A unified test for the intercept of a predictive regression model

XIAOHUI LIU,† YUZI LIU,† YAO RAO¶ AND FUCAI LU§

† *School of Statistics, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China(email: liuxiaohui@jxufe.edu.cn)*

¶ *Department of Economics, Management School, Chatham Street, The University of Liverpool, Liverpool L69 72H, UK(email: Y.Rao@liverpool.ac.uk)*

§ *School of Business Administration, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China(email: lu-fucai@263.net)*

Abstract

Testing the predictability of the predictive regression model is of great interest in economics and finance. Recently, Zhu et al. (2014) proposed a unified test to account for this issue. Their test has a desirable property that its limit distribution is standard regardless of the regressor being stationary, near unit root or unit root. However, this test depends on, a priori, whether there is an intercept in the predictive regression while this is usually unknown in practice. In this paper, using empirical likelihood inference, we develop a unified pretest for the intercept, as a pretest to determine the choice of the predictability test. Simulations studies confirm that the proposed pretest works well. Two real data examples are also provided to illustrate the importance of such pretest. The first revisits the S&P 500 index data and the second investigates stock return predictability and investor sentiment for six countries.

JEL classification: C12; C22; G1

Key words: Predictive regression model; Empirical likelihood; Intercept; Unified test.

(Total word: 6326)

* Our thanks go to Professor Liang Peng for stimulating this research and ShanShan Ping for doing some simulation studies. We also deeply grateful to referees for valuable suggestions and comments. Xiaohui Liu's research is supported by NSF of China (Grant No.11971208, 11601197), China Postdoctoral Science Foundation funded project (2016M600511, 2017T100475), the Postdoctoral Research Project of Jiangxi (No.2017KY10, xskt19393), NSF of Jiangxi Province (No.2018ACB21002, 20171ACB21030). Yuzi Liu's research is partly supported by the Postgraduate Innovation Project of Jiangxi Province (No.YC2019-S216). Fucai Lu's research is supported by NSF of China (No.71863015, 71463020)

I Introduction

As an important tool for modeling the relationship between a dependent variable and the lagged value of a regressor, predictive regression models have been widely used in economics and finance, especially when low frequency data are present. Seminal results about these models include Campbell and Yogo (2006), Cai and Wang (2014), and Kostakis et al. (2015), among others. See, e.g., the excellent summary Phillips (2015) and references therein for details.

A typical predictive regression model has usually the following linear structure:

$$\begin{cases} Y_t = \alpha + \beta X_{t-1} + U_t, \\ X_t = \mu + \phi X_{t-1} + \varepsilon_t, \\ B(L)\varepsilon_t = V_t, \end{cases} \tag{1}$$

where Y_t denotes the dependent variable such as the asset return at time t, and X_{t-1} denotes a financial variable, e.g., the log dividend-price ratio at time $t-1$. The initial value $X_0 = o_p(\sqrt{T})$ independent of $\{V_s\}_{s=1}^T$, $B(L) = 1 - (\sum_{i=1}^p b_i L^i)$ with $L^i \varepsilon_t = \varepsilon_{t-i}$, $B(1) \neq 0$, all roots of $B(L)$ are fixed and less than one in absolute value, and $\{(U_t, V_t)\}_{t=1}^T$ are independent and identically distributed (i.i.d.) random vectors with zeros means.

In practice, an interesting issue on this model is to test the predictability, i.e., the null hypothesis $H_0 : \beta = 0$, of Y_t by some lagged regressor X_{t-1}. It is known that the tests involving in the ordinary least squares estimator of β have different limit distributions. They depend on (i) whether $|\phi| < 1$ (stationary), or $\phi = 1$ (unit root), or $\phi = 1 + c/T$ (near unit root) for some $c \neq 0$, and (ii) whether μ is included or excluded in the AR process, and (iii) whether there exists the so-called embedded endogeneity, i.e., X_t and U_t may be correlated. They may depend on some non-estimable parameters, say c, when $\{X_t\}$ is non-stationary; See, e.g., Cai and Wang (2014); Campbell and Yogo (2006) and references therein for details.

In view of this, it is important to develop some unified tests which are robust against (i) – (iii). For the AR(1) process X_t's, uniform inference procedures for ϕ have already been discussed by many authors; See So and Shin (1999), Mikusheva (2007), Chan et al. (2012) and Hill et al. (2016). Recently, Zhu et al. (2014) investigated the unified predictability test for β in (1) by empirical likelihood method. The most outstanding property of this test is its robustness against (i)-(iii). However, this method depends in prior on whether or not $\alpha = 0$ in (1). A further data splitting technique has to be employed to get rid of the impact of the intercept on the resulting testing statistic if $\alpha \neq 0$. Although both empirical likelihood methods have the same limit distribution, the data splitting technique leads to a great loss of power (Zhu et al., 2014). They even may lead to different conclusions over the same data set; See Section IV in the sequel for an example. In practice, whether or not $\alpha = 0$ is unknown, and needs to be tested.

Unfortunately, it is nontrivial to construct a unified method for testing the existence of intercept in a time series model regardless of whether the regressors are stationary or non-stationary. For the AR(1) model, testing $H_0^* : \mu = 0$ was studied by separately considering the case $\phi = 1$ and the case $|\phi| < 1$ in Dickey and Fuller (1981) and Fuller et al. (1981), respectively. Recently Dios-Palomares and Roldan (2006) proposed to test $H_0^* : \mu = 0$ by combining the tests in Dickey and Fuller (1981) and Fuller et al. (1981), which depends on a prior test on testing $H_0^{**} : \mu = 0$ & $\phi = 1$ and does not work for the case of $\phi = 1 + c/n$ with some $c \neq 0$.

Motivated by the unified empirical likelihood inference in Chan et al. (2012) and Zhu et al. (2014), in this paper we develop a similar unified empirical likelihood inference for α and studies its power property. The proposed empirical likelihood function can be employed to test $H_0 : \alpha = 0$ and construct an interval for α without knowing whether $|\phi| < 1$ or $\phi = 1$ or $\phi = 1 - c/T$ for some $c \neq 0$; We refer to Owen (2001) for an overview on empirical likelihood method. It turns out that if the proposed test fails to reject $H_0 : \alpha = 0$, we will be able to use the first empirical likelihood method of Zhu et al. (2014), which explicitly forces $\alpha = 0$, to test the predictability. The testing power can be improved because no data splitting technique is involved.

It is worth mentioning that Georgiev et al. (2019) recently considers an interesting issue concerning the stability of the intercept in a predictive regression model. Such test is designed by investigating the impact of Z_{t-1}, an omitted but either manifest or unobserved latent variable, on the predictability of X_{t-1} *to* Y_t. A fixed regressor wild bootstrap procedure is developed to test the null hypothesis of $\beta_z = 0$. A related study by Georgiev et al. (2018) further extends the discussion for the instability test in the predictive regression with regards to whether the parameters are constant over time to all or a subset of the parameters, not just intercept as in Georgiev et al. (2019).

While our present paper is relevant to both two studies in that we also investigate the issue of parameters in the predictive regression model context, in particular, our paper is similar to Georgiev et al. (2019) since we both focus on the intercept parameter, however, our paper is different from Georgiev et al. (2019) because the intercept term in Georgiev et al. (2019) is

$\alpha_t = \alpha + \beta_z Z_{t-1}$ when Z_{t-1} is not included, whereas our paper mainly examines the existence of the intercept of whether $\alpha = 0$ which is different from testing if $\beta_z = 0$ in Georgiev et al. (2019).

Another important difference is that both Georgiev et al. (2018) and Georgiev et al. (2019) consider the case where the predictive regressor X_t is a unit root or near unit root process; while we provide a unified framework, i.e., we allow for the possibilities that not only unit root, near unit root, but also the case when X_t is stationary. Hence, the tests developed in our paper is able to accommodate these three cases without requiring any a priori information regarding the degree of persistence in the predictors.

We organize the rest of this paper as follows. Section 2 presents the methodologies and main asymptotic results of our proposed test. Section 3 contains the finite sample simulation studies. Two empirical applications are discussed in Section 4 and Section 5 concludes. All proofs are provided in the Appendix.

II Methodology and main results

Suppose the samples $\{(X_t, Y_t)\}_{t=1}^T$ are generated from model (1). The interesting issue is to consider the hypothesis

$$H_0 : \alpha = 0 \qquad \text{versus} \qquad H_1 : \alpha \neq 0,$$

for the following cases: $|\phi| < 1$ (stationary), or $\phi = 1$ (unit root), or $\phi = 1 + c/T$ (near unit root) for some $c \neq 0$. Throughout this paper, we assume that $\{(U_t, V_t)\}_{t=1}^T$ are i.i.d. random vectors, and satisfy

- (C1). $E(|U_1|^{2+\epsilon} + |V_1|^{2+\epsilon}) < \infty$ for some $\epsilon > 0$.

Since the least squares estimators $\hat{\alpha}$, $\hat{\beta}$ are solutions to

$$\begin{cases} \frac{1}{n}\sum_{t=1}^{T}(Y_t - \alpha - X_{t-1}\beta) = 0 \\ \frac{1}{n}\sum_{t=1}^{T}(Y_t - \alpha - X_{t-1}\beta)X_{t-1} = 0 \end{cases} \tag{2}$$

with respect to α and β, one may construct an empirical likelihood based test from (2) as did in Qin and Lawless (1994). However, as mentioned in Chan et al. (2012), the quantity

$$\frac{1}{T}\sum_{t=1}^{T} E\left(\left(\begin{matrix}(Y_t - \alpha - X_{t-1}\beta) \\ (Y_t - \alpha - X_{t-1}\beta)\frac{1}{\sqrt{T}}X_{t-1}\end{matrix}\right)\left(\begin{matrix}(Y_t - \alpha - X_{t-1}\beta) \\ (Y_t - \alpha - X_{t-1}\beta)\frac{1}{\sqrt{T}}X_{t-1}\end{matrix}\right)^{\top}\middle|\, \mathcal{F}_{t-1}\right)$$

does not converge in probability when $\mu = 0$ and $\phi = 1 - c/T$ for some c under the null hypothesis, where $\mathcal{F}_t$ denotes the σ-field generated by $\{(U_s, V_s) : 1 \leq s \leq t\}$. We hence are not able to obtain the asymptotic normality of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \binom{Y_t - \alpha - X_{t-1}\beta}{(Y_t - \alpha - X_{t-1}\beta)\frac{1}{\sqrt{T}}X_{t-1}}$. Consequently, Wilks' theorem does not hold for the related log-empirical likelihood ratio. A similar problem exists in the empirical likelihood function relying on the estimating equations

$$\begin{cases} \frac{1}{T} \sum_{t=1}^{T} (Y_t - \alpha - X_{t-1}\beta) = 0 \\ \frac{1}{T} \sum_{t=1}^{T} (Y_t - \alpha - X_{t-1}\beta) \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} = 0. \end{cases}$$

This is because the quantity $\frac{1}{T} \sum_{t=1}^{T} \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}$ does not converge in probability when $\mu = 0$ and $\phi = 1 - c/T$ for some c, which results in that the joint limit of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} U_t$ and $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{U_t X_{t-1}}{\sqrt{1+X_{t-1}^2}}$ is no longer a bivariate normal distribution (Zhu et al., 2014).

