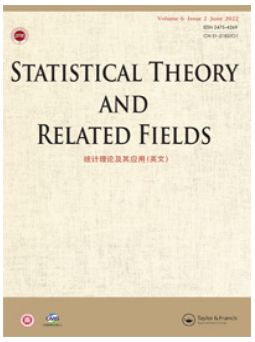

Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Empirical likelihood inference in autoregressive models with time-varying variances

Yu Han & Chunming Zhang

STATISTICAL THEORY AND RELATED FIELDS
2022, VOL. 6, NO. 2, 129–138
https://doi.org/10.1080/24754269.2021.1913977

Taylor & Francis
Taylor & Francis Group

Empirical likelihood inference in autoregressive models with time-varying variances

Yu Han [a] and Chunming Zhang [b]

[a]College of Science, Northeast Electric Power University, Jilin, People's Republic of China; [b]Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA

ABSTRACT
This paper develops the empirical likelihood (EL) inference procedure for parameters in autoregressive models with the error variances scaled by an unknown nonparametric time-varying function. Compared with existing methods based on non-parametric and semi-parametric estimation, the proposed test statistic avoids estimating the variance function, while maintaining the asymptotic chi-square distribution under the null. Simulation studies demonstrate that the proposed EL procedure (a) is more stable, i.e., depending less on the change points in the error variances, and (b) gets closer to the desired confidence level, than the traditional test statistic.

ARTICLE HISTORY
Received 25 December 2020
Revised 11 March 2021
Accepted 5 April 2021

KEYWORDS
Empirical likelihood; autoregressive model; unconditional heteroscedasticity; stable test

1. Introduction

In the literature of the macroeconomics and financial applications, the assumption of heteroscedasticity in many time series models revealed the facts that ignoring the issue of heteroscedasticity often leads to the inefficient estimation and unreliable inference. Thus, heteroscedasticity has been focused mainly on the effect of violations of homoscedasticity, usually in two forms, 'conditional heteroscedasticity' and 'unconditional heteroscedasticity'.

Non-constant volatility will be identified by 'conditional heteroscedasticity', when future periods of high and low volatility cannot be identified. Bollerslev (1986) and Engle (1982) proposed ARCH or GARCH models and provided the efficient estimation of the mean function by quasi-maximum likelihood based on other adaptive procedures. More complicated GARCH models had been proposed to allow for conditional heteroscedasticity, for instance, varying coefficient GARCH models (see Polzehl & Spokoiny, 2006) and spline GARCH models (see Engle & Rangel, 2008). The time-varying volatility is often used to describe the conditional heteroscedasticity. Drees and Starica (2002) and Starica (2003) made use of a non-stationary framework to analyse time series of S&P 500 returns, and found that this approach outperformed the GARCH-type models.

'Unconditional heteroscedasticity' will be used, when variables that have identifiable seasonal variability, such as electricity usage, are discussed. Hansen (1995) considered the linear regression model with deterministically trending regressors only, in which the error is an AR(p) process scaled by a continuous function of time. Nesting autoregressive model is also a special case when the conditional error variance of the model is a function of a covariate that has a form of a nearly integrated stochastic process with no deterministic drift. For the constant coefficient autoregressive model with time-varying variances (ARTV) which will be discussed in this article, Phillips and Xu (2006) utilised the ordinary least squares method and the nonparametric estimation of the variance function to provide three heteroscedasticity-robust test statistics, and proved their asymptotic standard normal distributions. Xu and Phillips (2008) proposed the heteroscedasticity-robust adaptive estimation for ARTV. Meanwhile, performances of methods in Phillips and Xu (2006) and Xu and Phillips (2008) relied on appropriately selecting the bandwidth used in the non-parametric function estimation.

Motivated from the 'empirical likelihood' (EL) approach, this article aims to develop a test statistic which is more stable, namely, depending less on the change points in the error variances, and avoiding the problem of selecting the bandwidth. In the literature, the EL approach was introduced by Owen (1988), Owen (1990) and Owen (1991) to construct confidence intervals in a nonparametric setting, which can be seen in Owen (2001). Since an EL approach possesses nonparametric properties, the distribution for the data is not required to be specified, and meanwhile more efficient estimates of the parameters can be yielded. The EL approach allows data to decide the shape of confidence regions without estimating the

CONTACT Chunming Zhang cmzhang@stat.wisc.edu

variance of the test statistic, and also is Bartlett correctable in DiCiccio et al. (1991). The EL approach has been applied to various situations, such as generalised linear models in Kolaczyk (1994), local linear smoother in Chen and Qin (2000), partially linear models in Shi and Lau (2000), parametric and semi-parametric models in multi response regression in Chen and Ingrid (2009); linear regression with censored data in Zhou and Li (2008), plug-in estimates of nuisance parameters in estimating equations in the context of survival analysis in Li and Wang (2003) and Qin and Jing (2001), heteroscedastic partially linear models in Lu (2009); GARCH models in Chan and Ling (2006); variable selection in Han et al. (2013) and Variyath and Chen (2010); analysis of longitudinal data in Qiu and Wu (2015). Qin and Lawless (1994) linked the EL with finitely many estimating equations, which served as finitely many equality constraints. To the best of our knowledge, there is no existing published work in the literature using the EL approach in the constant coefficient autoregressive models with time-varying variances. This article will also consider the constant coefficient autoregressive models with time-varying innovation variance by using the EL approach.

The remainder of the paper proceeds as follows. Section 2 describes the autoregressive model with time-varying variances and discusses main assumptions. Section 3 reviews the existing methods. Section 4 develops the empirical likelihood inference procedure with theoretical guarantees. Section 5 conducts simulation studies to evaluate the finite sample performance of the proposed method when compared with alternative methods. Section 6 briefly concludes. Technical details and proofs of the main results are relegated to Appendix.

2. Autoregressive model with time-varying variances

The constant coefficient autoregressive model with time-varying variances is described as follows,

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + u_t \\ &= \boldsymbol{X}_{t-1}^\top \boldsymbol{\beta}_o + u_t, \qquad (1) \\ u_t &= \sigma_t \varepsilon_t, \quad t = 1, \dots, T, \qquad (2) \end{aligned}$$

where $\top$ denotes transpose, $\boldsymbol{X}_{t-1} = (1, Y_{t-1}, \dots, Y_{t-p})^\top \in \mathbb{R}^{p+1}$ is the vector of covariates, and $\boldsymbol{\beta}_o = (\beta_0, \beta_1, \dots, \beta_p)^\top \in \mathbb{R}^{p+1}$ is the true parameter vector of interest, with $\beta_p \neq 0$, and the lag order p finite and known. We assume that $\{\sigma_t\}$ is a deterministic sequence of time t, satisfying

$$\sigma_t = g(t/T), \qquad (3)$$

and $\{\varepsilon_t\}$ is a martingale difference sequence with respect to $\mathscr{F}_t$, where $\mathscr{F}_t = \sigma(\varepsilon_s : s \le t)$ is the σ-field generated by $\{\varepsilon_s : s \le t\}$ with $\mathrm{E}(\varepsilon_t^2 \mid \mathscr{F}_{t-1}) = 1$, a.s., for all t. Thus, the conditional variance of $\{u_t\}$ is fully characterised by the multiplicative factor σ_t in (2), i.e.,

$$\mathrm{E}(u_t^2 \mid \mathscr{F}_{t-1}) = \sigma_t^2 = g^2(t/T), \quad \text{a.s..} \qquad (4)$$

Suppose that the data are generated from models (1)–(2), and we observe a sample containing $T+p$ observations, denoted by $\{Y_{-p+1}, Y_{-p+2}, \dots, Y_0, Y_1, \dots, Y_T\}$. The main goals are to make inferences about the true parameter vector $\boldsymbol{\beta}_o$ in models (1)–(2), i.e., testing the null hypothesis,

$$H_0 : \boldsymbol{\beta}_o = \boldsymbol{b}_0, \qquad (5)$$

where $\boldsymbol{b}_0 = (b_{0,0}, b_{0,1}, \dots, b_{0,p}) \in \mathbb{R}^{p+1}$, and constructing a confidence region for $\boldsymbol{\beta}_o$.

