boldsymbol{\Omega}^{-1} \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0) + o_p(1)$, which implies that

$$l(\hat{\boldsymbol{\theta}}, \mu_0) = [\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0)]' [\boldsymbol{I} - \boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{A}' (\boldsymbol{A} \boldsymbol{\Omega}^{-1} \boldsymbol{A}')^{-1} \boldsymbol{A} \boldsymbol{\Omega}^{-\frac{1}{2}}] [\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0)] + o_p(1).$$

By Lemma 4, $\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{\Delta}_n(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate standard normal distribution function, implying that $l(\hat{\boldsymbol{\theta}}, \mu_0) \xrightarrow{d} \chi^2_1$ by noting that

$$\begin{aligned}
tr(\boldsymbol{I} - \boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{A}' (\boldsymbol{A} \boldsymbol{\Omega}^{-1} \boldsymbol{A}')^{-1} \boldsymbol{A} \boldsymbol{\Omega}^{-\frac{1}{2}}) &= tr(\boldsymbol{I}) - tr(\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{A}' (\boldsymbol{A} \boldsymbol{\Omega}^{-1} \boldsymbol{A}')^{-1} \boldsymbol{A} \boldsymbol{\Omega}^{-\frac{1}{2}}) \\
&= p + q + s + r + 3 - rank(\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{A}') \\
&= 1.
\end{aligned}$$

□

Proof of Theorem 2. Following the proof of Theorem 1, we have

$$l(\hat{\boldsymbol{\theta}},0) = [\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\Delta}_n^*(\boldsymbol{\theta}_0)]'[\boldsymbol{I} - \boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{A}'(\boldsymbol{A}\boldsymbol{\Omega}^{-1}\boldsymbol{A}')^{-1}\boldsymbol{A}\boldsymbol{\Omega}^{-\frac{1}{2}}][\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\Delta}_n^*(\boldsymbol{\theta}_0)] + o_p(1), \qquad (5.12)$$

where $\boldsymbol{\Delta}_n^*(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta}_{\mathbf{0}},0)$.

Because $\mu_0 = M/\sqrt{n}$, we have

$$\begin{aligned}
\boldsymbol{\Delta}_n^*(\boldsymbol{\theta}_0) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta}_0,0) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\boldsymbol{D}'_{t,1}(\boldsymbol{\theta}_0), D_{t,2}(\boldsymbol{\theta}_0,0)) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[(\boldsymbol{D}'_{t,1}(\boldsymbol{\theta}_0), w_t(\eta_t-\mu_0))' + (\mathbf{0}, \frac{M}{\sqrt{n}}w_t)'\right] \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\boldsymbol{D}'_{t,1}(\boldsymbol{\theta}_0), w_t(\eta_t-\mu_0))' + \frac{1}{n}\sum_{t=1}^{n}(\mathbf{0}, Mw_t)' \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta}_0,\mu_0) + \frac{1}{n}\sum_{t=1}^{n}(\mathbf{0}, Mw_t)',
\end{aligned}$$

where $\mathbf{0}$ is a $(p+q+s+r+2)$-vector. By Lemma 4, we have $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{D}_t(\boldsymbol{\theta}_0,\mu_0) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Omega})$. By the weak laws of large number for stationary series, $\frac{M}{n}\sum_{t=1}^{n}w_t \xrightarrow{p} MEw_t$. Thus, $\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\Delta}_n^*(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution with mean $-(\mathbf{0}', MEw_t)'$ and covariance $\boldsymbol{I}$. By (5.12), we have $l(\hat{\boldsymbol{\theta}},0)$ converges to a noncentral chi-squared limit with one degree of freedom and the noncentral parameter $M^2(Ew_t)^2$ as $n \to \infty$. $\square$

Data Availability Statement

From the Federal Housing Finance Agency, we collect the state-level house price index (HPI) of California, Florida, Nevada, and Arizona from 1975 to 2018.

From Yahoo Finance, we collect the daily S&P500 and Microsoft Stock (MSFT) close prices from August 3, 2009 to July 29, 2019.