Journal of Econometrics 208 (2019) 141–159

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A unified test for predictability of asset returns regardless of properties of predicting variables

Xiaohui Liu [a], Bingduo Yang [b,c,*], Zongwu Cai [d,e], Liang Peng [f]

[a] School of Statistics, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China
[b] Lingnan (University) College, Sun Yat-sen University, Guangzhou, Guangdong 510275, China
[c] School of Finance, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China
[d] Department of Economics, University of Kansas, Lawrence, KS 66045, USA
[e] The Wang Yanan Institute for Studies in Economics and Fujian Provincial Key Lab of Statistical Sciences, Xiamen University, Xiamen, Fujian 361005, China
[f] Department of Risk Management and Insurance, Georgia State University, Atlanta, GA 30303, USA

ARTICLE INFO

Article history:
Available online 9 October 2018

JEL classification:
C12
C22

Keywords:
Empirical likelihood
Predictive regression
Weighted score

ABSTRACT

Some unified tests have been proposed recently in the literature for testing predictability of asset returns based on a simple linear predictive regression model, which has a drawback that predicted variable cannot be stationary if the predicting variable is nonstationary. To solve this issue, this paper includes the difference of the predicting variable into the simple linear predictive regression. Furthermore, a unified empirical likelihood inference is developed to test the predictability regardless of the properties of the predicting variable. A simulation study is conducted to confirm the efficiency of the proposed methods before applying to a real example.

1. Introduction

Recently, predictive regression models have gained much attention in economics and finance as well as actuarial science; see the monograph by Frees et al. (2014). Indeed, testing predictability of asset returns based on a simple predictive regression model has been studied for decades with many financial applications such as the mutual fund performance, the conditional capital asset pricing, and the optimal asset allocations. A simple linear predictive regression model assumes the following structural model:

$$Y_t = \alpha + \beta X_{t-1} + U_t, \quad X_t = \theta + \phi X_{t-1} + V_t, \quad 2 \le t \le n, \tag{1}$$

where $\{(U_t, V_t)^T\}$ is a sequence of independent and identically distributed random vectors with means zero and finite variances. Here A^T denotes the transpose of matrix or vector A. It is well documented in the literature that the least squares estimator for β based on the first equation in (1) is biased in finite sample behavior due to the correlation between U_t and V_t, and hence several bias-corrected estimators and tests for both stationary (e.g., $|\phi| < 1$) and nonstationary (e.g., $\phi = 1 - \rho/n$ for some ρ) for model (1) have been proposed in the literature; see, for example, Cavanagh et al. (1995), Stambaugh (1999), Amihud and Hurvich (2004), Chen and Deo (2009), Amihud et al. (2009), and references therein.

* Corresponding author.
E-mail address: bdyang2006@sina.com (B. Yang).

https://doi.org/10.1016/j.jeconom.2018.09.009

An issue related to the inference of β in model (1) is that the asymptotic distribution of an estimator or a test statistic depends heavily on whether X_t is stationary or nearly integrated or unit root, and θ is zero or nonzero. In practice, it is extremely challenging to distinguish between stationary and nearly integrated, and between nearly integrated and unit root since the level of persistence cannot be estimated consistently. Therefore, it is of importance to have a unified inference approach to avoid making a mistake in characterizing the predicting variable; see Campbell and Yogo (2006), Chen et al. (2013), Zhu et al. (2014), and the references therein. Recently, Choi et al. (2016) proposed a test for testing predictability (e.g., $H_0 : \beta = 0$) based on a so-called Cauchy estimation, which unifies the cases of nearly integrated and unit root, whereas Phillips and Lee (2013) and Kostakis et al. (2015) developed a new test for testing predictability by using an extended instrumental variable (dubbed as IVX) based inference. However, the construction of instrumental variables depends heavily on the persistence level of the predicting variable and the IVX approach reduces the convergence rate of the proposed estimator so that it might lose the power of the proposed test.

When $(U_t, V_t)^T$ has a bivariate normal distribution, one can write $U_t = V_t^* + \gamma V_t$ for some γ (a linear projection), where V_t^* and V_t are independent. Therefore, model (1) can be written as

$$Y_t = \alpha + \beta X_{t-1} + \gamma V_t + V_t^*, \quad X_t = \theta + \phi X_{t-1} + V_t. \tag{2}$$

Now, V_t^* and V_t become independent so that one could expect an inference based on the first equation in (2) with V_t replaced by $\hat{V}_t = X_t - \hat{\theta} - \hat{\phi} X_{t-1}$ is preferred to that based on the first equation in (1); see Amihud and Hurvich (2004) and Cai and Wang (2014) for details. Extension to modeling β and γ as a smooth function of t is available in Cai et al. (2015). Some recent developments on nonparametric inferences for predictive regressions, when X_t is nonstationary, include, but not limited to, the papers by Juhl (2014) and Kasparis et al. (2015).

Although model (1) has been studied and applied to testing predictability of asset returns, a common drawback with model (1) is that Y_t cannot be stationary if X_t is nonstationary. In such a case, it might lose the economic interpretation if Y_t is an asset return which is commonly assumed to be stationary. To overcome this issue, we propose including the difference of the predicting variable into the simple linear predictive regression. That is, we consider the following model

$$Y_t = \alpha + \beta_1 \Delta X_{t-1} + \beta_2 X_{t-2} + U_t, \quad X_t = \theta + \phi X_{t-1} + V_t, \tag{3}$$

where $\Delta X_{t-1} = X_{t-1} - X_{t-2}$. Clearly, when X_t is a unit root process, ΔX_{t-1} is stationary and Y_t in (3) can be stationary if $\beta_2 = 0$. Under model (3), Y_t is predictable by not only the difference of X_{t-1}'s but also the series X_{t-1}'s itself. Note that despite the fact that the difference $\Delta X_{t-1} = \theta + V_{t-1} + (\phi - 1)X_{t-2} = \theta + V_{t-1} - \rho X_{t-2}/n$, where $\phi = 1 - \rho/n$ for some ρ, is not an innovation unless the regressor belongs to the class of integrated processes (i.e., $\rho = 0$), it behaves asymptotically as an innovation for $\rho \neq 0$. To answer the important question on whether X_{t-1} and/or X_{t-2} can be used to directly predict Y_t, one may be interested in testing the null hypothesis $H_0 : \beta_1 = 0$ or $H_0 : \beta_2 = 0$ or $H_0 : \beta_1 = \beta_2 = 0$ under model (3).

Model (3) is different from the existing literature in the following ways. First, although model (3) can be viewed as a special case of multiple-predictor regressions in the literature such as Amihud et al. (2009) for the stationary case, there exists no unified test for testing predictability; that is, a test does not require knowing whether X_t is stationary or nearly integrated or unit root. Also, it does not need to estimate V_t as in model (2). Secondly, we remark that model (3) is different from that in Amihud and Hurvich (2004) and Cai and Wang (2014) since the predicting variables in (3) are X_{t-1} and X_{t-2} and we allow errors U_t and V_t to be correlated, while the predicting variables in (2) are X_t and X_{t-1} essentially and errors V_t^* and V_t are uncorrelated. Moreover, model (3) is different from that in Zhu et al. (2014) which is model (1) without the difference term. Finally, it is worth pointing out that the idea of adding the difference ΔX_{t-1} directly into (3) as a regressor is different from that in Kostakis et al. (2015) for using ΔX_t to construct the IVX instrument to obtain the two-stage least squares estimator of β_2 although both use the differencing idea.

To obtain a unified test for testing predictability in model (3) without knowing the properties of $\{X_t\}$, this paper investigates the possibility of extending the unified empirical likelihood inference in Zhu et al. (2014) for model (1). We refer to the book by Owen (2001) for an overview of empirical likelihood methods.

The paper is organized as follows. Section 2 presents the methodologies and main asymptotic results. A simulation study is conducted to illustrate the finite sample performance of the proposed robust test in Section 3. Section 4 reports a real data analysis. All proofs are relegated to Section 5.

2. Methodologies and asymptotic results

For testing predictability under model (3), we consider the testing hypothesis $H_0 : \beta_1 = 0$ or $H_0 : \beta_2 = 0$ or $H_0 : \beta_1 = \beta_2 = 0$ and propose a unified empirical likelihood test regardless of X_t being stationary or nearly integrated or unit root. More generally, the proposed empirical likelihood method can be employed to construct a unified confidence interval for β_1 or β_2 and a unified confidence region for $(\beta_1, \beta_2)^T$ without knowing the properties of $\{X_t\}$.

2.1. With a known intercept

To better appreciate the proposed unified test, we first consider the simple case of having a known intercept $\alpha = \alpha_0$, which may have an independent interest as well. That is to assume that observations $Y_1, \ldots, Y_n$ and $X_1, \ldots, X_{n-1}$ follow from the model

$$Y_t = \alpha_0 + \beta_1 \Delta X_{t-1} + \beta_2 X_{t-2} + U_t, \quad X_t = \theta + \phi X_{t-1} + \sum_{j=0}^{\infty} \psi_j V_{t-j}, \tag{4}$$

where, in what follows, the linear process $\sum_{j=0}^{\infty} \psi_j V_{t-j}$ is assumed to be strictly stationary,[1] and $\{(U_t, V_t)^T\}$ is a sequence of independent and identically distributed random vectors with means zero and finite variances.

It is known that a test statistic for testing predictability based on the least square estimators for β_1 and β_2 has an asymptotic distribution depending on the properties of $\{X_t\}$. Therefore, a direct application of the empirical likelihood method in Qin and Lawless (1994) fails, i.e., the so-called Wilks theorem does not hold. Motivated by the unified empirical likelihood inference idea in Zhu et al. (2014) for model (1), the following score equations are considered:

$$\begin{cases} \sum_{t=3}^{n} \{Y_t - \alpha_0 - \beta_1 \Delta X_{t-1} - \beta_2 X_{t-2}\} \Delta X_{t-1} = 0, \\ \sum_{t=3}^{n} \{Y_t - \alpha_0 - \beta_1 \Delta X_{t-1} - \beta_2 X_{t-2}\} X_{t-2} / \sqrt{1 + X_{t-2}^2} = 0, \end{cases} \tag{5}$$

which lead to the following empirical likelihood method. Clearly, the weight is only added to the second equation in (5) to ensure that $\frac{1}{n} \sum_{t=3}^{n} \{X_{t-2}/\sqrt{1 + X_{t-2}^2}\}^2$ converges in probability to a positive constant for $\{X_t\}$ being both stationary and nonstationary. Unlike the second equation, we do not need to add a weight to the first equation in (5) because $\frac{1}{n} \sum_{t=3}^{n} \{\Delta X_{t-1}\}^2$ always converges in probability to a positive constant. Now, define

$$\begin{cases} Z_{t1}(\beta_1, \beta_2) = \{Y_t - \alpha_0 - \beta_1 \Delta X_{t-1} - \beta_2 X_{t-2}\} \Delta X_{t-1}, \\ Z_{t2}(\beta_1, \beta_2) = \{Y_t - \alpha_0 - \beta_1 \Delta X_{t-1} - \beta_2 X_{t-2}\} X_{t-2} / \sqrt{1 + X_{t-2}^2}, \end{cases}$$

and let $Z_t(\beta_1, \beta_2) = (Z_{t1}(\beta_1, \beta_2), Z_{t2}(\beta_1, \beta_2))^T$ for $t = 3, \ldots, n$. Like Qin and Lawless (1994), we define the empirical likelihood function for $(\beta_1, \beta_2)^T$ via estimating equations as

$$L(\beta_1, \beta_2) = \sup \left\{ \prod_{t=3}^{n} (np_t) : p_3 \geq 0, \ldots, p_n \geq 0, \ \sum_{t=3}^{n} p_t = 1, \ \sum_{t=3}^{n} p_t Z_t(\beta_1, \beta_2) = 0 \right\}.$$

Note that the supremum is taken with respect to p_t's. It follows from the Lagrange multiplier method that the empirical likelihood ratio for β_1 and β_2 is

$$-2 \log L(\beta_1, \beta_2) = 2 \sum_{t=3}^{n} \log \left(1 + \lambda^T Z_t(\beta_1, \beta_2)\right),$$

where $\lambda = \lambda(\beta_1, \beta_2)$ satisfies

$$\frac{1}{n} \sum_{t=3}^{n} \frac{Z_t(\beta_1, \beta_2)}{1 + \lambda^T Z_t(\beta_1, \beta_2)} = 0.$$

If one is interested in testing $H_0 : \beta_1 = 0$ or constructing a unified confidence interval for β_1, then the profile empirical likelihood function $L^{P1}(\beta_1) = \max_{\beta_2} L(\beta_1, \beta_2)$ is considered. On the other hand, one considers $L^{P2}(\beta_2) = \max_{\beta_1} L(\beta_1, \beta_2)$ for testing $H_0 : \beta_2 = 0$ or constructing a unified interval for β_2.