Section 4 will present our proposed empirical likelihood inference, after Section 3 describes the estimation methods in Phillips and Xu (2006).

To facilitate the discussion of main results and comparison with related existing methods, the following conditions provided in Phillips and Xu (2006); Xu and Phillips (2008) are considered.

Conditions

(A1) $g(\cdot)$ in (3) and (4) is a measurable and strictly positive function on the interval $(0, 1]$ such that $0 < \inf_{r\in(0,1]} g(r) \le \sup_{r\in(0,1]} g(r) < \infty$, and $g(r)$ satisfies a Lipschitz condition except at a finite number of points of discontinuity;

(A2) Suppose that L is the lag operator. Then $0 = 1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_p L^p$ has all roots outside the unit circle;

(A3) $\{\varepsilon_t\}$ satisfies $\mathrm{E}(\varepsilon_t \mid \mathscr{F}_{t-1}) = 0$, and $\mathrm{E}(\varepsilon_t^2 \mid \mathscr{F}_{t-1}) = 1$, a.s., for all t;

(A4) $\sup_t \mathrm{E}(|\varepsilon_t^{4\nu}|) < \infty$ for some $\nu > 1$.

Remark 2.1: (i) In condition (A1), the function g is integrable on the interval $(0, 1]$ to any finite order. For brevity, we write $\int_0^1 g^m(x)\,\mathrm{d}x$ as $\int g^m$ for any finite positive integer $m \ge 1$.

(ii) Condition (A2) satisfies the stability conditions which, for a constant $g(\cdot)$ and homoskedastic $\{\varepsilon_t\}$, would ensure that $\{Y_t\}$ is stationary or asymptotically covariance-stationary. Under condition (A2), the mean μ of Y_t is given by

$$\mu = \frac{\beta_0}{1 - \beta_1 - \cdots - \beta_p},$$

and Y_t has the Wold representation,

$$Y_t = \mu + \sum_{i=1}^{\infty} \alpha_i u_{t-i},$$

where $\{\alpha_i\}$ satisfies that

$$\alpha_i - \beta_1 \alpha_{i-1} - \cdots - \beta_p \alpha_{i-p} = 0, \quad \text{as } i > 0,$$

and $\sum_{i=1}^{\infty} |\alpha_i| < \infty$. Define Ω to be the matrix with the (i,j)-th element $\gamma_{|i-j|}$, where $\gamma_k = \sum_{i=0}^{\infty} \alpha_i \alpha_{i+k} < \infty$.

(iii) Condition (A3) ensures that $\{\varepsilon_t\}$ is a martingale difference sequence and, at the same time, stipulates $\mathrm{E}(u_t^2 \mid \mathscr{F}_{t-1}) = g^2(t/T)$ doesn't depend on the past events, in other words, models (1)–(2) are unconditional heteroscedastic.

3. Existing methods

Regarding the estimation of $\boldsymbol{\beta}_{\mathrm{o}}$ in models (1)–(2), Phillips and Xu (2006) reviewed the ordinary least squares (OLS) estimator $\hat{\boldsymbol{\beta}}$, and showed that under the stated conditions, as $T \to \infty$,

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{\mathrm{o}}) = \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{X}_{t-1}^{\top} \boldsymbol{X}_{t-1} \right)^{-1} \times \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{X}_{t-1}^{\top} \varepsilon_t \right) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Lambda), \tag{6}$$

where $\xrightarrow{\mathcal{D}}$ stands for converges in distribution, $\Lambda = \Omega_1^{-1}\Omega_2\Omega_1^{-1}$, Ω_1 and Ω_2 are defined as the $(p+1) \times (p+1)$ matrices,

$$\Omega_1 = \begin{pmatrix} 1 & \mu\, \mathbf{l}_p^{\top} \\ \mu\, \mathbf{l}_p & \mu^2 + (\int g^2)\Omega \end{pmatrix}, \quad \Omega_2 = \begin{pmatrix} (\int g^2) & \mu\,(\int g^2)\, \mathbf{l}_p^{\top} \\ \mu\,(\int g^2)\, \mathbf{l}_p & \mu^2\,(\int g^2) + (\int g^4)\Omega \end{pmatrix}, \tag{7}$$

$\mathbf{l}_p = (1, \ldots, 1)^{\top} \in \mathbb{R}^p$ is a vector of ones, and μ and Ω are as defined in Remark 2.1.

Since g is typically unknown, the asymptotic covariance matrix Λ in (6) must be estimated and this can be done in several ways. First, by applying the weighted sum of squared OLS residuals using kernel smoothing, originally proposed by Nadaraya (1964) and Watson (1964) for estimation of regression functions, they proposed the consistent estimator of the function $g^2(r)$ non-parametrically for $r \in [0, 1]$,

$$\hat{g}^2(r) = \sum_{t=1}^{T} w_{r,t}\, \hat{u}_t^2, \tag{8}$$

where $\hat{u}_t = Y_t - \boldsymbol{X}_{t-1}^{\top}\hat{\boldsymbol{\beta}}$ is the OLS residual and the weights $w_{r,t}$, $t = 1, \ldots, T$, are defined as

$$w_{r,t} = \left\{ \sum_{t=1}^{T} K\left(\frac{[Tr] - t}{Th_T} \right) \right\}^{-1} K\left(\frac{[Tr] - t}{Th_T} \right), \tag{9}$$

where the kernel function $K(\cdot) : \mathbb{R} \mapsto [0, \infty)$ is assumed to satisfy $0 \le K(z) \le C_1 < \infty$ uniformly in z and

$$\int_{-\infty}^{\infty} K(z)\, \mathrm{d}z < C_2 < \infty,$$

for some constant C_1 and C_2; h_T is a bandwidth parameter depending on T. The selection of bandwidth parameter h_T uses the cross-validation procedure, i.e., minimises the averaged squared prediction errors (see Wong, 1983),

$$\mathrm{CV}(b) = \frac{1}{T} \sum_{s=1}^{T} \{\hat{u}_s^2 - \hat{g}_{-s}^2(s/T)\}^2, \tag{10}$$

with respect to b, where

$$\hat{g}_{-s}^2(r) = \sum_{t=1,\, t \neq s} w_{r,t}\, \hat{u}_t^2.$$

Phillips and Xu (2006) suggested the following three consistent estimators of the asymptotic covariance matrix Λ when g is unknown.