Similar to Li et al. (2014), we overcome this problem by adding a pseudo sample. Put

$$\begin{cases} Z_{t1}(\alpha, \beta) = Y_t - \alpha - \beta X_{t-1} \\ Z_{t2}(\alpha, \beta) = (Y_t - \alpha - \beta X_{t-1}) \frac{X_{t-1}}{(1+X_{t-1}^2)^\gamma} + W_t = 0, \ t = 1, 2, \cdots, n, \end{cases} \tag{3}$$

where W_t's are independent of X_t and Y_t, and $W_t \sim N(0, \bar{\sigma}^2), t = 1, 2, \cdots, n$, are i.i.d. for some known $\bar{\sigma}^2 > 0$. In Li et al. (2014), γ is chosen to be 0.75, but here we suggest to use some $\gamma \in (\frac{1}{2}, \frac{3}{4})$ due to the reason mentioned in Remark 4 below. Note that as stated in Section 2.1 of Hill et al. (2016), the choice of $\gamma > \frac{1}{2}$ guarantees the asymptotic normality of the $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_{t2}(\alpha, \beta)$ in the nonstationary case; while $\gamma < \frac{3}{4}$ ensures the testing statistic to have nontrivial power. To balance between the power and size, we suggest a simple choice of γ, i.e., using the average of $\frac{1}{2}$ and $\frac{3}{4}$ in practice.

We consider the following empirical likelihood function

$$\ell(\alpha) = \sup_{\beta} L(\alpha, \beta),$$

where

$$L(\alpha, \beta) = \sup \left\{ \prod_{t=1}^{T} T p_t, \ p_t \geq 0, \ t = 1, 2, \cdots, T, \ \sum_{t=1}^{T} p_t = 1, \ \sum_{t=1}^{T} p_t \boldsymbol{\zeta}_t(\alpha, \beta) = \mathbf{0} \right\}$$

with $\boldsymbol{\zeta}_t(\alpha, \beta) = (Z_{t1}(\alpha, \beta), Z_{t2}(\alpha, \beta))^\top$, $t = 1, 2, \cdots, T$.

This profile empirical likelihood function follows directly from Hill et al. (2016), and mainly serves as a benchmark for the other methods conducted in the sequel. Note that for some

$\gamma \in (\frac{1}{2}, \frac{3}{4})$, both quantities $\frac{1}{T}\sum_{t=1}^T E((Y_t - \alpha - \beta X_{t-1})^2 \frac{X_{t-1}}{(1+X_{t-1}^2)^\gamma} | \mathcal{F}_{t-1})$ and $\frac{1}{T}\sum_{t=1}^T E((Y_t - \alpha - \beta X_{t-1})^2 \frac{X_{t-1}^2}{(1+X_{t-1}^2)^{2\gamma}} | \mathcal{F}_{t-1})$ vanish as $T \to \infty$ when $\mu = 0$ and $\phi = 1 - c/T$ for some c. It directly follows that $\frac{1}{T}\sum_{t=1}^T E\left(\boldsymbol{\zeta}_t(\alpha,\beta)\boldsymbol{\zeta}_t(\alpha,\beta)^\top \middle| \mathcal{F}_{t-1}\right)$ converges in probability. It is then easy to check the following result. Hereafter, we denote $(\alpha_0, \beta_0)^\top$ as the true value of $(\alpha, \beta)^\top$.

Theorem 1. *Suppose (C1) holds. Then under the null hypothesis H_0 we have*

$$-2\log \ell(\alpha_0) \xrightarrow{d} \chi_1^2, \quad \text{as } T \longrightarrow \infty,$$

for $\mu \in R$, regardless of $|\phi| < 1$ (stationary), or $\phi = 1$ (unit root), or $\phi = 1 + c/T$ (near unit root) for some $c \neq 0$, where '$\xrightarrow{d}$' denotes the convergence in distribution, and χ_1^2 is a chi-squared distributed random variable with one degree of freedom.

Theorem 1 indicates that $-2\log \ell(\alpha_0)$ has an asymptotic chi-squared distribution uniformly holding for all possible cases of X_t's mentioned above. One may reject the null hypothesis H_0 if $-2\log \ell(\alpha) \geq \chi_1^2(1-a)$ at the significance level $a \in (0,1)$, where $\chi_1^2(1-a)$ denotes the $(1-a)$-th quantile of χ_1^2 though.

However, the existence of random W_t's results in the value of $-2\log \ell(\alpha_0)$ to be random for given observations. How to choose $\bar{\sigma}^2$ is also not trivial. Furthermore, the power of this test to detect the local alterative hypothesis departing from H_0 may be further improved for given $\gamma \in (\frac{1}{2}, \frac{3}{4})$ as stated in the sequel.

In view of this, we propose to consider the following profile empirical likelihood function

$$\widetilde{\ell}(\alpha) = \sup_{\beta} \widetilde{L}(\alpha, \beta),$$

with

$$\widetilde{L}(\alpha,\beta) = \sup\left\{\prod_{t=1}^T Tp_t,\ p_t \geq 0,\ t = 1,2,\cdots,T,\ \sum_{t=1}^T p_t = 1,\ \sum_{t=1}^T p_t \widetilde{\boldsymbol{\zeta}}_t(\alpha,\beta) = \mathbf{0}\right\}.$$

Here $\widetilde{\boldsymbol{\zeta}}_t(\alpha,\beta) = (\widetilde{Z}_{t1}(\alpha,\beta), \widetilde{Z}_{t2}(\alpha,\beta))^\top$, $t = 1, 2, \cdots, T$, with

$$\begin{cases} \widetilde{Z}_{t1}(\alpha,\beta) &= Y_t - \alpha - \beta X_{t-1} \\ \widetilde{Z}_{t2}(\alpha,\beta) &= (Y_t - \alpha - \beta X_{t-1})\Big(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)} + (Y_{t-1} - \alpha - \beta X_{t-2})\Big). \end{cases}$$

The following result states the limit distribution of $-2\log \widetilde{\ell}(\alpha_0)$ under the null hypothesis H_0.

Theorem 2. *Under the same condition of Theorem 1, we have*

$$-2\log\widetilde{\ell}(\alpha_0) \xrightarrow{d} \chi_1^2, \quad \textit{as } T \longrightarrow \infty,$$

for $\mu \in R$, regardless of $|\phi| < 1$ (stationary), or $\phi = 1$ (unit root), or $\phi = 1 + c/T$ (near unit root) for some $c \neq 0$, where χ_1^2 is a chi-squared distributed random variable with one degree of freedom.

Remark 1. *Theorems 1 and 2 still hold, under the same conditions of Theorem 3.1 in Basrak et al. (2002), when*

$$U_t = \sigma_t e_t,\ \sigma_t^2 = \omega + \sum_{i=1}^{r} a_i e_{t-i}^2 + \sum_{j=1}^{s} b_j \sigma_{t-j}^2,$$

with $\{(e_t, V_t)\}_{t=1}^T$ being i.i.d. random vectors.

Remark 2. *Theorems 1 and 2 still hold when the predictor X_t follows the so-called augmented Dickey-Fuller model, i.e., ADF(p):*

$$X_t = \mu + \phi X_{t-1} + \sum_{i=1}^{p} \psi_i (X_{t-i} - X_{t-1-i}) + \varepsilon_t, \quad t = 1, \cdots, n,$$

when

$$\begin{cases} \textit{all roots of } 1 - \sum_{i=1}^p \psi_i x^i = 0 \textit{ are outside of the unit circle for } \phi = 1 + c/T \textit{ with } c \in R, \\ \textit{all roots of } 1 - \phi x - \sum_{i=1}^p \psi_i x^i + \sum_{i=1}^p \psi_i x^{i+1} = 0 \textit{ are outside of the unit circle for } |\phi| < 1. \end{cases}$$

Similar to Theorem 1 above, Theorem 2 indicates that Wilks' theorem holds for $-2\log\widetilde{\ell}(\alpha_0)$ too. Hence, we can reject the null hypothesis H_0 if $-2\log\widetilde{\ell}(\alpha) \geq \chi_1^2(1-a)$ at the significance level $a \in (0,1)$ by Theorem 2.

The results above only analyze the size property. For a test statistic, another issue of interest is its power property, especially its ability to detect the local alterative hypothesis departing from H_0. For the proposed tests above, we have the following theorem.

Theorem 3. *Suppose condition C1 holds and $d \neq 0$. Then, as $T \to \infty$,*

(I) for the first empirical likelihood function, under $H_1 : \alpha = \alpha_0 - d a_T$, we have

$$-2\log\ell(\alpha_0) \xrightarrow{p} \begin{cases} \chi_1^2(\delta_{11}^2), & \textit{for } |\phi| < 1 \\ \frac{\left(\xi_1 - d\int_0^1 \frac{J_c(r)}{|J_c(r)|^{2\gamma}} dr\right)^2}{\bar{\sigma}^2}, & \textit{for } \phi = 1 + \frac{c}{T} \textit{ for some } c \in R \textit{ with } \mu = 0, \\ \chi_1^2(\delta_{12}^2), & \textit{for } \phi = 1 + \frac{c}{T} \textit{ for some } c \in R \textit{ with } \mu \neq 0, \end{cases}$$

where (i) $a_T = T^{-1/2}$, and δ_{11} denotes the second component of $\bar{\Sigma}_1^{-1/2}\boldsymbol{\theta}_1$ when $|\phi| < 1$ with

$$\boldsymbol{\theta}_1 = -d \begin{pmatrix} 1 \\ \lim\limits_{t\to\infty} E\left(\frac{X_{t-1}}{(1+X_{t-1}^2)^{\gamma}}\right) \end{pmatrix},$$

$$\bar{\Sigma}_1 = \begin{pmatrix} \sigma_u^2 & \sigma_u^2 \lim\limits_{t\to\infty} E\left(\frac{X_{t-1}}{(1+X_{t-1}^2)^{\gamma}}\right) \\ \sigma_u^2 \lim\limits_{t\to\infty} E\left(\frac{X_{t-1}}{(1+X_{t-1}^2)^{\gamma}}\right) & \sigma_u^2 \lim\limits_{t\to\infty} E\left(\frac{X_{t-1}}{(1+X_{t-1}^2)^{\gamma}}\right)^2 + \bar{\sigma}^2 \end{pmatrix}, \text{ where } \sigma_u^2 = E(U_1^2);$$

and (ii) $a_T = \frac{1}{T^{1-\gamma}}$, and $\xi_1 \sim N(0, \bar{\sigma}^2)$, and $J_c(s) = \int_0^s e^{c(s-r)} dW_\varepsilon(r)$ with $W_\varepsilon(r)$ being the Wiener process related to $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t$ for some $r \in (0,1]$, when $\phi = 1 + \frac{c}{T}$ for some $c \in R$ and $\mu = 0$; and (iii) $a_T = \frac{1}{T^{\frac{3}{2}-2\gamma}}$, and $\delta_{12} = \frac{-d}{\bar{\sigma}} \int_0^1 \frac{\mu \int_0^r e^{c(r-s)} ds}{|\mu \int_0^r e^{c(r-s)} ds|^{2\gamma}} dr$ when $\phi = 1 + \frac{c}{T}$ for some $c \in R$ and $\mu \neq 0$.