To prove the Wilks theorem for the aforementioned empirical likelihood method, we assume the following regularity conditions:

C1. $E(U_1) = 0$, $E(V_1) = 0$, $E(|U_1|^{2+\delta} + |V_1|^{2+\delta}) < \infty$ for some $\delta > 0$ and $\{(U_t, V_t)^T\}$ is a sequence of independent and identically distributed random vectors.

Theorem 1. *Suppose model (4) holds with known α_0 and coefficients ψ_j's satisfying that the linear process $\sum_{j=0}^{\infty} \psi_j V_{t-j}$ is a strictly stationary process, and either (i) $|\phi| < 1$ independent of n (stationary case), or (ii) $\phi = 1 - \rho/n$ for some $\rho \neq 0$ (near unit root case), or (iii) $\phi = 1$ (unit root case). Then, under Condition* **C1**, *we have $-2 \log L(\beta_{1,0}, \beta_{2,0}) \xrightarrow{d} \chi^2(2)$, $-2 \log L^{P1}(\beta_{1,0}) \xrightarrow{d} \chi^2(1)$ and $-2 \log L^{P2}(\beta_{2,0}) \xrightarrow{d} \chi^2(1)$ as $n \to \infty$, where $(\beta_{1,0}, \beta_{2,0})^T$ denotes the true value of $(\beta_1, \beta_2)^T$ and $\xrightarrow{d}$ denotes the convergence in distribution.*

It follows from Theorem 1 that a unified empirical likelihood test rejects the null hypothesis $H_0 : \beta_1 = \beta_2 = 0$ at the significance level a if $-2 \log L(0, 0) > \chi^2_{2,1-a}$, where $\chi^2_{2,1-a}$ denotes the $(1 - a)$th quantile of a chi-squared distribution with two degrees of freedom. Similarly, unified empirical likelihood tests can be obtained for testing $H_0 : \beta_1 = 0$ and $H_0 : \beta_2 = 0$ based on $-2 \log L^{P1}(0)$ and $-2 \log L^{P2}(0)$, respectively. A unified interval/region can be obtained as well by Theorem 1.

[1] If $\{\psi_j\}$ in model (4) satisfies some condition, say $\sum_{j=0}^{\infty} |\psi_j| < \infty$, it is easy to show that $\sum_{j=0}^{\infty} \psi_j V_{t-j}$ is strictly stationary; see Brockwell and Davis (1991) on Page 89 for details.

Remark 1. When $\{U_t\}$ in (4) is a strictly stationary AR(p) model, Theorem 1 fails due to correlated errors. In this case, we can follow the idea in Li et al. (2017) to develop some unified empirical likelihood tests and unified jackknife empirical likelihood tests by taking the structure of $U_t's$ into account.

2.2. With an unknown intercept

This section is devoted to considering the model with an unknown intercept α, i.e.,

$$Y_t = \alpha + \beta_1 \Delta X_{t-1} + \beta_2 X_{t-2} + U_t, \quad X_t = \theta + \phi X_{t-1} + \sum_{j=0}^{\infty} \psi_j V_{t-j} \tag{6}$$

where coefficients $\psi_j's$ are assumed to satisfy some condition so that $\sum_{j=0}^{\infty} \psi_j V_{t-j}$ is a strictly stationary process as before. By the same token, one may simply apply the empirical likelihood method to some weighted score equations. However, when X_t is nearly integrated and $\theta = 0$, the joint asymptotic distribution of the score equation with respect to α and the weighted score equation with respect to β_2 is no longer a bivariate normal distribution, which makes the Wilks theorem invalid. To solve this degenerate issue, Zhu et al. (2014) proposed splitting the data into two parts and then using the differencing idea with a big lag to get rid of the intercept before formulating an empirical likelihood function for model (1). Here, we generalize their idea for model (1) to the above model in (6).

Let $m = [n/2]$ and define $\tilde{Y}_t = Y_{t+m} - Y_t$, $\tilde{X}_t = X_{t+m} - X_t$, $\tilde{U}_t = U_{t+m} - U_t$ and $\tilde{V}_t = V_{t+m} - V_t$, where $[\cdot]$ denotes the floor function. Then, model (6) implies that

$$\tilde{Y}_t = \beta_1 \Delta\tilde{X}_{t-1} + \beta_2 \tilde{X}_{t-2} + \tilde{U}_t, \quad \tilde{X}_t = \phi\tilde{X}_{t-1} + \sum_{j=0}^{\infty} \psi_j \tilde{V}_{t-j}, \tag{7}$$

where $\Delta\tilde{X}_{t-1} = \tilde{X}_{t-1} - \tilde{X}_{t-2}$. The reason of using difference with a big lag is to ensure $|\tilde{X}_t| \xrightarrow{p} \infty$ when $|X_t| \xrightarrow{p} \infty$ as $t \to \infty$. Like the proposed unified empirical likelihood tests given in Section 2.1, we define the empirical likelihood function for β_1 and β_2 in the above model (7) as follows:

$$\tilde{L}(\beta_1, \beta_2) = \sup\left\{\prod_{t=3}^{m}(mp_t) : p_3 \geq 0, \ldots, p_m \geq 0, \ \sum_{t=3}^{m} p_t = 1, \ \sum_{t=3}^{m} p_t \tilde{Z}_t(\beta_1, \beta_2) = 0\right\},$$

where $\tilde{Z}_t(\beta_1, \beta_2) = \left(\tilde{Z}_{t1}(\beta_1, \beta_2), \tilde{Z}_{t2}(\beta_1, \beta_2)\right)^T$ with $\tilde{Z}_{t1}(\beta_1, \beta_2) = \{\tilde{Y}_t - \beta_1 \Delta\tilde{X}_{t-1} - \beta_2 \tilde{X}_{t-2}\}\Delta\tilde{X}_{t-1}$ and $\tilde{Z}_{t2}(\beta_1, \beta_2) = \{\tilde{Y}_t - \beta_1 \Delta\tilde{X}_{t-1} - \beta_2 \tilde{X}_{t-2}\}\tilde{X}_{t-2}/\sqrt{1 + \tilde{X}_{t-2}^2}$.

When the interest is in testing $H_0 : \beta_1 = 0$ or $H_0 : \beta_2 = 0$, one considers the profile empirical likelihood function $\tilde{L}^{P1}(\beta_1) = \max_{\beta_2} \tilde{L}(\beta_1, \beta_2)$ or $\tilde{L}^{P2}(\beta_2) = \max_{\beta_1} \tilde{L}(\beta_1, \beta_2)$, respectively. The following result shows that the Wilks theorem holds for the above proposed empirical likelihood method.

Theorem 2. *Under model (6), Condition* **C1** *and the same conditions on ϕ and $\psi_j's$ as in Theorem 1, we have that* $-2\log\tilde{L}(\beta_{1,0}, \beta_{2,0}) \xrightarrow{d} \chi^2(2)$, $-2\log\tilde{L}^{P1}(\beta_{1,0}) \xrightarrow{d} \chi^2(1)$ *and* $-2\log\tilde{L}^{P2}(\beta_{2,0}) \xrightarrow{d} \chi^2(1)$ *as* $n \to \infty$.

In view of Theorem 2, a unified empirical likelihood test rejects the null hypothesis $H_0 : \beta_1 = \beta_2 = 0$ at the significance level a for model (6) if $-2\log\tilde{L}(0, 0) > \chi^2_{2,1-a}$. Similarly, unified empirical likelihood tests can be obtained for testing $H_0 : \beta_1 = 0$ and $H_0 : \beta_2 = 0$ for model (6) based on $-2\log\tilde{L}^{P1}(0)$ and $-2\log\tilde{L}^{P2}(0)$, respectively. Again, a unified interval/region can be obtained as well by Theorem 2.

2.3. An extension to multiple regressors

When X_t in (4) is a multivariate vector, in which one of variables may be nonstationary and the rest of the variables are surely stationary, we could simply replace the weight $\sqrt{1 + X_{t-2}^2}$ in (5) by $\sqrt{1 + X_{t-2}^T X_{t-2}}$ and show that Theorem 1 still holds. However, this is not true when more than one variables in X_t may be nonstationary. Here, we propose an easy but inefficient way to extend the model in (4) to the case of a bivariate predicting vector by splitting the data into two parts and using each part to derive the score equations with respect to each variable in X_t. Therefore, the idea of splitting data here is different from the previous one, which uses differences to get rid of the intercept before formulating an empirical likelihood function.

Consider the following model

$$Y_t = \alpha_0 + \beta_1^T \Delta X_{t-1} + \beta_2^T X_{t-2} + U_t, \quad X_t = \theta + \phi X_{t-1} + \sum_{j=0}^{\infty} \psi_j V_{t-j}, \tag{8}$$

where β_1, β_2, X_t and θ are 2×1 vector, ϕ and ψ are 2×2 matrix, and α_0 is known. Note that we always write $\beta_i = (\beta_{i,1}, \beta_{i,2})^T$. Similar to the empirical likelihood function formulated in Section 2.1, one may use the following weighted score equations:

$$\begin{cases} \sum_{t=3}^{n}\{Y_t-\alpha_0-\beta_1^T\Delta X_{t-1}-\beta_2^T X_{t-2}\}\Delta X_{t-1}=0, \\ \sum_{t=3}^{n}\{Y_t-\alpha_0-\beta_1^T\Delta X_{t-1}-\beta_2^T X_{t-2}\}\frac{X_{t-2,1}}{\sqrt{1+X_{t-2,1}^2}}=0, \\ \sum_{t=3}^{n}\{Y_t-\alpha_0-\beta_1^T\Delta X_{t-1}-\beta_2^T X_{t-2}\}\frac{X_{t-2,2}}{\sqrt{1+X_{t-2,2}^2}}=0. \end{cases} \tag{9}$$

As before, the added weights ensure that

$$\text{both } \frac{1}{n}\sum_{t=3}^{n}\left\{X_{t-2,1}/\sqrt{1+X_{t-2,1}^2}\right\}^2 \text{ and } \frac{1}{n}\sum_{t=3}^{n}\left\{X_{t-2,2}/\sqrt{1+X_{t-2,2}^2}\right\}^2$$

converge in probability. However, the normalized product of the right sides of the second and third equations in (9) does not converge in probability in all cases. More specifically,

$$\frac{1}{n}\sum_{t=3}^{n}\{Y_t-\alpha_0-\beta_1^{(0)T}\Delta X_{t-1}-\beta_2^{(0)T}X_{t-2}\}^2\frac{X_{t-2,1}}{\sqrt{1+X_{t-2,1}^2}}\frac{X_{t-2,2}}{\sqrt{1+X_{t-2,2}^2}}$$

converges in distribution to a nondegenerate random variable rather than in probability to a constant when X_t is nonstationary and $\theta=0$ in (8), where $\beta_i^{(0)}$ to denote the true value of β_i. That is, the joint distribution of the normalized right sides of the score equations in (9) cannot be normal for all cases. Hence, a direct application of the empirical likelihood inference to the estimating equations in (9) fails to have a chi-squared limit for all cases.