- The first estimator of the asymptotic covariance matrix is

$$\hat{\Lambda}_1 = T \left(\sum_{t=1}^{T} \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}^{\top} \right)^{-1} \left(\sum_{t=1}^{T} \hat{u}_t^2 \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}^{\top} \right) \times \left(\sum_{t=1}^{T} \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}^{\top} \right)^{-1}. \tag{11}$$

- The second estimator of the asymptotic covariance matrix is

$$\hat{\Lambda}_2 = \hat{\Omega}_1^{-1} \left(\sum_{t=1}^{T} \hat{u}_t^2\, \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}^{\top} \right) \hat{\Omega}_1^{-1}, \tag{12}$$

  where the matrix $\hat{\Omega}_1$ is defined as

$$\hat{\Omega}_1 = \begin{pmatrix} 1 & \hat{\mu}\, \mathbf{l}_p^{\top} \\ \hat{\mu}\, \mathbf{l}_p & \hat{\mu}^2 + \left(T^{-1} \sum_{t=1}^{T} \hat{u}_t^2 \right) \hat{\Omega} \end{pmatrix},$$

  where $\hat{\mu}$ and $\hat{\Omega}$ correspond to replacing $\boldsymbol{\beta}_{\mathrm{o}}$, in the expressions of μ and Ω in Remark 2.1, with $\hat{\boldsymbol{\beta}}$.
- The third estimator of the asymptotic covariance matrix is

$$\hat{\Lambda}_3 = \hat{\Omega}_1^{-1}\tilde{\Omega}_2\hat{\Omega}_1^{-1}, \tag{13}$$

  where the matrix $\hat{\Omega}_2$ is defined as

$$\tilde{\Omega}_2 = \begin{pmatrix} \int \hat{g}^2 & \hat{\mu}(\int \hat{g}^2)\, \mathbf{l}_p^{\top} \\ \hat{\mu}(\int \hat{g}^2)\, \mathbf{l}_p & \hat{\mu}^2(\int \hat{g}^2) + (\int \hat{g}^4)\hat{\Omega} \end{pmatrix}.$$

Based on the above three estimators $\hat{\Lambda}_j$ of the true covariance matrix Λ, Phillips and Xu (2006) constructed three test statistics t_j, $j = 1, 2, 3$, for the true parameter vector $\boldsymbol{\beta}_{\mathrm{o}}$, stated as follows.

Lemma 3.1 (Theorem 2(ii) in Phillips and Xu (2006)): *Assume that $\hat{\boldsymbol{\beta}}$ is the OLS estimator of $\boldsymbol{\beta}_{\mathrm{o}}$. Then, under* *the above assumptions and null hypothesis* (5), *it follows that*

$$t_j = \frac{\sqrt{T}(\hat{\beta}_k - b_{0,k})}{((\hat{\Lambda}_j)_{kk})^{1/2}} \xrightarrow{\mathcal{D}} N(0,1), \quad \text{as } T \to \infty, \tag{14}$$

where $(\hat{\Lambda}_j)_{kk}$ *is the* (k,k)*-th element of the matrix* $\hat{\Lambda}_j$, $j = 1, 2, 3$, *defined in* (11), (12) *and* (13), *respectively.*

Hence, a large sample level $100(1-\alpha)\%$ confidence region for β_o based on the above Normal approximation (14) is given by

$$\Re_{j,\alpha} = \{\boldsymbol{b} : T(\hat{\boldsymbol{\beta}} - \boldsymbol{b})^\top \times [\mathrm{diag}\{(\hat{\Lambda}_j)_{k,k}, k = 0, 1, \ldots, p\}]^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{b}) \le \chi^2_{p;1-\alpha}\}, \tag{15}$$

where $\mathrm{diag}\{(\hat{\Lambda}_j)_{k,k}, k = 0, 1, \ldots, p\}$ is the main diagonal matrix of $\hat{\Lambda}_j$, $j = 1, 2, 3$, and $\chi^2_{p;1-\alpha}$ denotes the $100(1-\alpha)$th quantile of the chi-square distribution χ^2_p with p degrees of freedom.

4. Proposed method

In terms of the practical performance of the three tests t_j in (14), however, simulation results reveal two major issues arising from the estimation of the asymptotic covariance matrix and the selection of the bandwidth. In order to solve these problems, the proposed empirical likelihood approach will be applied to test parameters in models (1)–(2).

To construct an empirical likelihood function, the estimation equations will be defined by means of,

$$\boldsymbol{W}_t(\boldsymbol{b}) = \boldsymbol{X}_{t-1} \cdot (Y_t - \boldsymbol{X}_{t-1}^\top \boldsymbol{b}), \tag{16}$$

for a generic model parameter $\boldsymbol{b} \in \mathbb{R}^{p+1}$. According to condition (A3), we have that

$$\mathrm{E}(\boldsymbol{W}_t(\boldsymbol{\beta}_o)) = \mathrm{E}(\boldsymbol{X}_{t-1}\, g(t/T)\, \varepsilon_t) = g(t/T)\, \mathrm{E}(\boldsymbol{X}_{t-1}\, \varepsilon_t) = \boldsymbol{0}$$

holds for the true parameter vector $\boldsymbol{\beta}_o$. Based on (16), we define the empirical likelihood for the parameter $\boldsymbol{b}$ by

$$L(\boldsymbol{b}) = \sup\left\{\prod_{t=1}^T q_t : \sum_{t=1}^T q_t = 1, \sum_{t=1}^T q_t \boldsymbol{W}_t(\boldsymbol{b}) = \boldsymbol{0}\right\}.$$

By using the Lagrange multiplier, we have

$$\hat{q}_t(\boldsymbol{b}) = \frac{1}{T}\{1 + \hat{\boldsymbol{\lambda}}^\top \boldsymbol{W}_t(\boldsymbol{b})\}^{-1}, \quad t = 1, \ldots, T,$$

where $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}(\boldsymbol{b}) \in \mathbb{R}^{p+1}$ is the solution of equations,

$$\frac{1}{T}\sum_{t=1}^T \frac{\boldsymbol{W}_t(\boldsymbol{b})}{1 + \hat{\boldsymbol{\lambda}}^\top \boldsymbol{W}_t(\boldsymbol{b})} = \boldsymbol{0}. \tag{17}$$

We also note that $\prod_{t=1}^T q_t$, subject to constraints $q_t \ge 0$ and $\sum_{t=1}^T q_t = 1$, attains its maximum $(1/T)^T$ at $q_t = 1/T$. Thus, the empirical likelihood ratio at $\boldsymbol{b}$ is defined by

$$\mathrm{ELR}(\boldsymbol{b}) = \prod_{t=1}^T \{\hat{q}_t(\boldsymbol{b})\, T\}^{-1} = \prod_{t=1}^T \{1 + \hat{\boldsymbol{\lambda}}^\top \boldsymbol{W}_t(\boldsymbol{b})\}.$$

Taking the log transformation of the above equation, we get the corresponding empirical log-likelihood ratio,

$$\ell(\boldsymbol{b}) = 2\sum_{t=1}^T \log\{1 + \hat{\boldsymbol{\lambda}}^\top \boldsymbol{W}_t(\boldsymbol{b})\}. \tag{18}$$

In addition, Theorem 4.1 below provides the asymptotic null distribution of $\ell(\boldsymbol{\beta}_o)$.

Theorem 4.1: *Assume that conditions* (A1)–(A4) *hold. Then, under the null hypothesis* (5), *the limiting distribution of* $\ell(\boldsymbol{\beta}_o)$ *is the chi-square distribution with* p *degrees of freedom, i.e.,*

$$\ell(\boldsymbol{\beta}_o) \xrightarrow{\mathcal{D}} \chi^2_p, \quad \text{as } T \to \infty. \tag{19}$$

According to Theorem 4.1, the empirical likelihood ratio confidence interval for the true value $\boldsymbol{\beta}_o$ can be constructed as follows:

$$\Re_{\mathrm{EL},\alpha} = \{\boldsymbol{b} : \ell(\boldsymbol{b}) \le \chi^2_{p;1-\alpha}\}, \tag{20}$$

where $\chi^2_{p;1-\alpha}$ is defined below (15). Combined with (20), Theorem 4.1 implies Corollary 4.1.

Corollary 4.1: *Under the conditions of Theorem 4.1,*

$$\mathbb{P}(\boldsymbol{\beta}_o \in \Re_{\mathrm{EL},\alpha}) \to 1 - \alpha, \quad \text{as } T \to \infty.$$

5. Simulation evaluation

In this section, simulation studies are conducted to compare the finite sample performance of five methods described in Sections 3–4:

Ordinary least squares without the heteroscedasticity correction (OLS),
t_1, t_2, t_3,
the proposed empirical likelihood (EL) procedure.