(II) for the second empirical likelihood function, under $H_1 : \alpha = \alpha_0 - db_T$, we have

$$-2\log \widetilde{\ell}(\alpha_0) \xrightarrow{p} \begin{cases} \chi_1^2(\delta_{21}^2), & \text{for } |\phi| < 1 \\ \frac{\left(\xi_2 - d\int_0^1 sgn(J_c(r))dr\right)^2}{(E(U_1^2))^2}, & \text{for } \phi = 1 + \frac{c}{T} \text{ for some } c \in R \text{ with } \mu = 0, \\ \chi_1^2(\delta_{22}^2), & \text{for } \phi = 1 + \frac{c}{T} \text{ for some } c \in R \text{ with } \mu \neq 0, \end{cases}$$

where (i) $b_T = T^{-1/2}$, and δ_{21} denotes the second component of $\bar{\Sigma}_2^{-1/2}\boldsymbol{\theta}_2$ when $|\phi| < 1$ with

$$\boldsymbol{\theta}_2 = -d \begin{pmatrix} 1 \\ \lim\limits_{t\to\infty} E\left(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)}\right) \end{pmatrix},$$

$$\bar{\Sigma}_2 = \sigma_u^2 \begin{pmatrix} 1 & \lim\limits_{t\to\infty} E\left(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)}\right) \\ \lim\limits_{t\to\infty} E\left(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)}\right) & \lim\limits_{t\to\infty} E\left(\left(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)} + U_{t-1}\right)^2\right) \end{pmatrix};$$

and (ii) $b_T = T^{-1/2}\log(T)$, and $\xi_2 \sim N(0, (E(U_1^2))^2)$ when $\phi = 1 + \frac{c}{T}$ for some $c \in R$ and $\mu = 0$; and (iii) $b_T = T^{-1}\log(T)$, and $\delta_{22} = \frac{-2d\int_0^1 sgn(\mu\int_0^r e^{c(r-s)}ds)dr}{E(U_1^2)}$ when $\phi = 1 + \frac{c}{T}$ for some $c \in R$ and $\mu \neq 0$, where $sgn(\cdot)$ denotes the sign function.

Remark 3. *Theorem 3 shows that: when $|\phi| < 1$, the powers of $-2\log \ell(\alpha)$ and $-2\log \widetilde{\ell}(\alpha)$ are of the same order; when $\phi = 1 + \frac{c}{T}$ for some $c \in R$, the power of $-2\log \ell(\alpha)$ to detect* *the local alterative hypothesis is of order* $\geq T^{-\max(1-\gamma,\frac{3}{2}-2\gamma)}$*, while that of* $-2\log\widetilde{\ell}(\alpha)$ *is of order* $T^{-1/2}\log(T)$*. That is, it is impossible for* $-2\log\ell(\alpha)$ *to distinct* $\alpha_0+\frac{d}{T^\tau}$ *from* α_0 *when* $\tau\in(\max(1-\gamma,\frac{3}{2}-2\gamma),\frac{1}{2})=(1-\gamma,\frac{1}{2})$ *for given* $\gamma\in(\frac{1}{2},\frac{3}{4})$*. On the other hand,* $-2\log\widetilde{\ell}(\alpha)$ *is able to distinct* $\alpha_0+\frac{d}{T^\tau}$ *from* α_0 *for any* $\tau\in(0,\frac{1}{2})$*. Therefore,* $-2\log\widetilde{\ell}(\alpha)$ *is more powerful than* $-2\log\ell(\alpha)$ *in the cases of unit root and near unit root.*

Remark 4. *It is interesting to find that, when taking* $\gamma=0.75$ *as usually did in Li et al. (2014), the first test* $-2\log\ell(\alpha)$ *has no local power in the case of* $\phi=1+\frac{c}{T}$ *for some* $c\in R$ *when* $\mu\neq 0$*.*

Remark 5. *Observe that both* $\xi_1^2/\bar{\sigma}^2\sim\chi_1^2$ *and* $\xi_2^2/E(e_{t-1}^2\sigma_{t-1}^2\sigma_t^2)\sim\chi_1^2$ *in the case of* $\phi=1+\frac{c}{T}$ *for some* $c\in R$ *when* $\mu=0$*. Hence, the power of both tests above tend to 1 as* $|d|\to\infty$*.*

III Simulation Study

In this section we conduct Monte Carlo studies to examine the finite sample performance of our proposed log-empirical likelihood $-2\log\widetilde{\ell}(\alpha)$ in terms of coverage accuracy for testing $H_0:\alpha=0$ against $H_1:\alpha\neq 0$. In the first simulation study, we consider $\mu=\{0,\ 0.5\}$, $\beta_0=\{0,\ 1\}$, $\alpha_0=\{0,\ 1\}$, and $\phi=1+c/T^\delta$ with $(c,\delta)=\{(-0.1,0),(1,1),(-1,1),(-1,0.5),(0,0)\}$, which stand for five different types of stationary or nonstationary predictors. All computations are repeated 10000 times, and the sample size is taken to be 400 and 1000. The results are computed by the '*emplik*' and '*nlm*' packages, which are the most popular **R** packages for computing the profile empirical likelihood. The empirical coverage probabilities are reported in Table 1 at different significance levels, i.e., $a=0.25,\ 0.1,\ 0.05$. It turns out that the empirical coverage probabilities are close to their theoretical counterparts $1-a$, and tend to be closer as the sample size increases from 400 to 1000 in most cases. This indicates that the proposed test statistic $-2\log\widetilde{\ell}(\alpha)$ performs well under different parameter settings as reported in Table 1.

We follow Zhu et al. (2014) to denote EL1 as the empirical likelihood function for testing the predictability of model (1) when $\alpha=0$ and EL2 as the one involving data splitting when α is unknown. As mentioned in Section 1, if the test above fails to reject $H_0:\alpha=0$, we will be able to use EL1 to test the predictability of model (1). Note that EL1 is more powerful than EL2 as EL1 involves no data splitting technique as shown in Zhu et al. (2014). In fact, when the true intercept $\alpha_0=0$, it is easy to check that under the hypothesis $\tilde{H}_1:\beta=\beta_0+d/\sqrt{T}$, we have under condition C1 that

$$\frac{\frac{1}{\sqrt{m}}\sum_{t=1}^{m}(\tilde{Y}_t-\beta_0\tilde{X}_{t-1})\frac{\tilde{X}_{t-1}}{\sqrt{1+\tilde{X}_{t-1}^2}}}{\sqrt{\frac{1}{m}\sum_{t=1}^{m}\left((\tilde{Y}_t-\beta_0\tilde{X}_{t-1})\frac{\tilde{X}_{t-1}}{\sqrt{1+\tilde{X}_{t-1}^2}}\right)^2}} \xrightarrow{d} N(\Delta_1,1), \text{ and}$$

$$\frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(Y_t-\beta_0X_{t-1})\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left((Y_t-\beta_0X_{t-1})\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}\right)^2}} \xrightarrow{d} N(\Delta_2,1),$$

as $T\to\infty$ by using similar techniques as the proof of Theorem 3. Here m denotes the largest integer less than $T/2$, $\tilde{X}_t = X_{t+m}-X_t$ and $\tilde{Y}_t = Y_{t+m}-Y_t$, for $t=1,2,\cdots,m$. Note that $\Delta_1=\frac{d}{2\sigma_u}<\Delta_2=\frac{d}{\sigma_u}$ for $\phi=1+\frac{c}{T}$ for some $c\in R$ with $\mu=0$. ($\Delta_1<\Delta_2$ holds similarly for the other combinations of (μ,ϕ).) That is, it is more difficult to distinguish $N(\Delta_1,1)$ than $N(\Delta_2,1)$ from $N(0,1)$. Hence, the predictability testing method EL2 is less powerful than EL1. Therefore, if it is plausible to know a prior $\alpha=0$ explicitly, EL1 would be used in testing the predictability to improve the power performance.

To illustrate the benefit of adding a pretesting for $H_0:\alpha=0$ via using our proposed empirical likelihood method for the predictability test in model (1), we conduct another simulation study as follows. We investigate four scenarios based on different combinations with α_0 and β_0 being zero or nonzero. In each scenario, we first choose randomly one model from four candidate models with $(\alpha_0,\beta_0)\in\mathcal{A}\otimes\mathcal{B}$, where $\mathcal{A}\in\{\{0,1\},\{0,0.03\}\}$, $\mathcal{B}\in\{\{0,0.015\},\{0,0.03\}\}$ and $\otimes$ denotes the Cartesian product. For example, in the first scenario, we have (α_0,β_0) randomly generated from $\{0,1\}\otimes\{0,0.015\}=\{(0,0),(0,0.015),(1,0),(1,0.015)\}$ with equal probability. After that, we carry out the predictability test, i.e., hypothesis $\tilde{H}_0:\beta=0$, through the following procedures: (W1) a one-step procedure directly using EL2 of Zhu et al. (2014); (W2) a two-step procedure with a pretesting step for checking whether or not $\alpha=0$ as its first step, and then using EL2 if rejecting $\alpha=0$, or EL1 otherwise, as its second step. The other simulation settings are similar to the first simulation study above.

TABLE 1

Empirical coverage probabilities of $-2\log\widetilde{\ell}(\alpha)$*, where* T *denotes the sample size, and* a *the significance level*

(α_0, β_0)	(μ, c, δ)		$1-a$				$1-a$		
		T	0.75	0.90	0.95	T	0.75	0.90	0.95
(0, 0)	$(0.5, -0.1, 0)$	400	0.7496	0.8968	0.9464	1000	0.7535	0.8988	0.9497
	$(0.5, -1, 1)$	400	0.7457	0.8958	0.9459	1000	0.7429	0.8935	0.9438
	$(0.5, 1, 1)$	400	0.7485	0.8936	0.9449	1000	0.7448	0.8976	0.9480
	$(0.5, -1, 0.5)$	400	0.7474	0.8949	0.9474	1000	0.7404	0.8945	0.9489
	$(0.5, 0, 0)$	400	0.7338	0.8901	0.9425	1000	0.7480	0.8997	0.9471
(0, 1)	$(0, -0.1, 0)$	400	0.7319	0.8891	0.9433	1000	0.7377	0.8935	0.9495
	$(0, -1, 1)$	400	0.7496	0.8996	0.9473	1000	0.7514	0.9001	0.9498
	$(0, 1, 1)$	400	0.7461	0.9005	0.9500	1000	0.7437	0.9005	0.9479
	$(0, -1, 0.5)$	400	0.7488	0.8974	0.9480	1000	0.7524	0.9023	0.9534
	$(0, 0, 0)$	400	0.7457	0.8997	0.9481	1000	0.7476	0.8975	0.9503
	$(0.5, -0.1, 0)$	400	0.7445	0.8925	0.9470	1000	0.7481	0.8966	0.9498
	$(0.5, -1, 1)$	400	0.7529	0.9004	0.9475	1000	0.7477	0.9002	0.9516
	$(0.5, 1, 1)$	400	0.7500	0.8966	0.9453	1000	0.7502	0.8986	0.9486
	$(0.5, -1, 0.5)$	400	0.7405	0.8968	0.9477	1000	0.7536	0.8997	0.9510
	$(0.5, 0, 0)$	400	0.7397	0.8970	0.9465	1000	0.7496	0.9007	0.9482
(1, 1)	$(0, -0.1, 0)$	400	0.7324	0.8919	0.9458	1000	0.7405	0.8905	0.9460
	$(0, -1, 1)$	400	0.7586	0.9000	0.9516	1000	0.7557	0.9063	0.9510
	$(0, 1, 1)$	400	0.7493	0.8983	0.9509	1000	0.7538	0.9041	0.9526
	$(0, -1, 0.5)$	400	0.7389	0.8968	0.9495	1000	0.7458	0.8988	0.9486
	$(0, 0, 0)$	400	0.7597	0.9045	0.9487	1000	0.7558	0.9026	0.9516
	$(0.5, -0.1, 0)$	400	0.7405	0.8955	0.9468	1000	0.7448	0.8976	0.9503
	$(0.5, -1, 1)$	400	0.7466	0.8991	0.9497	1000	0.7448	0.8979	0.9474
	$(0.5, 1, 1)$	400	0.7476	0.8965	0.9477	1000	0.7406	0.8940	0.9464
	$(0.5, -1, 0.5)$	400	0.7470	0.8966	0.9483	1000	0.7502	0.9015	0.9486
	$(0.5, 0, 0)$	400	0.7467	0.8971	0.9485	1000	0.7466	0.8941	0.9455