In order to derive a unified inference when α is known, we propose to first splitting the data into two parts and then applying each part to the score equations with respect to each predicting variable. More specifically, let $m=[n/2]$ and define for $t=1,\ldots,m$

$$\begin{cases} \bar{Z}_{t,1}(\beta_1,\beta_2)=\{Y_t-\alpha_0-\beta_1^T\Delta X_{t-1}-\beta_2^T X_{t-2}\}\Delta X_{t-1,1}, \\ \bar{Z}_{t,2}(\beta_1,\beta_2)=\{Y_t-\alpha_0-\beta_1^T\Delta X_{t-1}-\beta_2^T X_{t-2}\}\frac{X_{t-2,1}}{\sqrt{1+X_{t-2,1}^2}}, \\ \bar{Z}_{t,3}(\beta_1,\beta_2)=\{Y_{t+m}-\alpha_0-\beta_1^T\Delta X_{t+m-1}-\beta_2^T X_{t+m-2}\}\Delta X_{t+m-1,2}, \\ \bar{Z}_{t,4}(\beta_1,\beta_2)=\{Y_{t+m}-\alpha_0-\beta_1^T\Delta X_{t+m-1}-\beta_2^T X_{t+m-2}\}\frac{X_{t+m-2,2}}{\sqrt{1+X_{t+m-2,2}^2}}, \\ \bar{Z}_t(\beta_1,\beta_2)=(\bar{Z}_{t,1}(\beta_1,\beta_2),\ldots,\bar{Z}_{t,4}(\beta_1,\beta_2))^T. \end{cases}$$

It is clear that the first part of data is used to derive scores with respect to $\beta_{1,1}$ and $\beta_{2,1}$, while the second part of data constructs scores with respect to $\beta_{1,2}$ and $\beta_{2,2}$. Now due to the independence of U_t and U_{t+m}, we have $\frac{1}{m}\sum_{t=3}^{m}\bar{Z}_{t,2}(\beta_1^{(0)},\beta_2^{(0)})\bar{Z}_{t,4}(\beta_1^{(0)},\beta_2^{(0)})\xrightarrow{p}0$ as $n\to\infty$, which is necessary to ensure that Wilks theorem holds for the application of the empirical likelihood method to the above $\bar{Z}_t(\beta_1,\beta_2)$, $t=3,\ldots,m$.

Define the empirical likelihood function for β_1 and β_2 as

$$\bar{L}(\beta_1,\beta_2)=\sup\left\{\prod_{t=3}^{m}(mp_t): p_3\geq 0,\ldots,p_m\geq 0,\ \sum_{t=3}^{m}p_t=1,\ \sum_{t=3}^{m}p_t\bar{Z}_t(\beta_1,\beta_2)=0\right\}.$$

To prove the Wilks theorem, we need the following regularity conditions:

- **D1.** $E(|U_1|^{2+\delta}+|V_{1,1}|^{2+\delta}+|V_{1,2}|^{2+\delta})<\infty$ for some $\delta>0$ and $\{(U_t,V_t)^T\}$ is a sequence of independent and identically distributed random vector with means zero.
- **D2.** $\{\psi_t\}$ satisfies that $\{\sum_{j=0}^{\infty}\psi_j V_{t-j}\}$ is a strictly stationary process.
- **D3.** Either ϕ given in (8) satisfies that $|I_2-\gamma\phi|\neq 0$ for all $|\gamma|\leq 1$ (stationary), or $\phi=I_2-\rho/n$, where I_2 denotes the 2×2 identity matrix and ρ is a 2×2 matrix (unit root or nearly integrated).

Theorem 3. *Under model (8) and Conditions* **D1–D3**, *we have that* $-2\log\bar{L}(\beta_1^{(0)},\beta_2^{(0)})\xrightarrow{d}\chi^2(4)$ *as* $n\to\infty$.

The consequence of Theorem 3 is that a unified empirical likelihood test rejects the null hypothesis $H_0 : (\beta_1^T, \beta_2^T)^T = 0$ at the significance level a for model (8) if $-2\log \bar{L}(0, 0) > \chi^2_{4,1-a}$. If one is interested in testing $H_0 : \beta_{1,1} = 0$, then a unified empirical likelihood test is based on the profile empirical likelihood ratio $-2\log(\max_{\beta_{1,2},\beta_2} \bar{L}(\beta_1, \beta_2))$, which has a chi-squared distribution with one degree of freedom under the null hypothesis. Similarly, a unified empirical likelihood test can be obtained for testing other null hypotheses such as $H_0 : \beta_1 = 0$ or $H_0 : \beta_2 = 0$.

Remark 2. When X_t is a d-dimensional vector, which may be nonstationary for each variable, the above proposed unified test has to split the data into d parts and use each part to construct score equations with respect to each variable in X_t, which becomes quite inefficient. Moreover, when the intercept α is unknown, getting rid of the intercept via differencing is needed too. Therefore, developing an efficient unified predictability test for a multivariate predictive regression is challenging and is believed to be beyond the scope of this paper. This will be one of our future research projects.

3. A simulation study

In this section, we investigate the finite sample performance of the proposed unified tests for predictive regressions using simulated data sets.

Let $\{(U_t, V_t)^T\}_{t=1}^n$ be a random sample from a bivariate Gaussian Copula $C(F_1(U_t), F_2(V_t), \eta)$, where the marginal distribution F_i is a Student's t distribution with degrees of freedom ν_i for $i = 1,2$. The dependence parameter is set to be $\eta = -0.5$, which captures the correlation between two innovations. The negative sign is based on a real example of predicting the stock return by the log dividend-price or the log earning-price in Cai and Wang (2014). For copula models and their applications, we refer to the books by Joe (1997), Jaworski et al. (2013, 2010). We take $\nu_1 = 5$ and $\nu_2 = 4$ and choose $\psi_j = 1$ for $j = 0, \ldots, q-1$ and $\psi_j = 0$ for $j \geq q$ with $q = 1$ or 5 in $\sum_{j=0}^{\infty} \psi_j V_{t-j}$, $\theta = 0, \alpha = 1, \phi = 0.9$ or $1 - 2/n$ or 1 with $n = 200$ or 500 or 2,000 in either model (4) or (6). For testing $H_0 : \beta_1 = 0$ in either model (4) or model (6), we consider $\beta_2 = 0.5$ and $\beta_1 = d/n^{0.5}$ for $d = 0, \pm1, \pm2, \pm3, \pm4$. On the other hand, for testing $H_0 : \beta_2 = 0$ in either model (4) or (6), we consider $\beta_1 = 0.5$ and $\beta_2 = d/\sqrt{n}$ for $d = 0, \pm1, \pm2, \pm3, \pm4$ when $\phi = 0.9$, and $\beta_1 = 0.5$ and $\beta_2 = d/n$ for $d = 0, \pm2, \pm4, \pm6, \pm8$ when $\phi = 1 - 2/n$ and 1. Therefore, results for $d = 0$ and $d \neq 0$ represent the empirical sizes and empirical powers of the proposed unified tests, respectively.

By drawing 10,000 random samples from either model (4) or model (6) under the above settings, we compute the unified profile empirical likelihood functions for testing $H_0 : \beta_1 = 0$ and $H_0 : \beta_2 = 0$ in Theorems 1 and 2 by using the package "emplik" in the statistical software R, and report the empirical sizes and powers in Tables 1–8. From these tables, we could summarize our observations as follows. Firstly, the test for $H_0 : \beta_2 = 0$ has an accurate size for all sample sizes considered, but the test for $H_0 : \beta_1 = 0$ does not have an accurate size when $n = 200$, which is because β_2 is estimated at the rate of n^{-1} while β_1 is estimated at the rate of $n^{-1/2}$. Secondly, tests based on Theorem 1 are more powerful than those based on Theorem 2 due to the employed technique of splitting data into two parts to get rid of the intercept. Thirdly, the proposed unified tests have an overall satisfactory power and their powers become larger as the alternative hypothesis is far away from the null hypothesis. Furthermore, for testing $H_0 : \beta_2 = 0$, the test for the case of unit root is more powerful than that for the nearly integrated case. Finally, results for $q = 1$ and $q = 5$ show that our method is robust against the dependent or independent errors in modeling the predicting variable.

4. A real example

This section is devoted to applying the proposed unified tests to testing the predictability of the asset returns with some U.S. equity data, where the sample period is January 1927 to December 2012 with monthly data as in Kostakis et al. (2015). The predicted variable is the CRSP value-weighted excess returns and some predicting variables are the difference between the log of dividends and the log of earnings (dividend payout ratio), the long-term US government bond yield (long-term yield), the difference between the log of dividends and the log of the lagged prices (dividend yield), the difference between the log of dividends and the log of stock prices (dividend price ratio), the 3-month US Treasury bill rate (T-bill rate), the difference between the log of earnings and the log of prices (earnings price ratio), the ratio of book value to market value for the Dow Jones Industrial Average (book-to-Market value ratio), the difference between the BAA and AAA-rated corporate bond yields (default yield spread), the ratio of 12-month moving sum of net equity issues by NYSE listed stocks divided by the total end-of-year market capitalization of these stocks (net equity expansion), and the difference between the long-term yield and the T-bill rate (term spread).

Before applying the proposed unified tests to the above data, we examine whether both $U_t's$ and $V_t's$ in models (4) and (6) are uncorrelated. By fitting model (6) with the least squares estimate, we plot the autocorrelation functions of the estimated $U_t's$ for the period 1/1927–12/2012 and the period 1/1952–12/2012, which suggest that the assumption of uncorrelated $U_t's$ in model (6) is reasonable for the period 1/1952–12/2012. Therefore our study will focus on the period 1/1952–12/2012. Plots of the autocorrelation functions also show that the assumption of uncorrelated $U_t's$ in model (4) with known $\alpha_0 = 0$

Table 1
Empirical sizes and powers for the test based on Theorem 1 and model (4) with $\theta = 0$ and known $\alpha = 1$ for testing $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$. We take $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$ in $\sum_{j=0}^{\infty} \psi_j V_{t-j}$.

(β_1, ϕ)	$n = 200$		$n = 500$		$n = 2000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.12	0.07	0.11	0.06	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.30	0.21	0.30	0.21	0.30	0.19
$(\frac{1}{\sqrt{n}}, 0.9)$	0.33	0.23	0.31	0.21	0.30	0.20
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.69	0.58	0.69	0.57	0.70	0.57
$(\frac{2}{\sqrt{n}}, 0.9)$	0.71	0.60	0.70	0.58	0.70	0.58
$(-\frac{3}{\sqrt{n}}, 0.9)$	0.92	0.86	0.93	0.87	0.94	0.90
$(\frac{3}{\sqrt{n}}, 0.9)$	0.92	0.88	0.93	0.88	0.94	0.90
$(-\frac{4}{\sqrt{n}}, 0.9)$	0.98	0.97	0.99	0.98	0.99	0.99
$(\frac{4}{\sqrt{n}}, 0.9)$	0.99	0.90	0.99	0.98	0.99	0.99
$(0, 1-\frac{2}{n})$	0.12	0.06	0.11	0.06	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 1-\frac{2}{n})$	0.30	0.21	0.30	0.20	0.30	0.20
$(\frac{1}{\sqrt{n}}, 1-\frac{2}{n})$	0.33	0.23	0.32	0.21	0.30	0.21
$(-\frac{2}{\sqrt{n}}, 1-\frac{2}{n})$	0.69	0.58	0.69	0.58	0.70	0.58
$(\frac{2}{\sqrt{n}}, 1-\frac{2}{n})$	0.71	0.60	0.70	0.59	0.70	0.59
$(-\frac{3}{\sqrt{n}}, 1-\frac{2}{n})$	0.92	0.86	0.93	0.88	0.94	0.90
$(\frac{3}{\sqrt{n}}, 1-\frac{2}{n})$	0.93	0.88	0.94	0.89	0.94	0.90
$(-\frac{4}{\sqrt{n}}, 1-\frac{2}{n})$	0.98	0.97	0.99	0.98	1.00	0.99
$(\frac{4}{\sqrt{n}}, 1-\frac{2}{n})$	0.99	0.97	0.99	0.98	0.99	0.99
$(0, 1)$	0.12	0.06	0.11	0.06	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 1)$	0.30	0.21	0.30	0.20	0.30	0.19
$(\frac{1}{\sqrt{n}}, 1)$	0.33	0.23	0.32	0.22	0.30	0.21
$(-\frac{2}{\sqrt{n}}, 1)$	0.69	0.58	0.69	0.58	0.70	0.58
$(\frac{2}{\sqrt{n}}, 1)$	0.72	0.61	0.70	0.60	0.70	0.59
$(-\frac{3}{\sqrt{n}}, 1)$	0.92	0.86	0.93	0.88	0.94	0.90
$(\frac{3}{\sqrt{n}}, 1)$	0.93	0.88	0.94	0.89	0.94	0.90
$(-\frac{4}{\sqrt{n}}, 1)$	0.98	0.97	0.99	0.98	0.99	0.99
$(\frac{4}{\sqrt{n}}, 1)$	0.99	0.97	0.99	0.98	0.99	0.99

is reasonable, and confirms the setting for the considered predicting variables is suitable. To save space as suggested by a reviewer, we only report the autocorrelation function for the predicting variable, dividend payout ratio in Fig. 1. Other plots are available upon request.