The zero-mean AR(1) with the time-varying variance is considered as follows:

$$Y_t = \beta_{0,1} Y_{t-1} + g(t/T)\varepsilon_t,$$

where $\{\varepsilon_t\} \overset{\text{i.i.d.}}{\sim} N(0,1)$. The kernel function $K(\cdot)$ is the standard Normal density function,

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad -\infty < x < \infty,$$

and the bandwidth parameter is selected by the cross-validation criterion (10). We consider $H_0 : \beta_{0,1} = \beta_1$ with known values of β_1.

Three kinds of the variance functions $g^2(r)$ are considered in the following simulations: a single abrupt point model, two abrupt points model, continuous function variance model as follows.

Model 1: A single abrupt point model,

$$g^2(r) = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\,\mathrm{I}_{\{r \geq \kappa\}}, \quad r \in [0,1].$$

Model 1 corresponds to the case of a single abrupt change of the error variance from σ_0^2 to σ_1^2 at time $[\kappa T]$, where κ is the break point within the value set $\{0.1, 0.5, 0.9\}$. The ratio of post-break and pre-break standard deviations $\delta = \sigma_1/\sigma_0$ is within the value set $\{0.2, 1, 5\}$ where $\sigma_0 = 1$.

Model 2: Two abrupt points model,

$$g^2(r) = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\,\mathrm{I}_{\{\kappa_0 < r \leq \kappa_1\}} + (\sigma_2^2 - \sigma_0^2)\,\mathrm{I}_{\{\kappa_1 < r\}}, \quad r \in [0,1].$$

Model 2 corresponds to the case of two abrupt points model which has the change of the error variance from σ_0^2 to σ_1^2 and σ_1^2 to σ_2^2. The time break points (κ_0, κ_1) take the values $(0.1, 0.9)$; $(\sigma_0^2, \sigma_1^2, \sigma_2^2)$ are from the set $\{(0.2, 5, 0.2), (5, 0.2, 5)\}$.

Model 3: Continuous function variance model,

$$g^2(r) = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2) r^m, \quad r \in [0,1].$$

Model 3 considers that the variance of the errors is the continuous function from σ_0^2 to σ_1^2. We suppose m to be within the value set $\{1, 2\}$ and $\delta = \sigma_1/\sigma_0$ within the value set $\{0.2, 5\}$ where $\sigma_0^2 = 1$.

Model 1 and Model 3 are the same as in Cavaliere (2004), Cavaliere and Taylor (2007) and Phillips and Xu (2006). Simulations are done when the parameter of interest β_1 increases on the set $\{0.1, 0.5, 0.9\}$, and the nominal size is 5%. The sample size T is from $\{60, 200\}$ respectively. The number of Monte Carlo replications is 5000.

Simulation results include two parts. The first part displayed in Tables 1, 2 and 3 assesses the rejection rates of five methods under the null hypothesis.

The second part includes Figures 1–3 to evaluate the rejection rates of methods OLS, t_1, t_2, t_3 and EL as the parameter β_1 under the alternatives increases.

From these simulations, we draw the following conclusions.

(a) First, the OLS-based test is the inefficient and unreliable test under the heteroscedastic innovations. From Table 1, the OLS-based test overrejects overwhelmingly the null hypothesis when the null is true, and has the largest distorted size under $(\kappa, \delta) \in \{(0.1, 0.2), (0.9, 5)\}$. In addition, the distorted size doesn't reduce except for the homoscedastic innovations with the increasing sample size which is also shown in Figures

Table 1. Comparison of the rejection rates of five methods in Model 1 for $\beta_1 \in \{0.1, 0.5, 0.9\}$, $\kappa \in \{0.1, 0.5, 0.9\}$, $\delta \in \{0.2, 1, 5\}$ and the sample size $T \in \{60, 200\}$, based on 5000 replications.

			$T = 60$					$T = 200$				
β_1	κ	δ	OLS	t_1	t_2	t_3	EL	OLS	t_1	t_2	t_3	EL
0.1	0.1	0.2	0.3386	0.1428	0.0560	0.1010	0.1724	0.3804	0.0972	0.0676	0.0840	0.0958
		1	0.0558	0.0692	0.0560	0.0450	0.0658	0.0436	0.0510	0.0476	0.0414	0.0492
		5	0.0644	0.0712	0.0628	0.0520	0.0654	0.0594	0.0558	0.0536	0.0464	0.0530
	0.5	0.2	0.1480	0.0812	0.0486	0.0506	0.0750	0.1528	0.0636	0.0524	0.0590	0.0610
		5	0.1526	0.0846	0.0682	0.0642	0.0798	0.1422	0.0608	0.0566	0.0528	0.0550
	0.9	0.2	0.0666	0.0716	0.0520	0.0432	0.0700	0.0588	0.0552	0.0498	0.0468	0.0524
		5	0.3400	0.1632	0.1624	0.1258	0.1712	0.3780	0.0966	0.0940	0.0754	0.0956
0.5	0.1	0.2	0.3146	0.1278	0.1066	0.1004	0.1632	0.3826	0.0924	0.0886	0.0884	0.0940
		1	0.0528	0.0666	0.0762	0.0514	0.0646	0.0494	0.0554	0.0648	0.0540	0.0536
		5	0.0622	0.0682	0.0822	0.0548	0.0644	0.0606	0.0564	0.0682	0.0560	0.0542
	0.5	0.2	0.1452	0.0766	0.0830	0.0608	0.0752	0.1600	0.0650	0.0692	0.0640	0.0614
		5	0.1396	0.0792	0.1014	0.0748	0.0736	0.1416	0.0572	0.0664	0.0552	0.0540
	0.9	0.2	0.0654	0.0692	0.0720	0.0474	0.0658	0.0630	0.0590	0.0636	0.0558	0.0574
		5	0.3232	0.1462	0.1952	0.1388	0.1638	0.3760	0.0948	0.01154	0.0796	0.0970
0.9	0.1	0.2	0.1954	0.0836	0.2328	0.1144	0.1278	0.3038	0.0664	0.1926	0.0914	0.0730
		1	0.0544	0.0650	0.1950	0.1172	0.0600	0.0548	0.0584	0.1160	0.0756	0.0560
		5	0.0596	0.0656	0.2086	0.1286	0.0604	0.0638	0.0570	0.1212	0.0792	0.0512
	0.5	0.2	0.1250	0.0724	0.2288	0.1334	0.0714	0.1478	0.0654	0.1492	0.0980	0.0636
		5	0.1442	0.0768	0.2552	0.1632	0.0728	0.1412	0.0580	0.1578	0.0982	0.0560
	0.9	0.2	0.0618	0.0646	0.1844	0.1116	0.0618	0.0650	0.0588	0.1162	0.0808	0.0568
		5	0.2778	0.1136	0.3962	0.2100	0.1454	0.3402	0.0780	0.2406	0.1108	0.0862

Table 2. Comparison of the rejection rates of five methods in Model 2 for $\beta_1 \in \{0.1, 0.5, 0.9\}$, $[\kappa_0, \kappa_1] = [0.1, 0.9]$, $[\sigma_0, \sigma_1, \sigma_2] \in \{[0.2, 5, 0.2], [5, 0.2, 5]\}$ and the sample size $T \in \{60, 200\}$, based on 5000 replications.