TABLE 2

Empirical ratios of correctly judging whether or the true parameter β_0 is equal to 0 when testing the predictability of model (1) at the significance level $a = 0.05$. Here W1 and W2 denotes the one-step and two-step procedure, respectively

(α_0, β_0)	(μ, c, δ)	$T = 400$		$T = 1000$	
		W1	*W2*	*W1*	*W2*
$\alpha_0 = \{0, 1\}, \beta_0 = \{0, 0.015\}$	$(0, -0.1, 0)$	0.513	0.513	0.524	0.544
	$(0, -1, 1)$	0.616	0.681	0.830	0.874
	$(0, 1, 1)$	0.733	0.787	0.942	0.959
	$(0, -1, 0.5)$	0.507	0.522	0.555	0.582
	$(0, 0, 0)$	0.662	0.727	0.906	0.920
$\alpha_0 = \{0, 1\}, \beta_0 = \{0, 0.03\}$	$(0, -0.1, 0)$	0.565	0.580	0.625	0.680
	$(0, -1, 1)$	0.797	0.854	0.954	0.968
	$(0, 1, 1)$	0.891	0.919	0.966	0.979
	$(0, -1, 0.5)$	0.578	0.610	0.798	0.859
	$(0, 0, 0)$	0.833	0.867	0.960	0.961
$\alpha_0 = \{0, 0.03\}, \beta_0 = \{0, 0.015\}$	$(0, -0.1, 0)$	0.518	0.535	0.526	0.560
	$(0, -1, 1)$	0.633	0.766	0.870	0.948
	$(0, 1, 1)$	0.738	0.852	0.914	0.945
	$(0, -1, 0.5)$	0.515	0.532	0.554	0.645
	$(0, 0, 0)$	0.659	0.790	0.890	0.946
$\alpha_0 = \{0, 0.03\}, \beta_0 = \{0, 0.03\}$	$(0, -0.1, 0)$	0.533	0.576	0.642	0.718
	$(0, -1, 1)$	0.791	0.907	0.969	0.961
	$(0, 1, 1)$	0.876	0.936	0.977	0.960
	$(0, -1, 0.5)$	0.572	0.661	0.773	0.913
	$(0, 0, 0)$	0.832	0.916	0.966	0.956

We compute the empirical ratios of correctly judging whether or not the true parameter β_0 is equal to 0, i.e., the predictor has predictability. Table 2 reports the corresponding results of W1 and W2 based on 1000 repeated computations. It turns out that the empirical ratio of correct judgement can be improved in most cases through adding a pretesting step by using our test of $-2\log\widetilde{\ell}(\alpha)$, especially when the sample size is relative small. That is, a pretesting step of using $-2\log\widetilde{\ell}(\alpha)$ can help to improve the finite performance of the empirical likelihood based predictability test suggested by Zhu et al. (2014).

IV Applications

In this Section, we examine two real data applications to illustrate how to apply our proposed log-empirical likelihood function $-2\log\widetilde{\ell}(\alpha)$ as a pretest for the intercept and the importance of adding such a test before considering the predictability in the regressions.

S&P 500 index data revisited

We first revisit the S&P 500 index data (S&P 500) given in Campbell and Yogo (2006). We use the monthly excess returns of S&P 500 during the period from January 1952 to December 2015 for the predicted variable, and the Long-term Yield during the same period as the predictor.

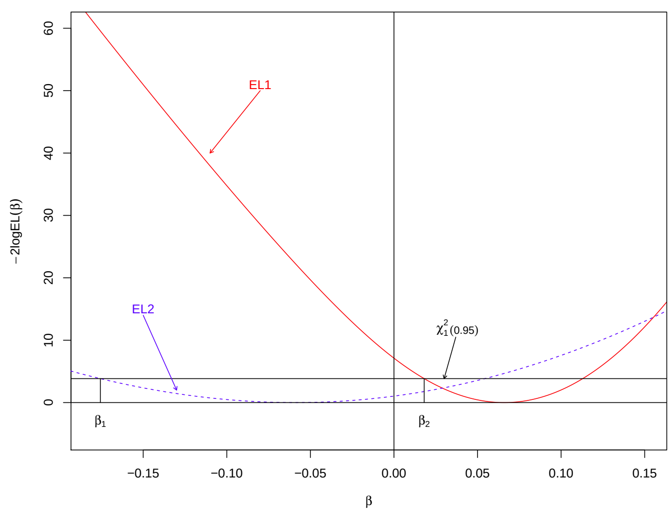

Figure 1 The curves of EL1 and EL2 with the predictor being Long-term Yield during 1952 to 2015. Here $\beta_1 = -0.1756$, $\beta_2 = 0.01809$

For this case, under the null of $\alpha = 0$, we compute that $-2\log\widetilde{\ell}(\alpha) = 12.0949 > \chi_1^2(0.95) = 3.8415$. This indicates that we can reject $H_0 : \alpha = 0$ at the significance level $a = 0.05$. The conclusion of rejecting H_0 can also be supplemented by the following fact : Note that when the true intercept is 0, both EL1 and EL2 of Zhu et al. (2014) can be used to test whether or not $\beta = 0$, and the minimizers of -2× *log-empirical likelihood* functions of β corresponding to EL1 and EL2, say $-2\,\text{logEL1}(\beta)$ and $-2\,\text{logEL2}(\beta)$, respectively, should tend to be close to each other, as they converge to the same value β_0 when the sample size goes to infinity. Otherwise, when $\alpha_0 \neq 0$, the minimizer of $-2\,\text{logEL1}(\beta)$ may differ from that of $-2\,\text{logEL2}(\beta)$ by observing that the existing of the intercept term have potential impacts on the behavior of $-2\,\text{logEL1}(\beta)$.

We plot both curves of $-2\log \mathrm{EL1}(\beta)$ and $-2\log \mathrm{EL2}(\beta)$ in Figure 1. It is shown that their minimizers differ from each other obviously. This supports us to conclude that $-2\log \widetilde{\ell}(\alpha)$ performs well in this case.

It is worth mentioning that in the range of (β_1, β_2), EL1 and EL2 of Zhu et al. (2014) come out different conclusions, and $\beta = 0$ lies in such interval means if we use EL1 and EL2 to test the predictability, we may obtain contradictive conclusions , because $-2\log \mathrm{EL1}(0) < \chi_1^2(0.95)$ while $-2\log \mathrm{EL2}\ (0) > \chi_1^2(0.95)$. In fact, this is not rare in practice. To this end, we further investigate several other commonly used financial variables, i.e. Term Spread, Long-term Yield and Earnings-price Ratio as the predictors, and test their predictability by applying EL1 and EL2. Considering the statement in Campbell and Yogo (2006) that (i) the valuation ratios are sensitive to whether the sample period includes data after 1994; (ii) the nature of interest rates in the United States changed after 1952 due to the Fed's policy of pegging the interest rate, we divide the collected data into four subsamples: 1926 - 2015, 1926 - 1994, 1994 - 2015 and 1952 - 2015.

Table 3 reports the statistic values of $-2\log \widetilde{\ell}(\alpha)$ under the null of $\alpha = 0$ as well as values of predictability test EL1 and EL2. Note that there are several cases that EL1 and EL2 imply different conclusions. For example, the null hypothesis of no predictability is rejected by EL2 for all periods of the Term Spread except for subsample 1994 - 2015, while EL1 fails to reject any of them. As for the variable Long-term Yield, for subsample 1926 - 2015, EL1 reject the null at 10% significance level while EL2 indicates non-rejection; for subsample 1952 - 2015, the consistency of EL1 and EL2 depends on the significance level. Finally, for Earnings-price Ratio, the null hypothesis of no predictability is not rejected by EL1 and EL2 for the subsample 1926-1994 and 1926-2015, which is in line with the conclusion given in Zhu et al. (2014). According to the values of $-2\log \widetilde{\ell}(\alpha)$ in Table 3, the null of $\alpha = 0$ is strongly rejected for all cases which suggest that EL2 is the appropriate test to use. Hence, if we rely on EL1 here, the conclusion would probably be incorrect.

TABLE 3

Test results for the monthly S&P 500 value-weighted index

Predictor	*Time period*	$-2\log\widetilde{\ell}(0)$	*Values of the predictability test statistics*	
			EL1	EL2
Term Spread	1926 - 2015	26.3214***	0.8273	5.7820**
	1926 - 1994	40.4477***	0.4221	3.5707*
	1994 - 2015	8.3356***	0.2249	0.0366
	1952 - 2015	43.8885***	2.6218	9.1772***
Long-term Yield	1926 - 2015	31.5969***	2.9637*	1.4705
	1926 - 1994	26.4610***	4.6973**	6.106254**
	1994 - 2015	6.3603**	0.2012	0.3590
	1952 - 2015	12.0950***	3.4470*	7.8816***
Earnings-price Ratio	1926 - 2015	43.6337***	0.1776	0.0507
	1926 - 1994	35.5870***	0.7258	1.2213
	1994 - 2015	7.1544***	0.4559	3.0020*
	1952 - 2015	33.1421***	0.3180	3.3220*

[1] Significance levels: *p <0.10, **p <0.05, ***p <0.01.

Monthly weighted dollars stock return data

In this subsection, we explore the predictive relation between investor sentiment and stock market returns. It has been well documented that investor sentiment plays an important role in international financial markets (see, e.g., Baker and Wurgler (2006), Baker et al. (2007), Lemmon and Portniaguina (2006), Schmeling (2009), Baker et al. (2012) and others).

Given that many studies in the literature mainly focus on U.S stock market, our sentiment indices look at six stock markets: the United States, Canada, the United Kingdom, Germany, France and Italy, with the objective to provide a wider international prospective of this issue. Among different sentiment measures, since Consumer Confidence Indices (CCI) are found to be highly adequate measures of investor sentiment (e.g. Fisher and Statman (2003); Qiu and Welch (2006)), we follow the literature to use this as a proxy for individual investor sentiment in our application. It is pointed out that an advantage of this proxy is that it covers the sentiment of investors at the aggregate level and not exclusively of financial markets. Similarly, Qiu and Welch (2006) finds CCI to be useful predictor of excess returns on small deciles stocks.