Next, we employ our proposed unified tests to model (4) with known $\alpha_0 = 0$ and to model (6) with unknown α for the period 1/1952–12/2012. P-values for testing $H_0 : \beta_1 = 0$, $H_0 : \beta_2 = 0$ and $H_0 : \beta_1 = \beta_2 = 0$ are reported in Tables 9 and 10, respectively. Results in Table 9 show predictability for each predicting variable, while results in Table 10 show no predictability except the difference of the long-term yield. This finding shows that a small intercept can change the predictability dramatically when the predicting variable is nearly integrated. This may be due to the drawback of our method in Theorem 2, which requires splitting the data into two parts so that efficiency is reduced. Without doubt, it is important and challenging to derive a unified test for testing zero intercept regardless of the predicting variable being stationary or nearly integrated or unit root, which is warranted as one of future research projects. On the other hand, results in Table 10 are in line with the conclusion of no predictability at level 5% for the considered predicting variables in Table 6 of Kostakis et al. (2015), which studies model (6) without the term ΔX_{t-1}. However, they found predictability for the predicting variables of T-bill rate and term spread at level 10%. Therefore, it remains interesting to extend the IVX-based test in Kostakis et al. (2015) to the larger model (6) and to see whether the predictability for these two predicting variables at level 10% will disappear due to the fact that an additional parameter has to be estimated. Since the IVX method involves the construction of an instrumental

Table 2
Empirical sizes and powers for the test based on Theorem 1 and model (4) with $\theta = 0$ and known $\alpha = 1$ for testing $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$. We take $\psi_0 = \cdots = \psi_4 = 1$ and $\psi_j = 0$ for $j \geq 5$ in $\sum_{j=0}^{\infty} \psi_j V_{t-j}$.

(β_1, ϕ)	$n = 200$		$n = 500$		$n = 2000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.12	0.07	0.12	0.06	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.31	0.21	0.30	0.20	0.30	0.19
$(\frac{1}{\sqrt{n}}, 0.9)$	0.32	0.22	0.31	0.21	0.30	0.20
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.68	0.57	0.69	0.58	0.70	0.58
$(\frac{2}{\sqrt{n}}, 0.9)$	0.70	0.59	0.69	0.58	0.70	0.58
$(-\frac{3}{\sqrt{n}}, 0.9)$	0.91	0.86	0.93	0.88	0.94	0.89
$(\frac{3}{\sqrt{n}}, 0.9)$	0.92	0.87	0.93	0.88	0.94	0.89
$(-\frac{4}{\sqrt{n}}, 0.9)$	0.98	0.97	0.99	0.98	0.99	0.98
$(\frac{4}{\sqrt{n}}, 0.9)$	0.98	0.97	0.99	0.98	0.99	0.99
$(0, 1-\frac{2}{n})$	0.12	0.07	0.11	0.06	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 1-\frac{2}{n})$	0.31	0.21	0.30	0.20	0.30	0.20
$(\frac{1}{\sqrt{n}}, 1-\frac{2}{n})$	0.33	0.22	0.31	0.22	0.31	0.20
$(-\frac{2}{\sqrt{n}}, 1-\frac{2}{n})$	0.69	0.58	0.69	0.58	0.70	0.59
$(\frac{2}{\sqrt{n}}, 1-\frac{2}{n})$	0.70	0.55	0.70	0.59	0.71	0.59
$(-\frac{3}{\sqrt{n}}, 1-\frac{2}{n})$	0.92	0.86	0.93	0.88	0.95	0.90
$(\frac{3}{\sqrt{n}}, 1-\frac{2}{n})$	0.93	0.88	0.94	0.88	0.94	0.90
$(-\frac{4}{\sqrt{n}}, 1-\frac{2}{n})$	0.98	0.97	0.99	0.98	0.99	0.99
$(\frac{4}{\sqrt{n}}, 1-\frac{2}{n})$	0.99	0.97	0.99	0.98	0.99	0.99
$(0, 1)$	0.12	0.07	0.11	0.06	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 1)$	0.31	0.21	0.30	0.20	0.30	0.20
$(\frac{1}{\sqrt{n}}, 1)$	0.33	0.22	0.32	0.22	0.31	0.20
$(-\frac{2}{\sqrt{n}}, 1)$	0.68	0.57	0.69	0.58	0.70	0.59
$(\frac{2}{\sqrt{n}}, 1)$	0.71	0.60	0.71	0.59	0.71	0.59
$(-\frac{3}{\sqrt{n}}, 1)$	0.92	0.86	0.93	0.88	0.94	0.90
$(\frac{3}{\sqrt{n}}, 1)$	0.93	0.88	0.94	0.89	0.94	0.90
$(-\frac{4}{\sqrt{n}}, 1)$	0.99	0.97	0.99	0.98	0.99	0.99
$(\frac{4}{\sqrt{n}}, 1)$	0.99	0.97	0.99	0.98	0.99	0.99

variable, such an extension to the larger model (6) is non-trivial at all although we expect the power of the IVX-based test might be reduced because of the larger model.

To compare with results in Campbell and Yogo (2006), we also apply our unified tests to the period 1/1952–12/2002, and report P-values in Tables 11 and 12. Again, when intercept is assumed to be known zero, results in Table 11 show predictability for each predicting variable. When the intercept is assumed to be unknown, unlike Table 10 for the longer period 1/1952–12/2012, results in Table 12 show that the CRSP value-weighted excess return can be predicted by the T-bill rate and its difference, the term spread and its difference, the difference of the long-term yield, and the difference of the net equity expansion. For model (1), Campbell and Yogo (2006) found predictability for T-bill rate and no predictability for dividend price ratio and earnings price ratio, which are in line with corresponding results in Table 12. On the other hand, they found predictability for the default yield spread while the corresponding result in Table 12 does not lead to this predictability.

5. Proofs

We provide only the proof of Theorem 1 under the following setting:

$$\theta = 0, \quad \phi = 1 - \rho/n \quad \text{for some} \quad \rho \in R \tag{10}$$

Table 3
Empirical sizes and powers for the test based on Theorem 2 and model (6) with $\theta = 0$ and unknown α for testing $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$. We take $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$ in $\sum_{j=0}^{\infty} \psi_j V_{t-j}$.

(β_1, ϕ)	$n = 200$		$n = 500$		$n = 2000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.13	0.07	0.11	0.06	0.11	0.06
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.22	0.14	0.21	0.13	0.20	0.12
$(\frac{1}{\sqrt{n}}, 0.9)$	0.24	0.15	0.22	0.14	0.21	0.13
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.47	0.35	0.47	0.35	0.45	0.33
$(\frac{2}{\sqrt{n}}, 0.9)$	0.49	0.38	0.48	0.35	0.47	0.34
$(-\frac{3}{\sqrt{n}}, 0.9)$	0.72	0.61	0.73	0.62	0.74	0.62
$(\frac{3}{\sqrt{n}}, 0.9)$	0.74	0.63	0.74	0.63	0.75	0.64
$(-\frac{4}{\sqrt{n}}, 0.9)$	0.89	0.82	0.90	0.84	0.91	0.85
$(\frac{4}{\sqrt{n}}, 0.9)$	0.90	0.84	0.91	0.85	0.91	0.85
$(0, 1-\frac{2}{n})$	0.13	0.07	0.11	0.06	0.11	0.05
$(-\frac{1}{\sqrt{n}}, 1-\frac{2}{n})$	0.22	0.14	0.21	0.13	0.20	0.12
$(\frac{1}{\sqrt{n}}, 1-\frac{2}{n})$	0.24	0.15	0.22	0.14	0.21	0.13
$(-\frac{2}{\sqrt{n}}, 1-\frac{2}{n})$	0.47	0.35	0.47	0.35	0.46	0.33
$(\frac{2}{\sqrt{n}}, 1-\frac{2}{n})$	0.50	0.38	0.49	0.36	0.47	0.35
$(-\frac{3}{\sqrt{n}}, 1-\frac{2}{n})$	0.72	0.62	0.73	0.62	0.74	0.63
$(\frac{3}{\sqrt{n}}, 1-\frac{2}{n})$	0.74	0.64	0.75	0.64	0.75	0.65
$(-\frac{4}{\sqrt{n}}, 1-\frac{2}{n})$	0.89	0.83	0.90	0.84	0.91	0.85
$(\frac{4}{\sqrt{n}}, 1-\frac{2}{n})$	0.91	0.85	0.91	0.85	0.92	0.86
$(0, 1)$	0.13	0.07	0.11	0.06	0.11	0.05
$(-\frac{1}{\sqrt{n}}, 1)$	0.21	0.13	0.20	0.13	0.20	0.12
$(\frac{1}{\sqrt{n}}, 1)$	0.24	0.16	0.22	0.14	0.21	0.13
$(-\frac{2}{\sqrt{n}}, 1)$	0.46	0.34	0.46	0.35	0.45	0.33
$(\frac{2}{\sqrt{n}}, 1)$	0.51	0.39	0.50	0.37	0.48	0.35
$(-\frac{3}{\sqrt{n}}, 1)$	0.71	0.61	0.73	0.62	0.74	0.63
$(\frac{3}{\sqrt{n}}, 1)$	0.75	0.65	0.75	0.65	0.76	0.65
$(-\frac{4}{\sqrt{n}}, 1)$	0.89	0.82	0.90	0.84	0.91	0.85
$(\frac{4}{\sqrt{n}}, 1)$	0.85	0.96	0.91	0.86	0.92	0.86

since proofs for other cases are similar and sometimes simpler. For example, when $\{X_t\}$ is stationary, Theorems 1 and 2 can be shown easily by using the weak law of large numbers and central limit theorem for martingales in Hall and Heyde (1980) and the standard arguments in proving the profile empirical likelihood method based on estimating equations in Qin and Lawless (1994). Before proving Theorem 1, we need some lemmas.