				$T = 60$					$T = 200$				
β_1	σ_0	σ_1	σ_2	OLS	t_1	t_2	t_3	EL	OLS	t_1	t_2	t_3	EL
0.1	0.2	5	0.2	0.0520	0.0888	0.0278	0.0234	0.0712	0.0438	0.0546	0.0360	0.0340	0.0496
	5	0.2	5	0.0570	0.0712	0.0628	0.0492	0.0664	0.0444	0.0504	0.0480	0.0418	0.0488
0.5	0.2	5	0.2	0.0514	0.0762	0.0408	0.0320	0.0714	0.0510	0.0570	0.0472	0.0426	0.0540
	5	0.2	5	0.0548	0.0704	0.0812	0.0498	0.0544	0.0480	0.0558	0.0650	0.0548	0.0540
0.9	0.2	5	0.2	0.0550	0.0698	0.1302	0.0806	0.0640	0.0552	0.0566	0.0976	0.0676	0.0536
	5	0.2	5	0.0522	0.0644	0.1996	0.1178	0.0598	0.0540	0.0598	0.1172	0.0752	0.0572

1 and 3. From Table 2, the OLS-based test has better performance than Table 1, however, the rejection rate reduces as the sample size increases. The results of the OLS-based test in Table 3 are similar to those in Table 1.

(b) Second, the performance of t_2 and t_3 depends on the numerical value of the true parameter and the pattern of the variance function. From Figures 1, 2, 3, an interesting phenomenon can be found that the rejection rates of t_2 and t_3 are

Table 3. Comparison of the rejection rates of five methods in Model 3 for $\beta_1 \in \{0.1, 0.5, 0.9\}$, $m \in \{1, 2\}$, $\delta \in \{0.2, 5\}$ and the sample size $t \in \{60, 200\}$, based on 5000 replications.

			$T = 60$					$T = 200$				
β_1	m	δ	OLS	t_1	t_2	t_3	EL	OLS	t_1	t_2	t_3	EL
0.1	1	0.2	0.0820	0.0812	0.0604	0.0446	0.0826	0.0820	0.0550	0.0476	0.0518	0.0546
		5	0.0780	0.0688	0.0624	0.0542	0.0632	0.0828	0.0582	0.0564	0.0540	0.0566
	2	0.2	0.0756	0.0784	0.0548	0.0498	0.0628	0.0676	0.0544	0.0482	0.0464	0.0520
		5	0.1200	0.0854	0.0778	0.0688	0.0716	0.1200	0.0620	0.0596	0.0582	0.0586
0.5	1	0.2	0.0800	0.0772	0.0780	0.0474	0.0782	0.0912	0.0620	0.0674	0.0620	0.0620
		5	0.0872	0.0800	0.0926	0.0670	0.0682	0.0804	0.0560	0.0674	0.0564	0.0544
	2	0.2	0.0766	0.0728	0.0728	0.0550	0.0656	0.0752	0.0602	0.0666	0.0598	0.0600
		5	0.1196	0.0806	0.0104	0.0844	0.0718	0.1212	0.0604	0.0734	0.0640	0.0590
0.9	1	0.2	0.0672	0.06640	0.1972	0.1112	0.0644	0.0830	0.0642	0.1268	0.0822	0.0606
		5	0.0789	0.0688	0.2238	0.1356	0.0578	0.0854	0.0564	0.1322	0.0850	0.0558
	2	0.2	0.0652	0.0676	0.1866	0.1166	0.0562	0.0732	0.0640	0.1198	0.0818	0.0628
		5	0.1258	0.0826	0.2572	0.1686	0.0668	0.1196	0.0622	0.1500	0.0958	0.0608

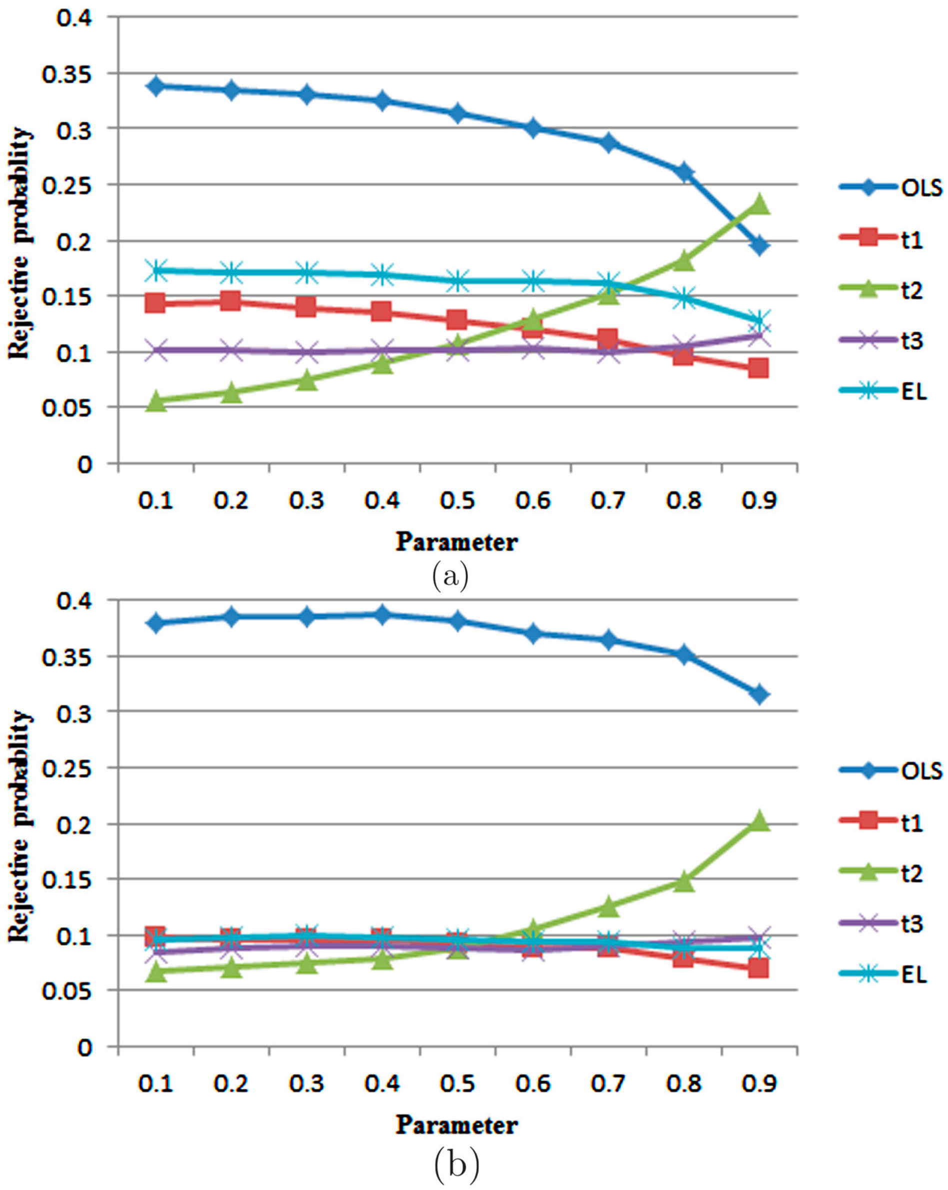

Figure 1. The relationship between the rejection rates of OLS, t_1, t_2, t_3, EL and the true coefficient β_1 in Model 1 (a single abrupt point model). The abrupt point $\kappa = 0.1$, $\delta = 0.2$. The true parameter β_1 increases gradually from 0.1 to 0.9. (a) The sample $T = 60$; (b) the sample $T = 200$.