We use monthly weighted stock return as the predicted variable and their CCI data as the predictor. The data description of each country has been summarized in Table 5. The monthly

CCI data for the U.S. market is taken from the University of Michigan Surveys of Consumers. The CCI data for Canada is obtained from the Conference Board of Canada, while data for the U.K., France, Germany and Italy are obtained from "Directorate Generale for Economic and Financial Affairs" (DG ECFIN). The stock return data for all the countries are extracted from Professor Kenneth French's website as they are collected across countries in a consistent manner (See Schmeling (2009)).

Figure 2 - 3 display the ACF and PACF plots of CCI for these six countries respectively. These plots indicate that all of these six CCI series have high degree of serial correlation and in some cases it is very close to 1. To check the persistence of the predictor CCI, in Table 4, we summarize the results of Augmented Dickey-Fuller (ADF) test for all the countries. The results of Ljung-Box test also confirm the number of lags included, as shown in the second column, are sufficient in all the series. The P-values of ADF test indicate that CCI in U.S. and Italy are clearly unit root process, while for France, CCI is stationary. The conclusions for U.K and Germany are not quite clear-cut, as the null of unit root is rejected at 10% significance level while we fail to reject the null when the significance level is 5%. Similarly, for the case in Canada, the null of unit root is only marginally failed to be rejected at 5% level. This again shows the benefit of having a unified approach that can accommodate regressors of different persistence without conducting any additional test.

TABLE 4

P-values of tests for the predictor and residual series

Country	*Number of lags**	*Ljung - Box test*		*ADF Test*
	$\{X_t\}$	$\{U_t\}$	$\{\epsilon_t\}$	$\{X_t\}$
United States	1	0.5852	0.1776	0.4700
Canada	4	0.8147	0.1066	0.0530
United Kingdom	1	0.8072	0.0519	0.0876
Germany	2	0.6361	0.1163	0.0841
France	1	0.4732	0.8851	0.0100
Italy	2	0.0936	0.4214	0.3537

*When the number of lag is greater than 1, the data is fitted via the augmented Dickey-Fuller representation mentioned in Remark 2 with $p =$ Number of lags - 1.

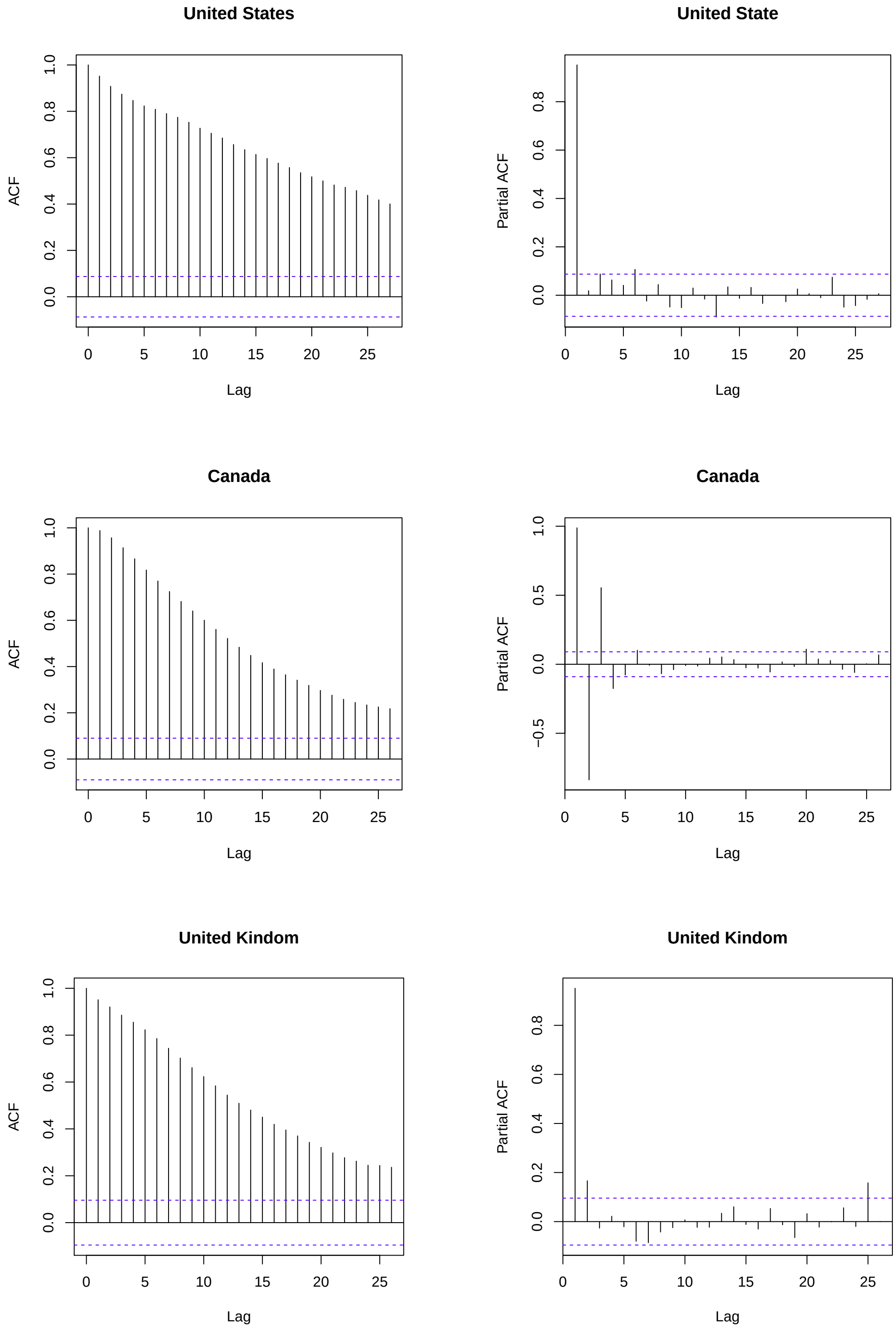

Figure 2 The autocorrelogram and partial autocorrelogram with variable being CCI of United State, Canada and United Kingdom

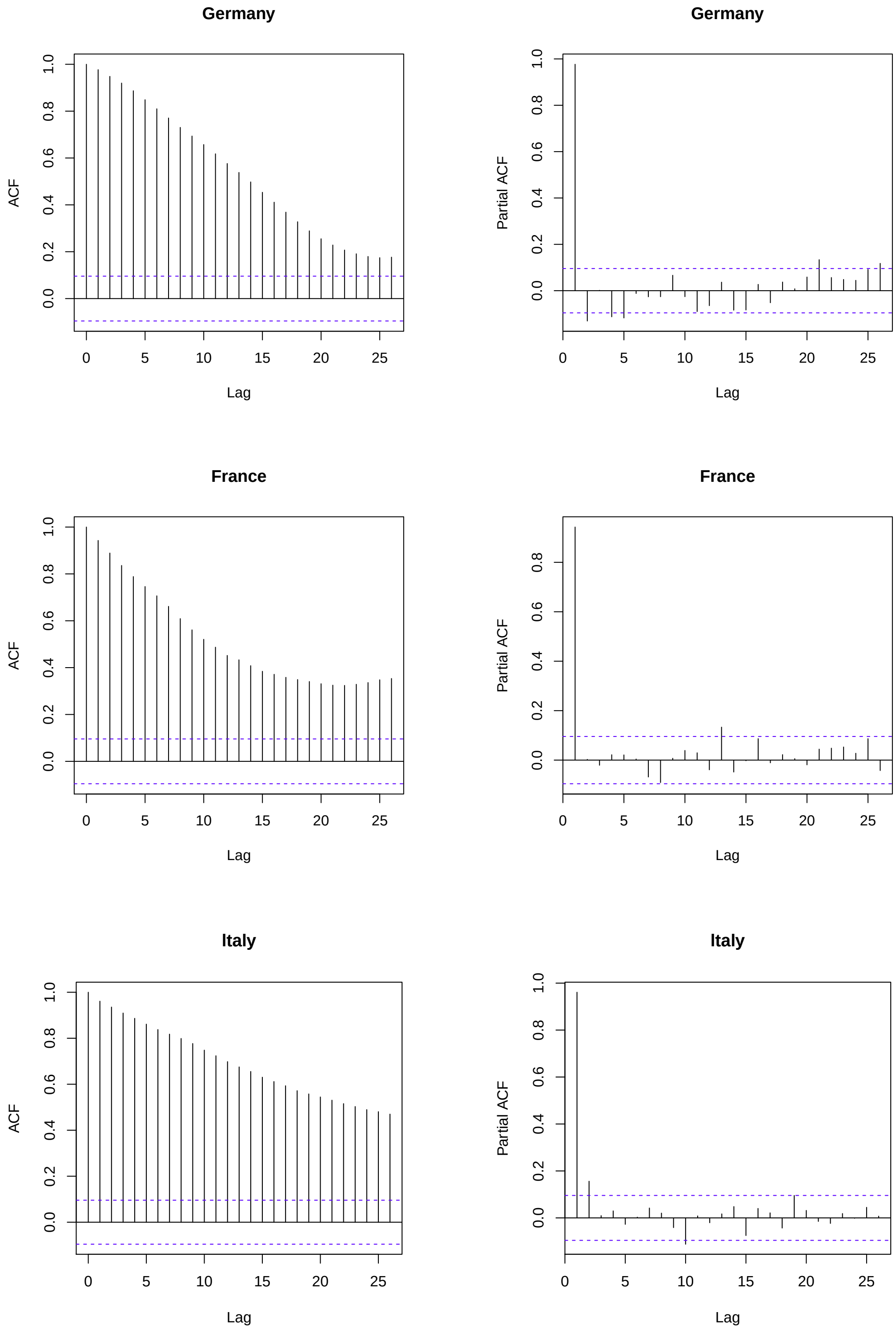

Figure 3 The autocorrelogram and partial autocorrelogram with variable being CCI of Germany, France and Italy

TABLE 5

Tests results for stock return predictability of investor sentiment

Country	*Time period*	*T*	$-2\log\widetilde{\ell}(0)$	*Values of the predictability test statistics*	
				EL1	*EL2*
United States	1978.01 - 2020.01	505	1.0364	23.4907***	0.0334
Canada	1980.07 - 2019.12	474	1.5663	9.7009***	0.5208
United Kingdom	1985.01 - 2019.12	420	0.9942	15.0027***	0.1009
Germany	1985.01 - 2019.12	420	0.2072	12.5109***	0.1774
France	1985.01 - 2019.12	420	0.9921	15.1127***	0.2594
Italy	1985.01 - 2019.12	420	0.7230	7.5849***	0.1369

[1] Significance levels: *p <0.10, **p <0.05, ***p <0.01.