Lemma 1. *Under conditions of Theorem 1 and (10), we have that*

$$\frac{1}{\sqrt{n}}\sum_{t=3}^{n} Z_{t1}(\beta_{1,0}, \beta_{2,0}) = \frac{1}{\sqrt{n}}\sum_{t=3}^{n} U_t\{\sum_{j=0}^{\infty} \psi_j V_{t-1-j}\} + o_p(1)$$

and

$$\frac{1}{\sqrt{n}}\sum_{t=3}^{n} Z_{t2}(\beta_{1,0}, \beta_{2,0}) = \{\frac{1}{\sqrt{n}}\sum_{t=3}^{n} U_t\}\frac{X_{n-2}}{\sqrt{1+X_{n-2}^2}} + o_p(1).$$

Table 4
Empirical sizes and powers for the test based on Theorem 2 and model (6) with $\theta = 0$ and unknown α for testing $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$. We take $\psi_0 = \cdots = \psi_4 = 1$ and $\psi_j = 0$ for $j \geq 5$ in $\sum_{j=0}^{\infty} \psi_j V_{t-j}$.

(β_1, ϕ)	$n = 200$		$n = 500$		$n = 2000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.12	0.07	0.11	0.06	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.22	0.14	0.20	0.13	0.20	0.12
$(\frac{1}{\sqrt{n}}, 0.9)$	0.24	0.16	0.22	0.14	0.21	0.13
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.46	0.35	0.46	0.34	0.46	0.33
$(\frac{2}{\sqrt{n}}, 0.9)$	0.49	0.37	0.47	0.36	0.46	0.33
$(-\frac{3}{\sqrt{n}}, 0.9)$	0.72	0.62	0.72	0.61	0.74	0.63
$(\frac{3}{\sqrt{n}}, 0.9)$	0.74	0.63	0.74	0.63	0.74	0.63
$(-\frac{4}{\sqrt{n}}, 0.9)$	0.89	0.83	0.90	0.84	0.91	0.85
$(\frac{4}{\sqrt{n}}, 0.9)$	0.90	0.84	0.91	0.85	0.92	0.85
$(0, 1-\frac{2}{n})$	0.12	0.07	0.11	0.06	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 1-\frac{2}{n})$	0.22	0.14	0.20	0.13	0.20	0.12
$(\frac{1}{\sqrt{n}}, 1-\frac{2}{n})$	0.24	0.16	0.22	0.14	0.22	0.13
$(-\frac{2}{\sqrt{n}}, 1-\frac{2}{n})$	0.47	0.35	0.46	0.35	0.46	0.34
$(\frac{2}{\sqrt{n}}, 1-\frac{2}{n})$	0.49	0.38	0.48	0.36	0.47	0.35
$(-\frac{3}{\sqrt{n}}, 1-\frac{2}{n})$	0.73	0.62	0.73	0.62	0.75	0.64
$(\frac{3}{\sqrt{n}}, 1-\frac{2}{n})$	0.75	0.65	0.75	0.64	0.75	0.64
$(-\frac{4}{\sqrt{n}}, 1-\frac{2}{n})$	0.89	0.83	0.91	0.84	0.92	0.86
$(\frac{4}{\sqrt{n}}, 1-\frac{2}{n})$	0.91	0.85	0.91	0.85	0.92	0.86
$(0, 1)$	0.12	0.07	0.11	0.06	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 1)$	0.22	0.14	0.20	0.12	0.20	0.12
$(\frac{1}{\sqrt{n}}, 1)$	0.25	0.17	0.22	0.14	0.21	0.13
$(-\frac{2}{\sqrt{n}}, 1)$	0.45	0.34	0.46	0.34	0.45	0.33
$(\frac{2}{\sqrt{n}}, 1)$	0.50	0.39	0.49	0.37	0.47	0.35
$(-\frac{3}{\sqrt{n}}, 1)$	0.72	0.61	0.72	0.61	0.74	0.63
$(\frac{3}{\sqrt{n}}, 1)$	0.76	0.65	0.75	0.65	0.75	0.64
$(-\frac{4}{\sqrt{n}}, 1)$	0.89	0.82	0.90	0.84	0.92	0.85
$(\frac{4}{\sqrt{n}}, 1)$	0.91	0.85	0.91	0.86	0.92	0.86

Proof. It follows from Phillips (1987) that under (10),

$$\frac{1}{\sqrt{n}} X_{[nr]} \Rightarrow J_\rho(r) \quad \text{in the space } D[0,1], \tag{11}$$

where $J_\rho(r) = \int_0^r e^{-(r-s)\rho} dW(s)$ and $W(s)$ is a centered Brownian process. Indeed, $J_\rho(\cdot)$ is a geometric Brownian process. Here, "$\Rightarrow$" denotes the weak convergence and the space $D[0, 1]$ is the collection of real-valued functions on $[0, 1]$ which are right continuous with left limits; see Billingsley (1999) for details. By (11), we can write that

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=3}^{n} Z_{t1}(\beta_{1,0}, \beta_{2,0}) = \frac{1}{\sqrt{n}} \sum_{t=3}^{n} U_t(X_{t-1} - X_{t-2}) \\ = & \frac{1}{\sqrt{n}} \sum_{t=3}^{n} U_t\{\sum_{j=0}^{\infty} \psi_j V_{t-1-j} - \frac{\rho}{n} X_{t-2}\} \\ = & \frac{1}{\sqrt{n}} \sum_{t=3}^{n} U_t\{\sum_{j=0}^{\infty} \psi_j V_{t-1-j}\} + o_p(1) \end{aligned} \tag{12}$$

Table 5
Empirical sizes and powers for the test based on Theorem 1 and model (4) with $\theta = 0$ and known $\alpha = 1$ for testing $H_0 : \beta_2 = 0$ vs $H_a : \beta_2 \neq 0$. We take $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$ in $\sum_{j=0}^{\infty} \psi_j V_{t-j}$.

(β_2, ϕ)	$n = 200$		$n = 500$		$n = 2000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.10	0.05	0.10	0.05	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.64	0.53	0.66	0.55	0.67	0.55
$(\frac{1}{\sqrt{n}}, 0.9)$	0.68	0.55	0.69	0.57	0.69	0.57
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.96	0.93	0.98	0.96	0.99	0.98
$(\frac{2}{\sqrt{n}}, 0.9)$	0.99	0.98	1.00	0.99	1.00	0.99
$(-\frac{3}{\sqrt{n}}, 0.9)$	1.00	0.99	1.00	1.00	1.00	1.00
$(\frac{3}{\sqrt{n}}, 0.9)$	1.00	1.00	1.00	1.00	1.00	1.00
$(-\frac{4}{\sqrt{n}}, 0.9)$	1.00	1.00	1.00	1.00	1.00	1.00
$(\frac{4}{\sqrt{n}}, 0.9)$	1.00	1.00	1.00	1.00	1.00	1.00
$(0, 1-\frac{2}{n})$	0.10	0.05	0.10	0.05	0.10	0.05
$(-\frac{2}{n}, 1-\frac{2}{n})$	0.24	0.16	0.24	0.16	0.23	0.16
$(\frac{2}{n}, 1-\frac{2}{n})$	0.19	0.11	0.18	0.10	0.17	0.10
$(-\frac{4}{n}, 1-\frac{2}{n})$	0.44	0.35	0.45	0.36	0.44	0.35
$(\frac{4}{n}, 1-\frac{2}{n})$	0.46	0.33	0.45	0.33	0.44	0.31
$(-\frac{6}{n}, 1-\frac{2}{n})$	0.62	0.54	0.64	0.55	0.62	0.54
$(\frac{6}{n}, 1-\frac{2}{n})$	0.73	0.62	0.72	0.61	0.72	0.60
$(-\frac{8}{n}, 1-\frac{2}{n})$	0.76	0.69	0.75	0.69	0.76	0.69
$(\frac{8}{n}, 1-\frac{2}{n})$	0.88	0.81	0.88	0.81	0.88	0.80
$(0, 1)$	0.10	0.05	0.10	0.05	0.10	0.05
$(-\frac{2}{n}, 1)$	0.36	0.28	0.37	0.29	0.36	0.28
$(\frac{2}{n}, 1)$	0.32	0.22	0.32	0.21	0.32	0.21
$(-\frac{4}{n}, 1)$	0.60	0.53	0.61	0.53	0.61	0.53
$(\frac{4}{n}, 1)$	0.70	0.59	0.70	0.60	0.69	0.58
$(-\frac{6}{n}, 1)$	0.76	0.70	0.76	0.70	0.75	0.70
$(\frac{6}{n}, 1)$	0.89	0.83	0.88	0.82	0.88	0.82
$(-\frac{8}{n}, 1)$	0.86	0.81	0.85	0.80	0.85	0.81
$(\frac{8}{n}, 1)$	0.96	0.92	0.95	0.92	0.96	0.93

and

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=3}^{n} Z_{t2}(\beta_{1,0},\beta_{2,0}) &= \frac{1}{\sqrt{n}}\sum_{t=3}^{n}\{\sum_{j=1}^{t}U_j - \sum_{j=1}^{t-1}U_j\}\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}} \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}U_j\frac{X_{n-2}}{\sqrt{1+X_{n-2}^2}} + \frac{1}{\sqrt{n}}\sum_{t=3}^{n-1}\{\sum_{j=1}^{t}U_j\}\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}} - \frac{1}{\sqrt{n}}\sum_{t=3}^{n}\{\sum_{j=1}^{t-1}U_j\}\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}} \qquad (13) \\
&= \{\frac{1}{\sqrt{n}}\sum_{t=3}^{n}U_t\}\frac{X_{n-2}}{\sqrt{1+X_{n-2}^2}} + \frac{1}{\sqrt{n}}\sum_{t=3}^{n-1}\{\sum_{j=1}^{t}U_j\}\{\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}} - \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}\} + o_p(1).
\end{aligned}
$$

It follows from Taylor expansion that

$$
\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}} - \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} = (1+\xi_t^2)^{-3/2}(X_{t-2}-X_{t-1}), \qquad (14)
$$

Table 6
Empirical sizes and powers for the test based on Theorem 1 and model (4) with $\theta = 0$ and known $\alpha = 1$ for testing $H_0 : \beta_2 = 0$ vs $H_a : \beta_2 \neq 0$. We take $\psi_0 = \cdots = \psi_4 = 1$ and $\psi_j = 0$ for $j \geq 5$ in $\sum_{j=0}^{\infty} \psi_j V_{t-j}$.

(β_2, ϕ)	$n = 200$		$n = 500$		$n = 2000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.10	0.05	0.10	0.05	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.64	0.53	0.66	0.55	0.66	0.55
$(\frac{1}{\sqrt{n}}, 0.9)$	0.68	0.55	0.60	0.57	0.69	0.57
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.95	0.93	0.98	0.96	0.99	0.98
$(\frac{2}{\sqrt{n}}, 0.9)$	0.99	0.98	1.00	0.99	1.00	0.99
$(-\frac{3}{\sqrt{n}}, 0.9)$	1.00	1.00	1.00	1.00	1.00	1.00
$(\frac{3}{\sqrt{n}}, 0.9)$	1.00	1.00	1.00	1.00	1.00	1.00
$(-\frac{4}{\sqrt{n}}, 0.9)$	1.00	1.00	1.00	1.00	1.00	1.00
$(\frac{4}{\sqrt{n}}, 0.9)$	1.00	1.00	1.00	1.00	1.00	1.00
$(0, 1 - \frac{2}{n})$	0.10	0.05	0.11	0.06	0.10	0.05
$(-\frac{2}{n}, 1 - \frac{2}{n})$	0.24	0.16	0.23	0.16	0.24	0.16
$(\frac{2}{n}, 1 - \frac{2}{n})$	0.18	0.11	0.18	0.10	0.18	0.11
$(-\frac{4}{n}, 1 - \frac{2}{n})$	0.46	0.37	0.45	0.35	0.44	0.35
$(\frac{4}{n}, 1 - \frac{2}{n})$	0.47	0.34	0.45	0.32	0.44	0.31
$(-\frac{6}{n}, 1 - \frac{2}{n})$	0.63	0.55	0.63	0.54	0.62	0.54
$(\frac{6}{n}, 1 - \frac{2}{n})$	0.73	0.62	0.72	0.60	0.72	0.61
$(-\frac{8}{n}, 1 - \frac{2}{n})$	0.76	0.70	0.76	0.69	0.76	0.69
$(\frac{8}{n}, 1 - \frac{2}{n})$	0.88	0.81	0.89	0.82	0.88	0.81
$(0, 1)$	0.10	0.05	0.11	0.05	0.10	0.05
$(-\frac{2}{n}, 1)$	0.36	0.28	0.36	0.28	0.36	0.28
$(\frac{2}{n}, 1)$	0.33	0.22	0.32	0.21	0.32	0.21
$(-\frac{4}{n}, 1)$	0.61	0.53	0.60	0.53	0.60	0.53
$(\frac{4}{n}, 1)$	0.71	0.60	0.70	0.60	0.69	0.59
$(-\frac{6}{n}, 1)$	0.76	0.70	0.76	0.70	0.76	0.70
$(\frac{6}{n}, 1)$	0.89	0.83	0.88	0.82	0.88	0.82
$(-\frac{8}{n}, 1)$	0.86	0.82	0.86	0.81	0.85	0.81
$(\frac{8}{n}, 1)$	0.96	0.93	0.96	0.93	0.96	0.93

where ξ_t lies between X_{t-2} and X_{t-1}. By (11), we have $|X_{t-2}|/t^a \xrightarrow{p} \infty$, $|X_{t-1}|/t^a \xrightarrow{p} \infty$ and $|X_{t-2} - X_{t-1}|/t^a \xrightarrow{p} 0$ for any $a \in (0, 1/2)$ as $t \to \infty$, which imply that

$$|\xi_t|/t^a \xrightarrow{p} \infty \quad \text{for any} \quad a \in (0, 1/2) \quad \text{as} \quad t \to \infty. \tag{15}$$