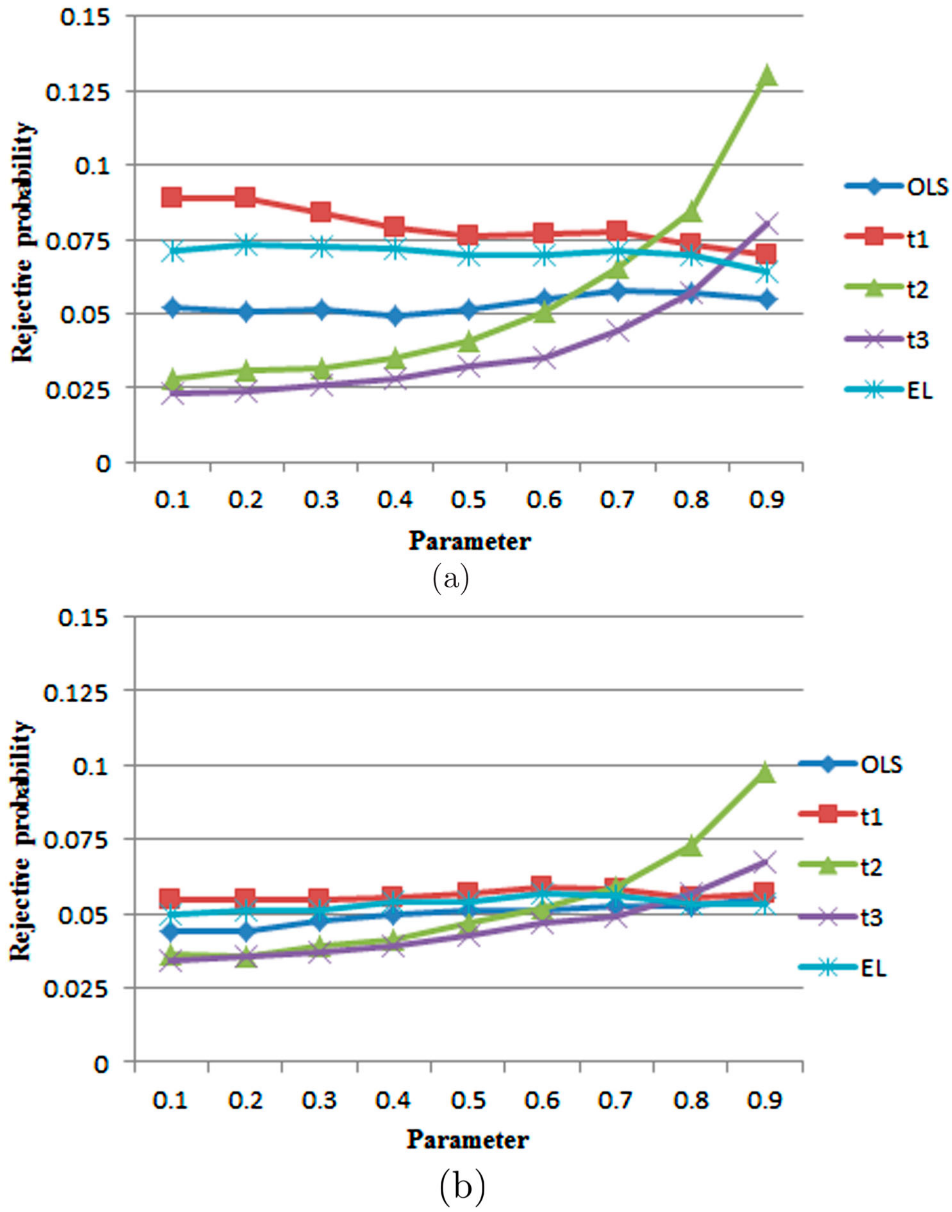

Figure 2. The relationship between the rejection rates of OLS, t_1, t_2, t_3, EL and the true coefficient β_1 in Model 2 (two abrupt points model). The abrupt points $\kappa_1 = 0.1$, $\kappa_2 = 0.9$, $[\sigma_0, \sigma_1, \sigma_2] = [0.2, 5, 0.2]$. The true parameter β_1 increases gradually from 0.1 to 0.9. (a) The sample $T = 60$; (b) the sample $T = 200$.

likely to be an increasing function of the parameter and grow bigger as $\beta_1 > 0.5$. The rejection rate of t_2 is far greater than the nominal size 5% when the numerical value of the parameter is close to unity, namely $\beta_1 = 0.9$. In particular, it is easy to see that t_2 and t_3 overaccept the null hypothesis when the parameter is less than or equal to 5% when $\beta_1 < 0.5$. On the contrary, t_2 and t_3 over-reject the null hypothesis when $\beta_1 > 0.9$. It also has the similar conclusions from Tables 1–3. So both t_2 and t_3 aren't the stable test for the ARTV model.

(c) Third, both EL and t_1 are the stable tests for the ARTV model and EL outperforms t_1. From Tables 1–3, we can find that EL and t_1 overreject the null hypothesis when the null is true. From Figures 1–3, the rejection rate of EL is almost a horizontal line and is closer to the nominal level 5% than t_1 except Figure 1(a) when the sample size is 60. When the sample size is 200, EL's rejection rate is nearly a nominal size of 5% and doesn't depend on the numerical value of the parameters even if the true value of β is close to unity ($\beta_1 = 0.9$). EL has the smallest size distortion overall and avoids correcting the variance. The simulation results generally support the asymptotic results. EL is more stable and has better performance than OLS, t_1, t_2, t_3 for testing the parameters of ARTV. So EL seems to be the better choice.