Since our proposed test is robust to whether or not the predictor is stationary, we may directly apply it to the current application. Table 5 summarizes the test statistic values for six countries. The results clearly show that using EL1 or EL2 leads to conflicting conclusion regarding the predictability of CCI to the stock returns. That is, all EL2 fail to reject the null $\beta = 0$ which suggests no predictability evidence is found, while all EL1 reject the null for all the countries at different level of significance and this provide evidence that such predictive relation is present. Hence, a decision is required to make a choice between these two tests which suggest that a pretest would be of great importance. Reported in the fourth column, the results of our proposed pretest indicate that zero intercept in model (1) cannot be rejected for all countries. Therefore, EL1 will be preferred here and the predictability relation is confirmed for all six countries. In fact, when we plot the corresponding performance of EL1 and EL2 for the case of the United States as an example in Figure 4, $-2\times$ log-empirical likelihood functions are much closer to each other, which means both EL1 and EL2 work under this scenario and implies that we cannot reject the null of $\alpha = 0$ in this case. This confirms with our U.S. result in Table 5 as the value of $-2\log\widetilde{\ell}(\alpha)$ is lower than the corresponding critical value. In summary, based on the analysis above, our conclusion shows that β is statistically significant in the predictive regression and the investor sentiment does have a predictive power to stock returns for all the countries we discussed in the given time periods.

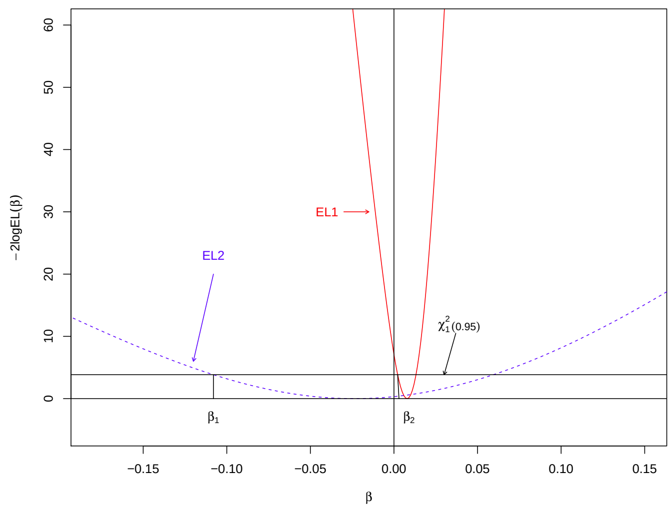

Figure 4 The curves of the EL1 and EL2 with predictor being US investor sentiments during 1982 to 2007. Here $\beta_1 = -0.1756$, $\beta_2 = 0.01809$

V Conclusions

A series of researches have been devoted to investigate whether asset returns can be predicted by some relevant financial or economic variables such as Dividend - price Ratio, Earnings - price Ratio, Interest Rate and so on. Zhu et al. (2014) proposed two unified methods for the predictive regression model which have an attractive advantage that they can allow for regressors with different degree of persistence, *i.e.*, whether it is stationary, nearly nonstationary or unit root. However, the testing statistics included in these procedures depend on a prior information of whether the intercept of the regression is zero or not.

In this paper, we develop a unified pretest for the intercept in the predictive regression to determine which predictability tests in Zhu et al. (2014) to be used to avoid misleading inferences. By revising the score functions in Hill et al. (2016), we prove that Wilks' theorem holds for the profile empirical likelihood function $-2\log\widetilde{\ell}(\alpha_0)$. In addition to its size property, the theoretical power property of our test is also discussed. The simulation studies show that our test of intercept performs very well in finite samples. It is also confirmed that it is beneficial to use a two-step approach with our pretest as the first step when testing the predictability. Furthermore, the importance of such pretest has been supported by our two real data applications. The first application, by re-visiting S&P 500 index data, illustrate when using our pretest, the evidence of predictability is found in most subsamples for the predictors Term Spread, Long-term Yield and Earnings-price Ratio. In the second application, we examined the predictive relation between investor sentiment and stock markets returns for six countries. Using Consumer Confidence Index as the proxy and applying our pretest as a first step, we find strong evidence of predictability in all countries which are consistent with findings in the literature. Note that, since our proposed pretest has no restrictions on the nature of predicting variables, this may imply wide applications in economics and finance.

Appendix: Proofs of the main results

Before proving the main results, we provide some preliminary lemmas. Since the proof of Theorem 1 is similar to that of Theorem 2, we only prove Theorem 2 in the sequel.

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

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{\boldsymbol{\zeta}}_t(\alpha_0,\beta_0)\xrightarrow{d}N(0,\Sigma)\quad \textit{as } T\to\infty,$$

where $\Sigma := (\sigma_{ij}^2)_{1\le i,j\le 2}$ *satisfying* $\Sigma=\bar{\Sigma}_2$*, which is specified in Theorem 3, when* $|\phi|<1$*, and*

$$\Sigma=\begin{pmatrix}\sigma_u^2 & 0\\ 0 & \sigma_u^4\end{pmatrix},$$

when $\phi = 1+c/T$ *for some* $c\in R$*. Here* $\sigma_u^2=E(U_1^2)$*.*

Proof. Note that, for $t=1,2,\cdots,T$,

$$\begin{aligned}\widetilde{Z}_{t1}(\alpha_0,\beta_0) &= U_t,\\ \widetilde{Z}_{t2}(\alpha_0,\beta_0) &= U_t\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)}+U_tU_{t-1}.\end{aligned}$$

Trivially,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{Z}_{t1}(\alpha_0,\beta_0)\xrightarrow{d}N(0,\sigma_{11}^2),\quad \text{as } T\to\infty. \tag{4}$$

When $|\phi|<1$, since $\{X_t\}$ is stationary, we have by the ergodic theorem that

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}E\left(\widetilde{Z}_{t2}^2(\alpha_0,\beta_0)|\mathcal{F}_{t-1}\right)\\ &=E(U_t^2)\cdot\frac{1}{T}\sum_{t=1}^{T}\left(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)}+U_{t-1}\right)^2\end{aligned}$$

$$\xrightarrow{p} E(U_1^2) \cdot \lim_{t\to\infty} E\left(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)} + U_{t-1}\right)^2,$$

where '$\xrightarrow{d}$' denotes the convergence in probability. When $\phi = 1+c/T$ for some $c \in R$, $|X_t| \to \infty$, and hence $\log(e + X_{t-1}^2) \to \infty$ in probability as $t \to \infty$. A simple derivation leads to

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} E\left(\widetilde{Z}_{t2}^2(\alpha_0,\beta_0)|\mathcal{F}_{t-1}\right) \\
&= E(U_t^2)\cdot\left[\frac{1}{T}\sum_{t=1}^{T}\frac{X_{t-1}^2}{(1+X_{t-1}^2)\log^2(e+X_{t-1}^2)}\right. \\
&\quad \left. + \frac{1}{T}\sum_{t=1}^{n}\frac{2X_{t-1}U_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)} + \frac{1}{T}\sum_{t=1}^{n} U_{t-1}^2\right] \\
&= E(U_t^2)\cdot\frac{1}{T}\sum_{t=1}^{T} U_{t-1}^2 + o_p(1).
\end{aligned}$$

Next, for any $0 < q < \epsilon$ and $\epsilon_1 > 0$, we have

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} E\left(\widetilde{Z}_{t2}^2(\alpha_0,\beta_0) I(\widetilde{Z}_{t2}^2(\alpha_0,\beta_0) > \epsilon_1^2 T)|\mathcal{F}_{t-1}\right) \\
&\le \frac{1}{(\epsilon_1\sqrt{T})^q}\frac{1}{T}\sum_{t=1}^{T} E\left(|\widetilde{Z}_{t2}(\alpha_0,\beta_0)|^{2+q}|\mathcal{F}_{t-1}\right) \\
&= \frac{E(|U_1|^{2+q})}{(\epsilon_1\sqrt{T})^q}\left\{\frac{1}{T}\sum_{t=1}^{T} E\left(\left|\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)} + U_{t-1}\right|^{2+q}\middle|\mathcal{F}_{t-1}\right)\right\} \\
&\le \frac{2^{1+q}E(|U_1|^{2+q})(1+E(|U_1|^{2+q}))}{(\epsilon_1\sqrt{T})^q} \to 0, \text{ as } T\to\infty,
\end{aligned}$$

by noting that

$$\left|\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)}\right| \le 1.$$

Hence, by Corollary 3.1 of Hall and Heyde (1980), we have as $n \to \infty$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{Z}_{t2}(\alpha_0,\beta_0) \xrightarrow{d} N(0,\sigma_{22}^2), \quad \text{as } T\to\infty. \tag{5}$$

On the other hand, by noting that

$$\frac{1}{T}\sum_{t=1}^{T} E(\widetilde{Z}_{t1}(\alpha_0,\beta_0)\widetilde{Z}_{t2}(\alpha_0,\beta_0)|\mathcal{F}_{t-1}) \tag{6}$$

$$= E(U_1^2) \cdot \frac{1}{T} \sum_{t=1}^{T} \left(\frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2} \log(e + X_{t-1}^2)} + U_{t-1} \right) \xrightarrow{p} \sigma_{12}^2$$

Therefore, this lemma follows directly from (4)–(6) and the Cramér-Wold device. □

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

$$\frac{1}{T} \sum_{t=1}^{T} \widetilde{\boldsymbol{\zeta}}_t(\alpha_0, \beta_0) \widetilde{\boldsymbol{\zeta}}_t(\alpha_0, \beta_0)^\top \xrightarrow{p} \Sigma,$$

$$\text{and} \qquad \max_{1 \le t \le T} \|\widetilde{\boldsymbol{\zeta}}_t(\alpha_0, \beta_0)\| = o_p(T^{1/2}), \quad \text{as } T \to \infty.$$

Proof. Like the proof of Lemma 1, the first part of this lemma can be shown by using the weak law of large numbers for martingale differences, and we skip details.

The second part of this lemma follows directly from the facts that $\max_{1 \le t \le T} |U_t| = o_p(T^{1/2})$ and

$$P\left(\max_{2 \le t \le T} |U_t U_{t-1}| > \epsilon_1 \sqrt{T} \right) \le \sum_{t=2}^{T} P\left(|U_t U_{t-1}| > \epsilon_1 \sqrt{T} \right) \le \frac{(E(|U_1|^{2+q}))^2}{\epsilon_1^{2+q} T^{1+q}} \to 0,$$

for any $\epsilon_1 > 0$. □

Before proving Theorem 1, we need some notations. Let

$$\vartheta := \vartheta(\beta) = \begin{cases} \beta & \text{if } |\phi| < 1 \\ \sqrt{T}\beta & \text{if } \phi = 1 + \frac{c}{T} \text{ or } \phi = 1 \text{ with } \mu = 0 \\ T\beta & \text{if } \phi = 1 + \frac{c}{T} \text{ or } \phi = 1 \text{ with } \mu \neq 0, \end{cases} \tag{7}$$

$\vartheta_0 := \vartheta(\beta_0)$, $\bar{\boldsymbol{\zeta}}_t(\alpha, \vartheta) = (\bar{Z}_{t1}(\alpha, \vartheta),\ \bar{Z}_{t2}(\alpha, \vartheta))^T$ for $t = 1, 2, \cdots, T$, where $\bar{Z}_{t1}(\alpha, \vartheta) = \widetilde{Z}_{t1}(\alpha, \beta)$ and $\bar{Z}_{t2}(\alpha, \vartheta) = \widetilde{Z}_{t2}(\alpha, \beta)$, and denote $\bar{L}(\alpha, \vartheta) = L(\alpha, \beta)$.