By (11), (13)–(15), we have

$$\frac{1}{\sqrt{n}} \sum_{t=3}^{n} Z_{t2}(\beta_{1,0}, \beta_{2,0}) = \{\frac{1}{\sqrt{n}} \sum_{t=3}^{n} U_t\} \frac{X_{n-2}}{\sqrt{1 + X_{n-2}^2}} + o_p(1). \tag{16}$$

Hence, the lemma follows from (12) and (16). □

Lemma 2. *Under conditions of Theorem 1 and (10), we have that as $n \to \infty$*

$$\frac{1}{n} \sum_{t=3}^{n} Z_t(\beta_{1,0}, \beta_{2,0}) Z_t^T(\beta_{1,0}, \beta_{2,0}) \xrightarrow{p} \Sigma = \begin{pmatrix} E(U_1^2)E(\sum_{j=0}^{\infty} \psi_j V_{-j})^2 & 0 \\ 0 & E(U_1^2) \end{pmatrix},$$

where $\xrightarrow{p}$ denotes the convergence in probability.

Table 7
Empirical sizes and powers for the test based on Theorem 2 and model (6) with $\theta = 0$ and unknown α for testing $H_0 : \beta_2 = 0$ vs $H_a : \beta_2 \neq 0$. We take $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$ in $\sum_{j=0}^{\infty} \psi_j V_{t-j}$.

(β_2, ϕ)	$n = 200$		$n = 500$		$n = 2000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.11	0.06	0.10	0.05	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.43	0.32	0.43	0.33	0.43	0.31
$(\frac{1}{\sqrt{n}}, 0.9)$	0.43	0.30	0.44	0.31	0.43	0.31
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.81	0.74	0.84	0.77	0.88	0.81
$(\frac{2}{\sqrt{n}}, 0.9)$	0.89	0.81	0.92	0.85	0.92	0.86
$(-\frac{3}{\sqrt{n}}, 0.9)$	0.95	0.92	0.98	0.96	0.99	0.98
$(\frac{3}{\sqrt{n}}, 0.9)$	0.99	0.98	1.00	0.99	1.00	1.00
$(-\frac{4}{\sqrt{n}}, 0.9)$	0.99	0.98	1.00	1.00	1.00	1.00
$(\frac{4}{\sqrt{n}}, 0.9)$	1.00	1.00	1.00	1.00	1.00	1.00
$(0, 1 - \frac{2}{n})$	0.10	0.05	0.09	0.05	0.09	0.05
$(-\frac{2}{n}, 1 - \frac{2}{n})$	0.14	0.07	0.13	0.07	0.13	0.07
$(\frac{2}{n}, 1 - \frac{2}{n})$	0.15	0.08	0.15	0.08	0.13	0.07
$(-\frac{4}{n}, 1 - \frac{2}{n})$	0.25	0.17	0.24	0.15	0.24	0.16
$(\frac{4}{n}, 1 - \frac{2}{n})$	0.29	0.18	0.28	0.17	0.27	0.17
$(-\frac{6}{n}, 1 - \frac{2}{n})$	0.38	0.29	0.38	0.29	0.36	0.27
$(\frac{6}{n}, 1 - \frac{2}{n})$	0.45	0.33	0.45	0.33	0.45	0.32
$(-\frac{8}{n}, 1 - \frac{2}{n})$	0.51	0.42	0.50	0.42	0.50	0.41
$(\frac{8}{n}, 1 - \frac{2}{n})$	0.63	0.51	0.63	0.51	0.61	0.49
$(0, 1)$	0.11	0.06	0.10	0.05	0.11	0.06
$(-\frac{2}{n}, 1)$	0.15	0.08	0.15	0.08	0.14	0.08
$(\frac{2}{n}, 1)$	0.23	0.14	0.22	0.14	0.22	0.13
$(-\frac{4}{n}, 1)$	0.30	0.22	0.30	0.21	0.30	0.21
$(\frac{4}{n}, 1)$	0.45	0.33	0.44	0.32	0.44	0.32
$(-\frac{6}{n}, 1)$	0.46	0.37	0.47	0.38	0.46	0.37
$(\frac{6}{n}, 1)$	0.64	0.53	0.65	0.54	0.64	0.53
$(-\frac{8}{n}, 1)$	0.60	0.52	0.59	0.52	0.60	0.50
$(\frac{8}{n}, 1)$	0.78	0.69	0.79	0.70	0.78	0.69

Proof. Let $D_t = U_t^2(X_{t-1} - X_{t-2})$. Using (11), (14) and (15), we can show that

$$\frac{1}{n}\sum_{t=3}^{n} Z_{t1}^2(\beta_{1,0}, \beta_{2,0}) = \frac{1}{n}\sum_{t=3}^{n} U_t^2(\sum_{j=0}^{\infty} \psi_j V_{t-1-j} - \frac{\rho}{n} X_{t-2})^2$$

$$= \frac{1}{n}\sum_{t=3}^{n} U_t^2(\sum_{j=0}^{\infty} \psi_j V_{t-1-j})^2 + o_p(1) = E(U_1^2)E(\sum_{j=0}^{\infty} \psi_j V_{-j})^2 + o_p(1),$$

and

$$\frac{1}{n}\sum_{t=3}^{n} Z_{t2}^2(\beta_{1,0}, \beta_{2,0}) = \frac{1}{n}\sum_{t=3}^{n} U_t^2 \frac{X_{t-2}^2}{1 + X_{t-2}^2} = \frac{1}{n}\sum_{t=3}^{n} U_t^2 + o_p(1) = E(U_1^2) + o_p(1),$$

as well as

$$\frac{1}{n}\sum_{t=3}^{n} Z_{t1}(\beta_{1,0}, \beta_{2,0}) Z_{t2}(\beta_{1,0}, \beta_{2,0})$$

$$= \frac{1}{n}\sum_{t=3}^{n} U_t^2(X_{t-1} - X_{t-2}) \frac{X_{t-2}}{\sqrt{1 + X_{t-2}^2}}$$

Table 8
Empirical sizes and powers for the test based on Theorem 2 and model (6) with $\theta = 0$ and unknown α for testing $H_0 : \beta_2 = 0$ vs $H_a : \beta_2 \neq 0$. We take $\psi_0 = \cdots = \psi_4 = 1$ and $\psi_j = 0$ for $j \geq 5$ in $\sum_{j=0}^{\infty} \psi_j V_{t-j}$.

(β_2, ϕ)	$n = 200$		$n = 500$		$n = 2000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.11	0.06	0.10	0.05	0.10	0.05
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.42	0.33	0.44	0.33	0.44	0.32
$(\frac{1}{\sqrt{n}}, 0.9)$	0.43	0.30	0.43	0.31	0.43	0.31
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.81	0.73	0.84	0.77	0.88	0.81
$(\frac{2}{\sqrt{n}}, 0.9)$	0.89	0.81	0.92	0.85	0.92	0.85
$(-\frac{3}{\sqrt{n}}, 0.9)$	0.95	0.92	0.98	0.96	0.99	0.98
$(\frac{3}{\sqrt{n}}, 0.9)$	0.99	0.98	1.00	0.99	1.00	1.00
$(-\frac{4}{\sqrt{n}}, 0.9)$	0.99	0.98	1.00	1.00	1.00	1.00
$(\frac{4}{\sqrt{n}}, 0.9)$	1.00	1.00	1.00	1.00	1.00	1.00
$(0, 1-\frac{2}{n})$	0.10	0.05	0.10	0.05	0.09	0.05
$(-\frac{2}{n}, 1-\frac{2}{n})$	0.14	0.08	0.13	0.07	0.14	0.08
$(\frac{2}{n}, 1-\frac{2}{n})$	0.15	0.09	0.14	0.08	0.14	0.07
$(-\frac{4}{n}, 1-\frac{2}{n})$	0.24	0.16	0.23	0.15	0.25	0.16
$(\frac{4}{n}, 1-\frac{2}{n})$	0.28	0.18	0.27	0.17	0.26	0.17
$(-\frac{6}{n}, 1-\frac{2}{n})$	0.38	0.29	0.38	0.28	0.37	0.28
$(\frac{6}{n}, 1-\frac{2}{n})$	0.46	0.34	0.46	0.33	0.45	0.32
$(-\frac{8}{n}, 1-\frac{2}{n})$	0.51	0.42	0.51	0.42	0.50	0.41
$(\frac{8}{n}, 1-\frac{2}{n})$	0.63	0.51	0.63	0.51	0.61	0.49
$(0, 1)$	0.11	0.06	0.10	0.05	0.11	0.06
$(-\frac{2}{n}, 1)$	0.15	0.08	0.15	0.08	0.15	0.08
$(\frac{2}{n}, 1)$	0.23	0.15	0.22	0.14	0.22	0.14
$(-\frac{4}{n}, 1)$	0.30	0.22	0.29	0.21	0.30	0.21
$(\frac{4}{n}, 1)$	0.45	0.33	0.44	0.33	0.43	0.31
$(-\frac{6}{n}, 1)$	0.46	0.38	0.46	0.38	0.46	0.38
$(\frac{6}{n}, 1)$	0.63	0.53	0.64	0.53	0.64	0.53
$(-\frac{8}{n}, 1)$	0.60	0.52	0.60	0.52	0.60	0.52
$(\frac{8}{n}, 1)$	0.78	0.69	0.79	0.70	0.78	0.69

Table 9
P-values for testing $H_0 : \beta_1 = 0, H_0 : \beta_2 = 0$ and $H_0 : \beta_1 = \beta_2 = 0$ for model (4) with known $\alpha_0 = 0$ for the period 1/1952–12/2012.