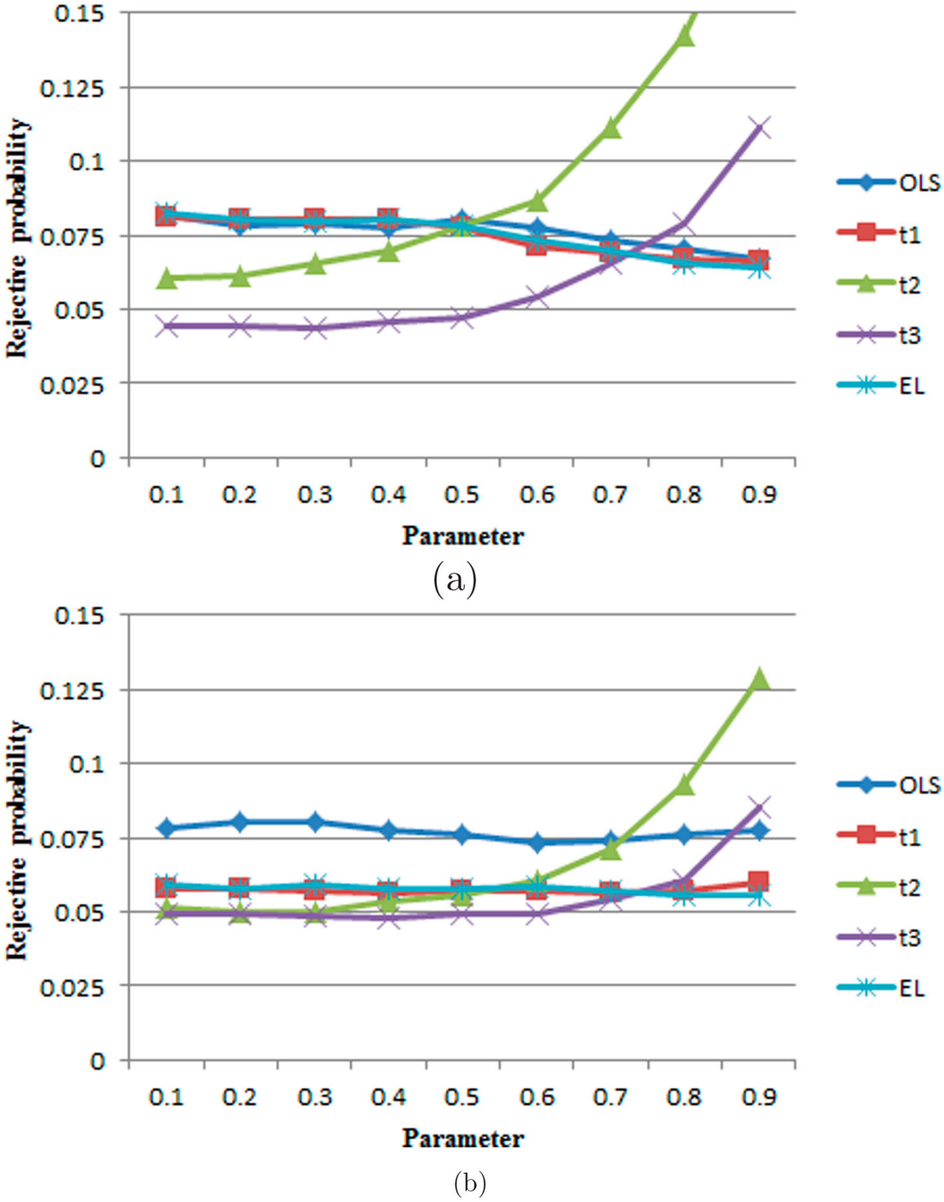

Figure 3. The relationship between the rejection rates of OLS, t_1, t_2, t_3, EL corresponding to the true coefficient β_1 in Model 3 (continuous function variance model), and $m = 1, \delta = 0.2$. The true parameter β_1 increases gradually from 0.1 to 0.9. (a) The sample $T = 60$; (b) the sample $T = 200$.

6. Conclusion

This article focuses on the empirical likelihood appro ach for autoregressive models with error terms scaled by an unknown nonparametric time-varying function. The empirical likelihood ratio test statistic avoids estimating the unknown variance function, in the presence of heteroscedastic error terms. The results of simulations of three different models show that the empirical likelihood is more stable than the other four test statistics. In addition, some extensions include improving the efficiency of statistic based on the different equations, and locating the abrupt time points when they exist.

Acknowledgments

The authors thank the editor, Prof. Jun Shao, and two anonymous reviewers for helpful comments. Yu Han was supported by the Scientific Research Foundation of Jilin Education (JJKH20200102KJ). The work of C. Zhang was partially supported by U.S. National Science Foundation grants DMS-2013486 and DMS-1712418, and provided by the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors thank the editor, Prof. Jun Shao, and two anonymous reviewers for helpful comments. Yu Han was supported by the Scientific Research Foundation of Jilin Education [grant number JJKH20200102KJ]. The work of C. Zhang was partially supported by U.S. National Science Foundation [grant numbers DMS-2013486 and DMS-1712418], and provided by the University of Wisconsin-Madison Office of the

Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

Notes on contributors

Yu Han, received Ph.D. degree in mathematical statistics from Jilin University in 2012. He is currently an associate research fellow in the Educational Supervision and Evaluation Center of Northeast Electrical Power University. His current research interests are Time Series Analysis, Nonparametric and semi-parametric estimation & inference. He worked at Department of Statistics, Wisconsin University-Madison between 2013 and 2014 as a visiting scholar. He has published 12 papers. He has accomplished 3 projects as a principle investigator and as a participator. One was accomplished, and two are ongoing.

Chunming Zhang is Professor of Statistics at the University of Wisconsin-Madison. Her research interests range from statistical learning and data mining, statistical methods with applications to imaging data, neuroinformatics and bioinformatics, multiple testing, large-scale simultaneous inference and applications, statistical methods in financial econometrics, non- and semi-parametric estimation and inference, to functional and longitudinal data analysis.

ORCID

Yu Han http://orcid.org/0000-0002-7728-6217
Chunming Zhang http://orcid.org/0000-0002-3153-2662

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Cavaliere, G. (2004). Unit root tests under time-varying variance shifts. *Econometric Reviews*, *23*, 259–292. https://doi.org/10.1081/ETC-200028215

Cavaliere, G., & Taylor, A. M. R. (2007). Testing for unit roots in time series models with nonstationary volatility. *Journal of Econometrics*, *140*(2), 919–947. https://doi.org/10.1016/j.jeconom.2006.07.019

Chan, N. H., & Ling, S. Q. (2006). Empirical likelihood for Garch models. *Econometric Theory*, *3*, 403–428. https://doi.org/10.1017/S0266466606060208

Chen, S. X., & Ingrid, V. K. (2009). A review on empirical likelihood methods for regression. *Test*, *18*(3), 415–447. https://doi.org/10.1007/s11749-009-0159-5

Chen, S. X., & Qin, Y. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika*, *87*, 946–953. https://doi.org/10.1093/biomet/87.4.946

DiCiccio, T., Hall, P., & Romano, J. (1991). Empirical likelihood is Bartlett-Correctable. *Annals of Statistics*, *19*(2), 1053–1061. https://doi.org/10.1214/aos/1176348137

Drees, H., & Starica, C. (2002). *A simple non-stationary model for stock returns* (Working paper). Chalmers University of Technology.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, *50*, 987–1008. https://doi.org/10.2307/1912773

Engle, R. F., & Rangel, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies*, *21*(3), 1187–1222. https://doi.org/10.1093/rfs/hhn004

Han, Y., Jin, Y. H., & Chen, M. (2013). Empirical likelihood-based subset selection for partially linear autoregressive models. *Acta Mathematicae Applicatae Sinica, English Series*, *29*(4), 793–808. https://doi.org/10.1007/s10255-013-0256-9

Hansen, B. E. (1995). Regression with nonstationary volatility. *Econometrica*, *63*, 1113–1132. https://doi.org/10.2307/2171723

Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statistica Sinica*, *4*, 199–218. http://www3.stat.sinica.edu.tw/statistica/oldpdf/A4n111.pdf

Li, G., & Wang, Q. H. (2003). Empirical likelihood regression analysis for right censored data. *Statistica Sinica*, *13*, 51–68. https://www.jstor.org/stable/24307094?seq=1

Lu, X. W. (2009). Empirical likelihood for heteroscedastic partially linear models. *Journal of Multivariate Analysis*, *100*, 387–396. https://doi.org/10.1016/j.jmva.2008.05.006

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications*, *9*(1), 141–142. https://doi.org/10.1137/1109020

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*, 237–249. https://doi.org/10.1093/biomet/75.2.237

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics*, *18*, 90–120. https://doi.org/10.1214/aos/1176347494

Owen, A. B. (1991). Empirical likelihood for linear models. *Annals of Statistics*, *19*(4), 1725–1747. https://doi.org/10.1214/aos/1176348368

Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall.

Phillips, P. C. B., & Xu, K. L. (2006). Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis*, *27*, 289–308. https://doi.org/10.1111/jtsa.2006.27.issue-2

Polzehl, J., & Spokoiny, V. (2006). *Varying coefficient GARCH versus local constant volatility modeling: Comparison of predictive power* (Working paper). Weierstrass Institute for Applied Analysis and Stochastics.

Qiu, J., & Wu, L. (2015). A moving blocks empirical likelihood method for longitudinal data. *Biometrics*, *71*, 616–624. https://doi.org/10.1111/biom.12317

Qin, G., & Jing, B. Y. (2001). Empirical likelihood for censored linear regression. *Scandinavian Journal of Statistics*, *28*, 661–673. https://doi.org/10.1111/sjos.2001.28.issue-4

Qin, J., & Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, *22*, 300–325. https://doi.org/10.1214/aos/1176325370

Shi, J., & Lau, T. S. (2000). Empirical likelihood for partially linear models. *Journal of Multivariate Analysis*, *72*(1), 132–148. https://doi.org/10.1006/jmva.1999.1866

Starica, C. (2003). *Is GARCH (1,1) as good a model as the Nobel prize accolades would imply* (Working paper). Chalmers University of Technology.