Lemma 3. *Under conditions of Theorem 1, $L(\alpha_0, \vartheta)$ attains its maximum value with probability tending to one at some point $\bar{\vartheta}$ such that $|\bar{\vartheta} - \vartheta_0| < n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2 + \delta)$ as $n \to \infty$, and $\bar{\vartheta}$ and $\bar{\boldsymbol{\lambda}}$ satisfy $Q_{1n}(\bar{\vartheta}, \bar{\boldsymbol{\lambda}}) = 0$ and $Q_{2n}(\bar{\vartheta}, \bar{\boldsymbol{\lambda}}) = 0$, where*

$$Q_{1T}(\vartheta, \boldsymbol{\lambda}) := \frac{1}{T} \sum_{t=1}^{T} \frac{\bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta)}{1 + \boldsymbol{\lambda}^\top \bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta)},$$

$$Q_{2T}(\vartheta, \boldsymbol{\lambda}) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{1 + \boldsymbol{\lambda}^\top \bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta)} \left(\frac{\partial \bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta)}{\partial \vartheta} \right)^\top \boldsymbol{\lambda}.$$

Proof. Let

$$\breve{X}_t = \begin{cases} X_t & \text{if } |\phi| < 1 \\ \frac{1}{\sqrt{T}} X_t & \text{if } \phi = 1 + \frac{c}{T} \text{ or } \phi = 1 \text{ with } \mu = 0 \\ \frac{1}{T} X_t & \text{if } \phi = 1 + \frac{c}{T} \text{ or } \phi = 1 \text{ with } \mu \neq 0, \quad t = 1, 2, \cdots, T. \end{cases}$$

In the sequel we only give a detailed proof of the case $\phi = 1 + \frac{c}{T}$ with some $c \neq 0$ and $\mu = 0$ since other cases can be proved similarly.

It is easy to check that

$$\begin{aligned} \sup_{|\vartheta - \vartheta_0| < T^{-1/\delta_0}} \left| \frac{1}{T} \sum_{t=1}^{T} \bar{Z}_{t1}(\alpha_0, \vartheta) \right| &\leq \left| \frac{1}{T} \sum_{t=1}^{T} U_t \right| + \sup_{|\vartheta - \vartheta_0| < T^{-1/\delta_0}} \left| \frac{1}{T} \sum_{t=1}^{T} (\vartheta - \vartheta_0) \breve{X}_{t-1} \right| \\ &= O_p(T^{-1/2}) + O_p(T^{-1/\delta_0}) \\ &= O_p(T^{-1/\delta_0}). \end{aligned}$$

Similarly, $\sup_{|\vartheta - \vartheta_0| < T^{-1/\delta_0}} |\frac{1}{T} \sum_{t=1}^{T} \bar{Z}_{t2}(\alpha_0, \vartheta)| = O_p(T^{-1/\delta_0})$, and hence we obtain

$$\sup_{|\vartheta - \vartheta_0| < T^{-1/\delta_0}} \left\| \frac{1}{T} \sum_{t=1}^{n} \bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta) \right\| = O_p(T^{-1/\delta_0}). \tag{8}$$

Next, observe that

$$\begin{aligned} P\left(\max_{1 \leq t \leq T} |\breve{X}_t| \geq b T^{1/2+\delta_1} \right) &\leq \sum_{t=1}^{T} P\left(|\breve{X}_t| \geq b T^{1/2+\delta_1} \right) \\ &\leq \frac{E\left(\sum_{t=1}^{T} \breve{X}_t^2 \right)}{b^2 T^{1+2\delta_1}} \\ &= \frac{E\left(\int_0^1 J_c^2(s) ds \right)}{b^2 T^{2\delta_1}} + o(1) \to 0, \end{aligned}$$

for some $1/\delta_0 > \delta_1 > 0$ as $T \to \infty$. Therefore

$$\sup_{|\vartheta - \vartheta_0| < T^{-1/\delta_0}} \max_{1 \leq t \leq T} |(\vartheta - \vartheta_0) \breve{X}_t| = o_p(T^{1/2}),$$

implies that as $T \to \infty$,

$$\sup_{|\vartheta - \vartheta_0| < T^{-1/\delta_0}} \max_{1 \leq t \leq T} |\bar{Z}_{t1}(\mu_0, \vartheta)| = o_p(T^{1/2}).$$

Similarly, $\sup_{|\vartheta - \vartheta_0| < T^{-1/\delta_0}} \max_{1 \leq t \leq T} |\bar{Z}_{t2}(\alpha_0, \vartheta)| = o_p(T^{1/2})$, and hence

$$\sup_{|\vartheta - \vartheta_0| < T^{-1/\delta_0}} \max_{1 \leq t \leq T} \left\| \bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta) \right\| = o_p(T^{1/2}), \quad \text{as } T \to \infty. \tag{9}$$

By (8) and (9), this lemma can be proved by using the same arguments as in the proof of Lemma 1 in Qin and Lawless (1994). □

Proof of Theorem 2. We only prove the case of $\mu = 0$ and $\phi = 1 + \frac{c}{T}$ for some $c \neq 0$. Using similar arguments in proving Lemma 2, we have

$$\frac{\partial Q_{1T}(\vartheta_0, 0)}{\partial \boldsymbol{\lambda}^\top} = -\frac{1}{T}\sum_{t=1}^{T} \bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta_0)\bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta_0)^\top \xrightarrow{p} -\Sigma,$$

$$\frac{1}{T}\sum_{t=1}^{T} -\breve{X}_t \frac{X_t}{\sqrt{1 + X_t^2 \log(e + X_{t-1}^2)}} \xrightarrow{p} 0.$$

Next, observe that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} \breve{X}_{t-1}U_{t-1} &= \frac{1}{T}\sum_{t=1}^{T} \phi\breve{X}_{t-2}U_{t-1} + \frac{1}{T}\sum_{t=1}^{T} V_{t-1}U_{t-1} \\
&\xrightarrow{d} \int_0^1 J_c(r)dW_u(r) + E(V_1U_1),
\end{aligned}$$

by Hansen (1992), where $W_u(r)$ denotes the Wiener process related to $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} U_t$ for any $r \in (0, 1]$. Using this, it is easy to check that

$$\begin{aligned}
\frac{\partial Q_{1T}(\vartheta_0, 0)}{\partial \vartheta} &= \left(\frac{\partial Q_{2T}(\vartheta_0, 0)}{\partial \boldsymbol{\lambda}^\top}\right)^\top \\
&= \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} -\breve{X}_t \\ -\breve{X}_t \frac{X_t}{\sqrt{1+X_t^2 \log(e+X_{t-1}^2)}} - \breve{X}_{t-1}U_{t-1} - \breve{X}_{t-2}U_t \end{pmatrix} \\
&\xrightarrow{d} \begin{pmatrix} -\int_0^1 J_c(s)ds \\ 0 \end{pmatrix},
\end{aligned}$$

where $J_c(s)$ is defined in the proof of Theorem 3. It follows from Lemmas 1-3 and the same arguments in the proof of Theorem 1 in Qin and Lawless (1994) that

$$\begin{pmatrix} \bar{\boldsymbol{\lambda}} \\ \bar{\vartheta} - \vartheta_0 \end{pmatrix} = S_T^{-1} \begin{pmatrix} -Q_{1T}(\vartheta_0, 0) + o_p(T^{-1/2}) \\ o_p(T^{-1/2}) \end{pmatrix},$$

where

$$S_T = \begin{pmatrix} \frac{\partial Q_{1T}(\vartheta_0,0)}{\partial \lambda^\top} & \frac{\partial Q_{1T}(\vartheta_0,0)}{\partial \vartheta} \\ \frac{\partial Q_{2T}(\vartheta_0,0)}{\partial \lambda^\top} & 0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^\top & 0 \end{pmatrix}$$

with $S_{11} = -\Sigma$ and $S_{12} = \begin{pmatrix} -\int_0^1 J_c(s)ds \\ 0 \end{pmatrix}$.

Next, by the standard arguments in empirical likelihood method, we can show under H_0 that, as $T \to \infty$,

$$-2\log \widetilde{\ell}(\alpha_0) = 2\sum_{t=1}^{T} \log\{1 + \bar{\boldsymbol{\lambda}}^\top \bar{\boldsymbol{\zeta}}_t(\alpha_0, \bar{\vartheta})\}$$

$$
\begin{aligned}
&= 2T(\bar{\boldsymbol{\lambda}}^\top,\ \bar{\vartheta}-\vartheta_0)\left(Q_{1T}^\top(\vartheta_0,0),\ 0\right)^\top \\
&\quad +T(\bar{\boldsymbol{\lambda}}^\top,\ \bar{\vartheta}-\vartheta_0)S_T(\bar{\boldsymbol{\lambda}},\ \bar{\vartheta}-\vartheta_0)^\top + o_p(1) \\
&= -n(Q_{1T}^\top(\vartheta_0,0),\ 0)S_T^{-1}\left(Q_{1T}^\top(\vartheta_0,0),\ 0\right)^\top + o_p(1) \\
&\xrightarrow{d} -(\boldsymbol{\eta}^\top,\ 0)\begin{pmatrix} S_{11} & S_{12} \\ S_{12}^\top & 0 \end{pmatrix}^{-1}(\boldsymbol{\eta}^\top,\ 0)^\top \\
&= -\boldsymbol{\eta}^\top\left(S_{11}^{-1}-(S_{11}^{-1}S_{12})(S_{12}^\top S_{11}^{-1}S_{12})^{-1}(S_{12}^\top S_{11}^{-1})\right)\boldsymbol{\eta} \\
&= (\Sigma^{-\frac{1}{2}}\boldsymbol{\eta})^\top\left(I_{2\times 2}-(\Sigma^{-\frac{1}{2}}S_{12})(S_{12}^\top\Sigma^{-1}S_{12})^{-1}(S_{12}^\top\Sigma^{-\frac{1}{2}})\right)(\Sigma^{-\frac{1}{2}}\boldsymbol{\eta}) \\
&= (\Sigma^{-\frac{1}{2}}\boldsymbol{\eta})^\top\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}(\Sigma^{-\frac{1}{2}}\boldsymbol{\eta}) \\
&:= \upsilon^2,
\end{aligned}
$$

where $\boldsymbol{\eta} \sim N(0,\Sigma)$, and υ denotes the second component of $\Sigma^{-\frac{1}{2}}\boldsymbol{\eta} \sim N(0, I_{2\times 2})$. Then this theorem follows immediately. □

Proof of Theorem 3. Here, we only prove Part (II) when $\phi = 1+\frac{c}{T}$ for some $c \in R$ and $\mu = 0$, since the rest proofs follow a similar fashion.