Regressor	$H_0 : \beta_1 = 0$	$H_0 : \beta_2 = 0$	$H_0 : \beta_1 = \beta_2 = 0$
Dividend payout ratio	0.1316	0.0049	0.0097
Long-term yield	0.0155	0.0085	0.0013
Dividend yield	0.5236	0.0007	0.0032
Dividend price ratio	0.2418	0.0010	0.0002
T-bill rate	0.0090	0.0608	0.0053
Earnings price ratio	0.6748	0.0008	0.0035
Book-to-Market value ratio	0.2458	0.0006	0.0002
Default yield spread	0.6182	0.0032	0.0129
Net equity expansion	0.0956	0.0477	0.0301
Term spread	0.1459	0.0007	0.0026

$$= \frac{1}{n}\sum_{t=3}^{n}\{\sum_{j=2}^{t} D_j - \sum_{j=2}^{t-1} D_j\}\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}}$$

$$= \frac{1}{n}\sum_{j=2}^{n} D_j \frac{X_{n-2}}{\sqrt{1+X_{n-2}^2}} + \frac{1}{n}\sum_{t=3}^{n-1}\{\sum_{j=2}^{t} D_j\}\frac{X_{t-2}}{\sqrt{1+X_{j-2}^2}}$$

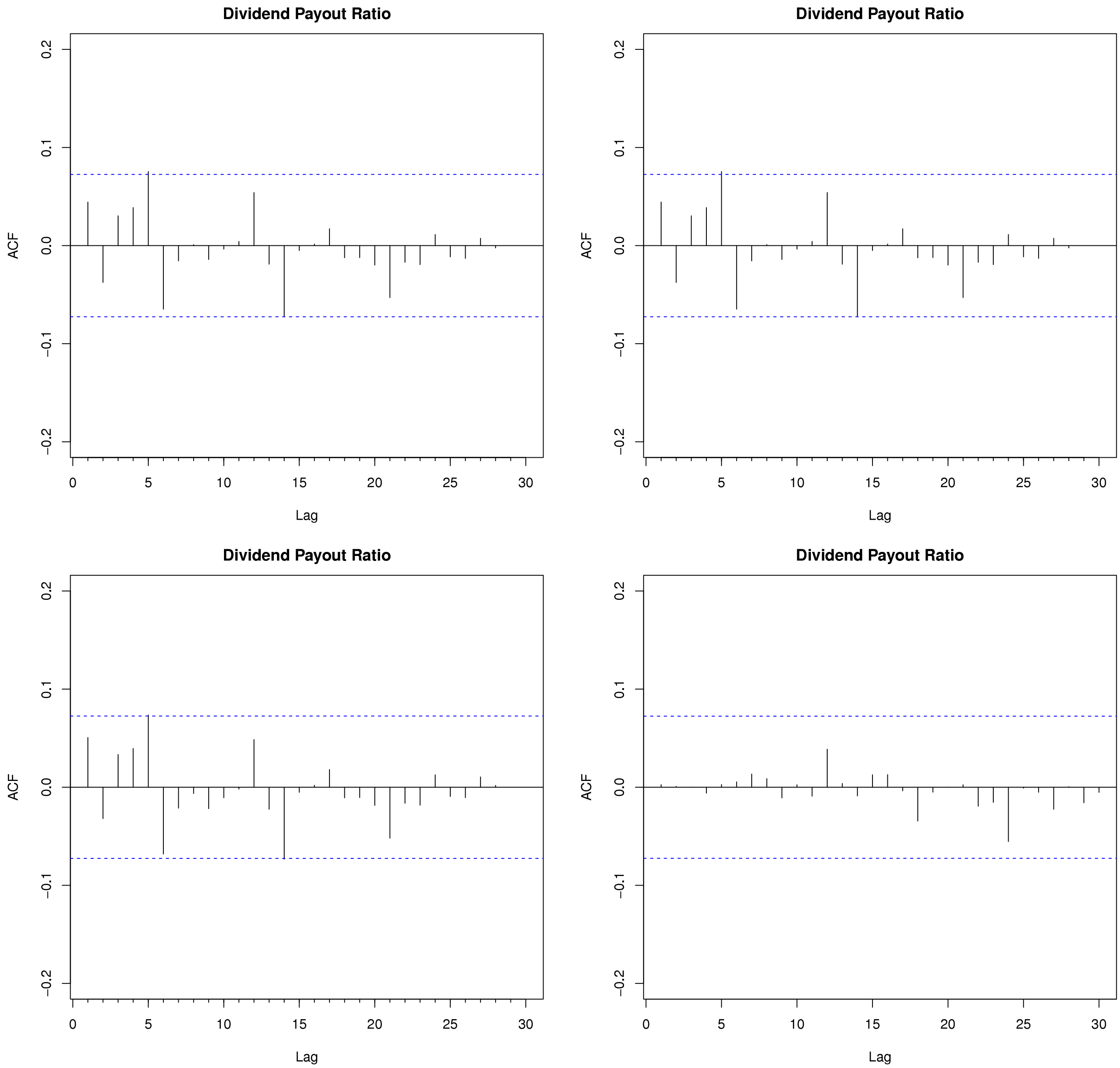

Fig. 1. (a) Upper left panel: autocorrelation function of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}_1(X_{t-1} - X_{t-2}) - \hat{\beta}_2 X_{t-2}$ for the period 1/1927–12/2012, where $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)^T$ is the least squares estimate. (b) Upper right panel: autocorrelation function of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}_1(X_{t-1} - X_{t-2}) - \hat{\beta}_2 X_{t-2}$ for the period 1/1952–12/2012, where $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)^T$ is the least squares estimate. (c) Lower left panel: autocorrelation function of $\hat{U}_t = Y_t - \hat{\beta}_1(X_{t-1} - X_{t-2}) - \hat{\beta}_2 X_{t-2}$ for the period 1/1952–12/2012, where $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)^T$ is the least squares estimate. (d) Lower right panel: autocorrelation function of the estimated $V_t's$ for the period 1/1952–12/2012 based on the ARMA(1,15) model, i.e., $X_t = \theta + \phi X_{t-1} + V_t - \sum_{j=1}^{15} \psi_j V_{t-j}$.

$$
\begin{aligned}
& -\frac{1}{n}\sum_{t=3}^{n}\{\sum_{j=2}^{t-1} D_j\}\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}} \\
= \quad & \{\frac{1}{n}\sum_{t=2}^{n} U_t^2(\sum_{j=0}^{\infty}\psi_j V_{t-1-j} - \frac{\rho}{n}X_{t-2})\}\frac{X_{n-2}}{\sqrt{1+X_{t-2}^2}} \\
& + \frac{1}{n}\sum_{t=3}^{n-1}\{\sum_{j=2}^{t} D_j\}\{\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}} - \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}\} + o_p(1)
\end{aligned}
$$

Table 10
P-values for testing $H_0 : \beta_1 = 0, H_0 : \beta_2 = 0$ and $H_0 : \beta_1 = \beta_2 = 0$ for model (6) with unknown α for the period 1/1952–12/2012.

Regressor	$H_0 : \beta_1 = 0$	$H_0 : \beta_2 = 0$	$H_0 : \beta_1 = \beta_2 = 0$
Dividend payout ratio	0.2921	0.6481	0.5674
Long-term yield	0.0866	0.2758	0.1623
Dividend yield	0.2631	0.8195	0.5064
Dividend price ratio	0.6891	0.8703	0.8965
T-bill rate	0.1292	0.1328	0.1268
Earnings price ratio	0.5483	0.9997	0.8293
Book-to-Market value ratio	0.6159	0.8268	0.8694
Default yield spread	0.6675	0.6552	0.8641
Net equity expansion	0.1182	0.5829	0.2559
Term spread	0.7083	0.1631	0.3477

Table 11
P-values for testing $H_0 : \beta_1 = 0, H_0 : \beta_2 = 0$ and $H_0 : \beta_1 = \beta_2 = 0$ for model (4) with known $\alpha_0 = 0$ for the period 1/1952–12/2002.

Regressor	$H_0 : \beta_1 = 0$	$H_0 : \beta_2 = 0$	$H_0 : \beta_1 = \beta_2 = 0$
Dividend payout ratio	0.7029	0.0039	0.0145
Long-term yield	0.0184	0.0148	0.0026
Dividend yield	0.5365	0.0019	0.0079
Dividend price ratio	0.7253	0.0025	0.0036
T-bill rate	0.0051	0.0682	0.0034
Earnings price ratio	0.6481	0.0025	0.0036
Book-to-Market value ratio	0.5214	0.0016	0.0014
Default yield spread	0.0235	0.0021	0.0015
Net equity expansion	0.0937	0.1459	0.0889
Term spread	0.0488	0.0016	0.0030

Table 12
P-values for testing $H_0 : \beta_1 = 0, H_0 : \beta_2 = 0$ and $H_0 : \beta_1 = \beta_2 = 0$ for model (6) with unknown α for the period 1/1952–12/2002.

Regressor	$H_0 : \beta_1 = 0$	$H_0 : \beta_2 = 0$	$H_0 : \beta_1 = \beta_2 = 0$
Dividend payout ratio	0.7281	0.6122	0.8451
Long-term yield	0.0297	0.4013	0.0804
Dividend yield	0.7425	0.6885	0.8778
Dividend price ratio	0.9720	0.6550	0.8853
T-bill rate	0.0046	0.0683	0.0074
Earnings price ratio	0.8708	0.8010	0.9389
Book-to-Market value ratio	0.5021	0.7541	0.7957
Default yield spread	0.1966	0.7089	0.4129
Net equity expansion	0.0223	0.1918	0.0241
Term spread	0.0751	0.0025	0.0059

$$
\begin{aligned}
&= \{\frac{1}{n}\sum_{t=2}^{n} U_t^2(\sum_{j=0}^{\infty}\psi_j V_{t-1-j}) + o_p(1)\}\frac{X_{n-2}}{\sqrt{1+X_{n-2}^2}} \\
&\quad + \frac{1}{n}\sum_{t=3}^{n-1}\{\sum_{j=2}^{t} D_j\}(1+\xi_t^2)^{-3/2}(X_{t-2}-X_{t-1}) + o_p(1) \\
&= o_p(1),
\end{aligned}
$$

where ξ_t lies between X_{t-2} and X_{t-1}. Hence, the lemma follows from the above equations. □

Lemma 3. *Under conditions of Theorem 1 and (10), $L(\beta_1, \beta_{2,0})$ attains its maximum value with probability tending to one at some point $\bar{\beta}_1$ such that $|\bar{\beta}_1 - \beta_{1,0}| < n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2+\delta)$ as $n \to \infty$, and $\bar{\beta}_1$ and $\bar{\lambda}$ satisfy $Q_{1n}(\bar{\beta}_1, \bar{\lambda}) = 0$ and $Q_{2n}(\bar{\beta}_1, \bar{\lambda}) = 0$, where*

$$
Q_{1n}(\beta_1, \lambda) := \frac{1}{n}\sum_{t=3}^{n}\frac{Z_t(\beta_1, \beta_{2,0})}{1+\lambda^T Z_t(\beta_1, \beta_{2,0})},
$$

$$Q_{2n}(\beta_1,\lambda)=\frac{1}{n}\sum_{t=3}^{n}\frac{1}{1+\lambda^T Z_t(\beta_1,\beta_{2,0})}\left(\frac{\partial Z_t(\beta_1,\beta_{2,0})}{\partial\beta_1}\right)^T\lambda.$$

Proof. Like the proof of Lemma 1 of Qin and Lawless (1994), the lemma follows from Lemmas 1 and 2. □

Lemma 4. *Under conditions of Theorem 1 and (10), $L(\beta_{1,0},\beta_2)$ attains its maximum value with probability tending to one at some point $\bar\beta_2$ such that $|\sqrt{n}\bar\beta_2-\sqrt{n}\beta_{2,0}|<n^{-1/\delta_0}$ for some $\delta_0\in(2,2+\delta)$ as $n\to\infty$, and $\bar\beta_2$ and $\bar\lambda$ satisfy $Q^*_{1n}(\bar\beta_2,\bar\lambda)=0$ and $Q^*_{2n}(\bar\beta_2,\bar\lambda)=0$, where*

$$Q^*_{1n}(\beta_2,\lambda):=\frac{1}{n}\sum_{t=3}^{n}\frac{Z_t(\beta_{1,0},\beta_2)}{1+\lambda^T Z_t(\beta_{1,0},\beta_2)},$$
$$Q^*_{2n}(\beta_2,\lambda)=\frac{1}{n}\sum_{t=3}^{n}\frac{1}{1+\lambda^T Z_t(\beta_{1,0},\beta_2)}\left(\frac{\partial Z_t(\beta_{1,0},\beta_2)}{\partial(\sqrt{n}\beta_2)}\right)^T\lambda.$$