Xu, K. L., & Phillips, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics*, *142*, 265–280. https://doi.org/10.1016/j.jeconom.2007.06.001

Variyath, A. M., & Chen, J. H. (2010). Abraham B. Empirical likelihood based variable selection. *Journal of Statistical Planning and Inference*, *140*, 971–981. https://doi.org/10.1016/j.jspi.2009.09.025

Watson, G. S. (1964). Smooth regression analysis. *Sankhya Series A*, *26*, 359–372.

Wong, W. H. (1983). On the consistency of cross validation in kernel nonparametric regression. *Annals of Statistics*, *11*, 1136–1141. https://doi.org/10.1214/aos/1176346327

Zhou, M., & Li, G. (2008). Empirical likelihood analysis of the Buckley-James estimator. *Journal of Multivariate Analysis*, *99*, 649–664. https://doi.org/10.1016/j.jmva.2007.02.007

Appendix. Proofs of main results

Before proving Theorem 4.1, we first show Lemmas A.1–A.2. To simplify notations, we denote $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}(\boldsymbol{\beta}_o)$ and $\boldsymbol{W}_t = \boldsymbol{W}_t(\boldsymbol{\beta}_o)$.

Lemma A.1: *Assume that conditions (A1)–(A4) hold. Then*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_t \xrightarrow{\mathcal{D}} N(\mathbf{0},\Omega_2), \tag{A1}$$

$$\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{W}_t\boldsymbol{W}_t^\top \xrightarrow{\text{P}} \Omega_2, \tag{A2}$$

where $\xrightarrow{\text{P}}$ *denotes converges in probability.*

Proof: According to Phillips and Xu (2006) (Lemma 1(iii)–(iv)), the proof of Lemma A.1 completes. ■

Lemma A.2: *Assume that conditions (A1)–(A3) hold. Then*

$$\hat{\boldsymbol{\lambda}} = O_{\mathbb{P}}(T^{-1/2}).$$

Proof: From (17), we have

$$\mathbf{0} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{W}_t - \frac{1}{T}\sum_{t=1}^{T}\frac{\boldsymbol{W}_t\boldsymbol{W}_t^\top}{1+\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t}\hat{\boldsymbol{\lambda}}.$$

By (A1) of Lemma 3.1,

$$\frac{\|\hat{\boldsymbol{\lambda}}\|_2}{1+\|\hat{\boldsymbol{\lambda}}\|_2\max_t\|\boldsymbol{W}_t\|_2}\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{W}_t\boldsymbol{W}_t^\top\right\| \le \left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{W}_t\right\|_2 = O_{\mathbb{P}}(T^{-1/2}).$$

According to conditions (A1) and (A4), we have $\mathrm{E}(|Y_t|^{4\nu}) < \infty$ for some $\nu > 1$, and then

$$\begin{aligned}\max_t\|\boldsymbol{W}_t\|_2 &= \max_t\|\boldsymbol{X}_{t-1}(Y_t-\boldsymbol{\beta}_o^\top\boldsymbol{X}_{t-1})\|_2 = \max_t\|\boldsymbol{X}_{t-1}u_t\|_2\\ &= \max_t\|\boldsymbol{X}_{t-1}g(t/T)\varepsilon_t\|_2 = o_{\mathbb{P}}(T^{\frac{1}{4\nu}}).\end{aligned} \tag{A3}$$

From (A2) of Lemma A.1 and a similar argument used in Owen (1991), the proof of Lemma A.2 is completed. ■

Proof: Noticing that if $\boldsymbol{\beta}_o$ is the true parameters, applying Taylor's expansion to (18), we have

$$\begin{aligned}\ell(\boldsymbol{\beta}_o) &= 2\sum_{t=1}^{T}\log(1+\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t)\\ &= 2\sum_{t=1}^{T}\left\{\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t - \frac{1}{2}(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t)^2\right\} + r_T,\end{aligned} \tag{A4}$$

where r_T, in probability, satisfies the following inequality in light of Lemma A.1 (A2) and Lemma A.2 for some constant $C > 0$,

$$|r_T| \le C\sum_{t=1}^{T}|\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t|^3 \le C\|\hat{\boldsymbol{\lambda}}\|_2^3\max_{1\le t\le T}\|\boldsymbol{W}_t\|_2\sum_{t=1}^{T}\|\boldsymbol{W}_t\|_2^2 = o_{\mathbb{P}}(1).$$

By Lemma A.1 (A2), Lemma A.2 and similar arguments as above, we have

$$\sum_{t=1}^{T}\frac{(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t)^3}{1+\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t} = o_{\mathbb{P}}(1). \tag{A5}$$

By (17), we obtain

$$\begin{aligned}0 &= \sum_{t=1}^{T}\frac{\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t}{1+\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t} = \sum_{t=1}^{T}(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t) - \sum_{t=1}^{T}(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t)^2\\ &\quad+\sum_{t=1}^{T}\frac{(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t)^3}{1+\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t}.\end{aligned} \tag{A6}$$

By (A5) and (A6), we obtain

$$\sum_{t=1}^{T}(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t) = \sum_{t=1}^{T}(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t)^2 + o_{\mathbb{P}}(1). \tag{A7}$$

Again by (17), we obtain

$$\begin{aligned}\mathbf{0} &= \sum_{t=1}^{T}\frac{\boldsymbol{W}_t}{1+\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t} = \sum_{t=1}^{T}\boldsymbol{W}_t\left\{1-\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t+\frac{(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t)^2}{1+\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t}\right\}\\ &= \sum_{t=1}^{T}\boldsymbol{W}_t - \sum_{t=1}^{T}(\boldsymbol{W}_t\boldsymbol{W}_t^\top)\hat{\boldsymbol{\lambda}} + \sum_{t=1}^{T}\frac{\boldsymbol{W}_t\,(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t)^2}{1+\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t}.\end{aligned}$$

By Lemma A.1 and (A3), we have

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{\boldsymbol{W}_t(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t)^2}{1+\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t}\right\|_2\\ &\le C\|\hat{\boldsymbol{\lambda}}^2\|_2\max_t\|\boldsymbol{W}_t\|_2\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{W}_t\|_2^2 = o_{\mathbb{P}}(T^{-1/2}).\end{aligned}$$

Thus, we have

$$\begin{aligned}\hat{\boldsymbol{\lambda}} &= \left(\sum_{t=1}^{T}\boldsymbol{W}_t\boldsymbol{W}_t^\top\right)^{-1}\sum_{t=1}^{T}\boldsymbol{W}_t + \left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{W}_t\boldsymbol{W}_t^\top\right)^{-1}\\ &\quad\times\left\{\frac{1}{T}\sum_{t=1}^{T}\frac{\boldsymbol{W}_t\,(\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t)^2}{1+\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t}\right\}\\ &= \left(\sum_{t=1}^{T}\boldsymbol{W}_t\boldsymbol{W}_t^\top\right)^{-1}\sum_{t=1}^{T}\boldsymbol{W}_t + o_{\mathbb{P}}(T^{-1/2}).\end{aligned}$$

By substituting $\hat{\boldsymbol{\lambda}}$ of the above equation into (A4) and (A7), we have

$$\begin{aligned}\ell(\boldsymbol{\beta}_o) &= \sum_{t=1}^{T}\hat{\boldsymbol{\lambda}}^\top\boldsymbol{W}_t\boldsymbol{W}_t^\top\hat{\boldsymbol{\lambda}} + o_{\mathbb{P}}(1)\\ &= \left(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{W}_t\right)^\top\left(T^{-1}\sum_{t=1}^{T}\boldsymbol{W}_t\boldsymbol{W}_t^\top\right)^{-1}\\ &\quad\times\left(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{W}_t\right) + o_{\mathbb{P}}(1).\end{aligned}$$

The proof of Theorem 4.1 is completed by using Lemma A.1. ■