Note that, under $H_1 : \alpha = \alpha_0 - db_T$, for $t = 1, 2, \cdots, T$,

$$
\begin{aligned}
\widetilde{Z}_{t1}(\alpha_0,\beta_0) &= U_t - db_T = \widetilde{Z}_{t1}(\alpha_0 - db_T, \beta_0) - db_T, \\
\widetilde{Z}_{t2}(\alpha_0,\beta_0) &= U_t\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)} + U_tU_{t-1} \\
&\quad -db_T\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)} - db_TU_{t-1} - db_TU_t + d^2b_T^2 \\
&= \widetilde{Z}_{t2}(\alpha_0 - db_T,\beta_0) - db_T\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)} - db_T(U_{t-1}+U_t) + d^2b_T^2.
\end{aligned}
$$

Then, similar Lemmas 1-3, we have as, $T \to \infty$,

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^T \boldsymbol{\zeta}_t(\alpha_0,\beta_0) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^T \boldsymbol{\zeta}_t(\alpha_0 - db_T,\beta_0) + \begin{pmatrix} -\frac{1}{\sqrt{T}}\sum\limits_{t=1}^T db_T \\ -db_T\frac{1}{\sqrt{T}}\sum\limits_{t=1}^T \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}\log(e+X_{t-1}^2)} + O_p(b_T) \end{pmatrix}
\end{aligned}
$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{\zeta}_t(\alpha_0 - db_T, \beta_0) + \begin{pmatrix} -\frac{1}{T} \sum_{t=1}^{n} d \log(T) \\ -d\frac{1}{T} \sum_{t=1}^{T} \frac{\frac{1}{\sqrt{T}} X_{t-1} \log(T)}{\sqrt{1+\left(\frac{X_{t-1}}{\sqrt{T}}\right)^2 \left[\log\left(\frac{e}{T}+\left(\frac{X_{t-1}}{\sqrt{T}}\right)^2\right)+\log(T)\right]}} + O_p(b_T) \end{pmatrix}$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{\zeta}_t(\alpha_0 - db_T, \beta_0) + \begin{pmatrix} -\frac{1}{T} \sum_{t=1}^{T} d \log(T) \\ -d \int_0^1 \operatorname{sgn}(J_c(r)) dr + O_p(b_T) \end{pmatrix}$$

$$= \begin{pmatrix} O_p(\log(T)) \\ Y_2 - d \int_0^1 \operatorname{sgn}(J_c(r)) dr + O_p(b_T) \end{pmatrix},$$

and

$$\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\zeta}_t(\alpha_0, \beta_0) \boldsymbol{\zeta}_t(\alpha_0, \beta_0)^\top \xrightarrow{p} \Sigma,$$

$$\text{and} \qquad \max_{1 \le t \le T} \|\boldsymbol{\zeta}_t(\alpha_0, \beta_0)\| = o_p(T^{1/2}), \quad \text{as } T \to \infty,$$

and $L(\alpha_0, \vartheta)$ attains its maximum value with probability tending to one at some point $\tilde{\vartheta}$ such that $|\tilde{\vartheta} - \vartheta_0| < T^{-1/\delta_0}$ for some $\delta_0 \in (2, 2+\delta)$ as $T \to \infty$, and $\tilde{\vartheta}$ and $\tilde{\boldsymbol{\lambda}}$ satisfy $\tilde{Q}_{1T}(\tilde{\vartheta}, \tilde{\boldsymbol{\lambda}}) = 0$ and $\tilde{Q}_{2T}(\tilde{\vartheta}, \tilde{\boldsymbol{\lambda}}) = 0$, where

$$\tilde{Q}_{1T}(\vartheta, \boldsymbol{\lambda}) := \frac{1}{T} \sum_{t=1}^{T} \frac{\bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta)}{1 + \boldsymbol{\lambda}^\top \bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta)},$$

$$\tilde{Q}_{2T}(\vartheta, \boldsymbol{\lambda}) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{1 + \boldsymbol{\lambda}^\top \bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta)} \left(\frac{\partial \bar{\boldsymbol{\zeta}}_t(\alpha_0, \vartheta)}{\partial \vartheta} \right)^\top \boldsymbol{\lambda}.$$

Then under $H_1 : \alpha = \alpha_0 - db_T$, similar to the proof of Theorem 2, it is easy to check that, as $T \to \infty$,

$$\begin{aligned}
&-2 \log \widetilde{\ell}(\alpha_0) \\
&= 2 \sum_{t=1}^{n} \log\{1 + \tilde{\boldsymbol{\lambda}}^\top \bar{\boldsymbol{\zeta}}_t(\alpha_0, \bar{\vartheta})\} \\
&= -T(\tilde{Q}_{1T}^\top(\vartheta_0, 0),\ 0) \tilde{S}_T^{-1} \left(\tilde{Q}_{1T}^\top(\vartheta_0, 0),\ 0 \right)^\top + o_p(1) \\
&= - \begin{pmatrix} O_p(\log(n)) \\ \xi_2 - d \int_0^T \operatorname{sgn}(J_c(r)) dr + O_p(b_T) \\ 0 \end{pmatrix}^\top \tilde{S}_T^{-1} \begin{pmatrix} O_p(\log(T)) \\ \xi_2 - d \int_0^n \operatorname{sgn}(J_c(r)) dr + O_p(b_T) \\ 0 \end{pmatrix} + o_p(1) \\
&= \begin{pmatrix} O_p(\log(T)) \\ \xi_2 - d \int_0^T \operatorname{sgn}(J_c(r)) dr + O_p(b_T) \end{pmatrix}^\top \begin{pmatrix} O_p(b_T) & O_p(b_T) \\ O_p(b_T) & \sigma_{22}^{-2} + O_p(b_T) \end{pmatrix} \times
\end{aligned}$$

$$\begin{pmatrix} O_p(\log(T)) \\ \xi_2 - d\int_0^T \text{sgn}(J_c(r))dr + O_p(b_T) \end{pmatrix} + o_p(1)$$

$$\xrightarrow{d} \frac{\left(\xi_2 - d\int_0^1 \text{sgn}(J_c(r))dr\right)^2}{(E(U_1^2))^2},$$

by noting that $O_p(\log^2(T)b_T) = o_p(1)$ and $(\Sigma + O_p(b_T))^{-1} = \Sigma^{-1} + O_p(b_T)$, which can be obtained by applying the Taylor expansion to the matrix inverse. Here

$$\tilde{S}_T := \begin{pmatrix} \tilde{S}_{11,T} & \tilde{S}_{12,T} \\ \tilde{S}_{12,T}^\top & 0 \end{pmatrix} = \begin{pmatrix} \frac{\partial \tilde{Q}_{1T}(\vartheta_0,0)}{\partial \lambda^\top} & \frac{\partial \tilde{Q}_{1T}(\vartheta_0,0)}{\partial \vartheta} \\ \frac{\partial \tilde{Q}_{2T}(\vartheta_0,0)}{\partial \lambda^\top} & 0 \end{pmatrix}.$$

This completes the proof of this theorem. □

References

Basrak, B., Davis, R. A. and Mikosch T. (2002). 'Regular variation of GARCH process', *Stochastic Processes and Their Applications,* Vol. **99**, pp. 95-115.

Baker, M.and Wurgler, J. (2006). 'Investor sentiment and the cross - section of stock returns'. *The Journal of Finance*, Vol. **61**(4), pp. 1645-1680.

Baker, M. and Wurgler, J. (2007). 'Investor sentiment in the stock market'. *Journal of Economic Perspectives*, Vol. **21**, pp. 129-151.

Baker, M., Wurgler, J. and Yuan, Y. (2012). 'Global, local, and contagious investor sentiment', *Journal of Financial Economics*, Vol. **104**(2), pp. 272-287.

Cai, Z. and Wang, Y. (2014).'Testing predictive regression models with nonstationary regressors'. *Journal of Econometrics,* Vol. **178**, pp. 4-14.

Campbell, J.Y., and Yogo, M., (2006). 'Efficient tests of stock return predictability', *Journal of Financial Economics,* Vol. **81**(1), pp. 27-60.

Chan, N.H., Li, D. and Peng, L. (2012). 'Toward a unified interval estimation of autoregressions', *Econometric Theory,* Vol. **28**, pp. 705–717.

Dickey, D.A. and Fuller, W.A. (1981). 'Likelihood ratio statistics for autoregressive time series with a unit root', *Econometrica,* Vol. **49**, pp. 1057-1072.

Dios-Palomares, R. and Roldan, J.A. (2006). 'A strategy for testing the unit root in AR(1) model with intercept: a Monte Carlo experiment', *Journal of Statistical Planning and Inference,* Vol. **136**, pp. 2685-2705.

Fei, Y. (2018). 'Limit theory for mildly interated process with intercept', *Economics Letters*, Vol. **163**, pp. 98-101.

Fisher, K. L., and Statman, M. (2003). 'Consumer confidence and stock returns', *Journal of Portfolio Management*, Vol. **30**, pp. 115-127.

Georgiev, I., Harvey, D. I., Leybourne, S. J., and Taylor, A. R. (2018). 'Testing for parameter instability in predictive regression models'. *Journal of Econometrics*, Vol. **204**(1), pp. 101-118.

Georgiev, I., Harvey, D. I., Leybourne, S. J., and Taylor, A. R. (2019). 'A bootstrap stationarity test for predictive regression invalidity'. *Journal of Business & Economic Statistics*, **37**(3), pp. 528-541.

Fuller, W.A., Hasza, P.D. and Goebel, J.J. (1981). 'Estimation of the parameters of stochastic difference equations', *The Annals of Statististics,* Vol. **9**, pp. 531-543.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* NY: Academic Press. New York.

Hansen, B.E. (1992). 'Convergence to Stochastic Integrals for Dependent Heterogeneous Processes', *Econometric Theory,* Vol. **7**, pp. 489-500.

Hill, J.B., Li, D. and Peng, L. (2016). 'Uniform Interval Estimation for an AR(1) Process with AR Errors', *Statistica Sinica,* Vol. **26**, pp. 119-136.

Kostakis, A., Magdalinos, T., Stamatogiannis, M. P. (2014). 'Robust econometric inference for stock return predictability', *The Review of Financial Studies*, Vol. **28**(5), pp. 1506-1553.

Lemmon, M. , and Portniaguina, E. (2006). 'Consumer confidence and asset prices: some empirical evidence', *The Review of Financial Studies*, Vol. **19**(4), p.1499-1529.

Li, D., Chan, N.H. and Peng, L. (2014). 'Empirical likelihood test for causality for bivariate AR(1) processes', *Econometric Theory*, Vol. **30**, pp. 357-371.

Mikusheva, A. (2007). 'Uniform inference in autoregressive models', *Econometrica* Vol. **75**, pp. 1411-1452.

Owen, A. (2001). *Empirical Likelihood*. Chapman and Hall/CRC, Boca Raton, FL.

Phillips, P.C.B. (2015). 'Pitfalls and possibilities in predictive regression', *Journal of Financial Economics,* Vol. **13**, pp. 521-555.

Qin, J., and Lawless, J. (1994). 'Empirical likelihood and general estimating equations', *The Annals of Statististics,* Vol. **22**, pp. 300-325.

Qiu, L. and Welch, I. (2006). Investor Sentiment Measures, NBER working paper no 10794.

Schmeling, M. (2009). 'Investor sentiment and stock returns: some international evidence', *Journal of Empirical Finance*, Vol. **16**(3), pp. 0-408.

So, B.S. and Shin D.W. (1999). 'Cauchy estimators for autoregressive processes with applications to unit root tests and confidence intervals', *Econometric Theory*, Vol. **15**, pp. 165-176.

Stambaugh, R.F. (1999). 'Predictive regressions', *Journal of Financial Economics,* Vol. **54**(3), pp. 375-421.

Zhu, F., Cai, Z., and Peng, L. (2014). 'Predictive regressions for macroeconomic data', *The Annals of Applied Statistics,* Vol. **8**, pp. 577-594.