Proof. Since $X_t=O_p(\sqrt{n})$, we write $\beta_2X_{t-2}=(\sqrt{n}\beta_2)(X_{t-2}/\sqrt{n})$. Hence, following the proof of Lemma 1 of Qin and Lawless (1994), the lemma can be shown by using Lemmas 1 and 2. □

Proof of Theorem 1. We only prove the case of (10) since other cases can be proved in a similar way. First, it follows easily from (11) and Condition C1 that

$$\max_{3\le t\le n}\|Z_t(\beta_{1,0},\beta_{2,0})\|=o_p(n^{1/2}). \tag{17}$$

Hence, using Lemmas 1 and 2, (17) and the same arguments in the proof of Theorem 1 in Owen (1990), we can show that $-2\log L(\beta_{1,0},\beta_{2,0})\xrightarrow{d}\chi^2(2)$ as $n\to\infty$. Let $e_t=\sum_{j=0}^{\infty}\psi_jV_{t-j}$. By (11), (14) and (15), we have

$$\frac{1}{n}\sum_{t=3}^{n}(X_{t-1}-X_{t-2})^2=\frac{1}{n}\sum_{t=3}^{n}e_{t-1}^2+o_p(1)=E(\sum_{j=0}^{\infty}\psi_jV_{-j})^2+o_p(1),$$

and

$$\begin{aligned}
&\frac{1}{n}\sum_{t=3}^{n}(X_{t-1}-X_{t-2})\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}}=\frac{1}{n}\sum_{t=3}^{n}e_{t-1}\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}}+o_p(1)\\
=\;&\frac{1}{n}\sum_{t=2}^{n-1}e_t\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}+o_p(1)=\frac{1}{n}\sum_{t=2}^{n-1}\{\sum_{j=1}^{t}e_j-\sum_{j=1}^{t-1}e_j\}\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}+o_p(1)\\
=\;&\frac{1}{n}\sum_{j=1}^{n-1}e_j\frac{X_{n-2}}{\sqrt{1+X_{n-2}^2}}+\frac{1}{n}\sum_{t=2}^{n-1}\{\sum_{j=1}^{t}e_j\}\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}-\frac{X_t}{\sqrt{1+X_t^2}}\}+o_p(1)\\
=\;&o_p(1)+\frac{1}{n}\sum_{t=2}^{n-1}\{\sum_{j=1}^{t}e_j\}(1+\xi_t^2)^{-3/2}(X_{t-1}-X_t)+o_p(1)\\
=\;&o_p(1),
\end{aligned}$$

where ξ_t lies between X_{t-1} and X_t. Hence, we have

$$\begin{aligned}
\frac{\partial Q_{1n}(\beta_{1,0},0)}{\partial\beta_1}&=\frac{1}{n}\sum_{t=3}^{n}\begin{pmatrix}-(X_{t-1}-X_{t-2})^2\\-(X_{t-1}-X_{t-2})\frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}}\end{pmatrix}\\
&=\begin{pmatrix}-E(\sum_{j=0}^{\infty}\psi_jV_{-j})^2\\0\end{pmatrix}+o_p(1)\\
&=S_1+o_p(1).
\end{aligned}\tag{18}$$

Furthermore, we have

$$\frac{\partial Q_{1n}(\beta_{1,0},0)}{\partial\lambda}=\Sigma+o_p(1),\quad\frac{\partial Q_{2n}(\beta_{1,0},0)}{\partial\beta_1}=0,\quad\frac{\partial Q_{2n}(\beta_{1,0},0)}{\partial\lambda^T}=\frac{\partial Q_{1n}(\beta_{1,0},0)}{\partial\beta_1},\tag{19}$$

where Σ is given in Lemma 2. Using Lemma 3 and expanding $Q_{1n}(\bar\beta_1, \bar\lambda)$ and $Q_{2n}(\bar\beta_1, \bar\lambda)$ around $(\beta_{1,0}, 0)^T$, we have

$$\begin{aligned} 0 &= Q_{1n}(\beta_{1,0}, 0) + \frac{\partial Q_{1n}(\beta_{1,0}, 0)}{\partial \beta_1}(\bar\beta_1 - \beta_{1,0}) + \frac{\partial Q_{1n}(\beta_{1,0}, 0)}{\partial \lambda}\bar\lambda + o_p(|\bar\beta_1 - \beta_{1,0}| + \|\bar\lambda\|) \\ &= Q_{1n}(\beta_{1,0}, 0) + S_1(\bar\beta_1 - \beta_{1,0}) - \Sigma\bar\lambda + o_p(|\bar\beta_1 - \beta_{1,0}| + \|\bar\lambda\|) \end{aligned} \tag{20}$$

and

$$\begin{aligned} 0 &= Q_{2n}(\beta_{1,0}, 0) + \frac{\partial Q_{2n}(\beta_{1,0}, 0)}{\partial \beta_1}(\bar\beta_1 - \beta_{1,0}) + \frac{\partial Q_{2n}(\beta_{1,0}, 0)}{\partial \lambda}\bar\lambda + o_p(|\bar\beta_1 - \beta_{1,0}| + \|\bar\lambda\|) \\ &= S_1^T\lambda + o_p(|\bar\beta_1 - \beta_{1,0}| + \|\bar\lambda\|). \end{aligned} \tag{21}$$

By (20) and (21), we have

$$S_1^T\Sigma^{-1}S_1\sqrt{n}(\bar\beta_1 - \beta_{1,0}) = -S_1^T\Sigma^{-1}\sqrt{n}Q_{1n}(\beta_{1,0}, 0) + o_p(1) \tag{22}$$

and

$$\sqrt{n}\bar\lambda = \Sigma^{-1}\sqrt{n}Q_{1n}(\beta_{1,0}, 0) + \Sigma^{-1}S_1\sqrt{n}(\bar\beta_1 - \beta_{1,0}) + o_p(1), \tag{23}$$

which imply that

$$\sqrt{n}\bar\lambda = \{\Sigma^{-1} - \Sigma^{-1}S_1S^{-1}S_1^T\Sigma^{-1}\}\sqrt{n}Q_{1n}(\beta_{1,0}, 0) + o_p(1), \tag{24}$$

where $S = S_1^T\Sigma^{-1}S_1$. Since $\Sigma = \mathrm{diag}\{\sigma_1^2, \sigma_2^2\}$ and $S_1^T = (-\sigma_1^2, 0)$, we have $S = \sigma_1^2$ and

$$M = \Sigma^{-1} - \Sigma^{-1}S_1S^{-1}S_1^T\Sigma^{-1} = \mathrm{diag}\{0, \sigma_2^{-2}\}. \tag{25}$$

It follows from Lemmas 1, 3, (23)–(25) and Taylor expansion that

$$\begin{aligned} & -2\log L^{P2}(\beta_{2,0}) = -2\log L(\bar\beta_1, \beta_{2,0}) \\ = \ & 2\sum_{t=3}^{n}\bar\lambda^T Z_t(\bar\beta_1, \beta_{2,0}) - \sum_{t=3}^{n}\bar\lambda^T Z_t(\bar\beta_1, \beta_{2,0})Z_t^T(\bar\beta_1, \beta_{2,0})\bar\lambda + o_p(1) \\ = \ & 2n\bar\lambda^T Q_{1n}(\bar\beta_1, 0) - n\bar\lambda^T\Sigma\lambda + o_p(1) \\ = \ & 2n\bar\lambda^T Q_{1n}(\beta_{1,0}, 0) + 2n\bar\lambda^T\frac{\partial Q_{1n}(\beta_{1,0}, 0)}{\partial\beta_1}(\bar\beta_1 - \beta_{1,0}) - n\bar\lambda^T\Sigma\bar\lambda + o_p(1) \\ = \ & 2n\bar\lambda^T\Sigma\bar\lambda - n\bar\lambda^T\Sigma\bar\lambda + o_p(1) \\ = \ & \{\sqrt{n}Q_{1n}(\beta_{1,0}, 0)\}^T M\Sigma M\{\sqrt{n}Q_{1n}(\beta_{1,0}, 0)\} + o_p(1) \\ = \ & \{\frac{1}{\sqrt{n}}\sum_{t=3}^{n} Z_{t2}(\beta_{1,0}, \beta_{2,0})\}^2/\sigma_2^2 + o_p(1) \\ \xrightarrow{d} \ & \chi^2(1) \end{aligned}$$

as $n \to \infty$. Similarly, we can show that $-2\log L^{P1}(\beta_{1,0}) \xrightarrow{d} \chi^2(1)$ as $n \to \infty$ by using Lemma 4 instead of Lemma 3. Hence the theorem follows. □

Proof of Theorem 2. It can be shown in the same way as in Theorem 1. □

Proof of Theorem 3. Note that a key fact in proving Theorems 1 and 2 is that $|X_t| \xrightarrow{p} \infty$ in case of nonstationary, which still holds under conditions of Theorem 3. Also, due to the independence of U_t and U_{t+m}, it is clear that

$$\frac{1}{m}\sum_{t=3}^{m}\bar Z_{t,2}(\beta_1^{(0)}, \beta_2^{(0)})\bar Z_{t,4}(\beta_1^{(0)}, \beta_2^{(0)}) \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty.$$

Therefore Theorem 3 can be shown by following the arguments in proving Theorem 1. □

Acknowledgments

We thank the editor, Professor Oliver Linton, an associate editor and three reviewers for their constructive and helpful comments and suggestions that improved the paper significantly. Liu's research was supported by the National Natural Science Foundation of China (11601197, 11461029), China Postdoctoral Science Foundation funded project (2016M600511, 2017T100475), NSF of Jiangxi Province (2018ACB21002, 20171ACB21030). Yang's research was partly supported by NSF of Jiangxi Province (20171BAA208002) and the science and technology project of Jiangxi Province Education Department (GJJ170329). Cai's research was, in part, supported by the National Natural Science Foundation of China (71131008, 71631004). Peng's research was partly supported by Simons Foundation.

References

Amihud, Y., Hurvich, C.M., 2004. Predictive regressions: a reduced-bias estimation method. J. Financ. Quant. Anal. 39, 813–841.
Amihud, Y., Hurvich, C.M., Wang, Y., 2009. Multiple-predictor regressions: hypothesis testing. Rev. Financ. Stud. 22, 413–434.
Billingsley, P., 1999. Convergence of Probability Measures, second ed. Wiley, New York.
Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, second ed. Springer, New York.
Cai, Z., Wang, Y., 2014. Testing predictive regression models with nonstationary regressors. J. Econometrics 178, 4–14.
Cai, Z., Wang, Y., Wang, Y., 2015. Testing instability in a predictive regression model with nonstationary regressors. Econom. Theory 31, 953–980.
Campbell, J.Y., Yogo, M., 2006. Efficient tests of stock return predictability. J. Financ. Econ. 81, 27–60.
Cavanagh, C.L., Elliott, G., Stock, J.H., 1995. Inference in models with nearly integrated regressors. Econom. Theory 11, 1131–1147.
Chen, W.W., Deo, R.S., 2009. Bias reduction and likelihood-based almost exactly sized hypothesis testing in predictive regressions using the restricted likelihood. Econom. Theory 25, 1143–1179.
Chen, W.W., Deo, R.S., Yi, Y., 2013. Uniform inference in predictive regression models. J. Bus. Econom. Statist. 31, 525–533.
Choi, Y., Jacewitz, S., Park, J.Y., 2016. A reexamination of stock return predictability. J. Econometrics 192, 168–189.
Frees, E.W., Derrig, R.A., Meyers, G., 2014. Predictive Modeling Applications in Actuarial Science. Cambridge University Press.
Hall, P., Heyde, C., 1980. Martingale Limit Theory and Its Applications. Academic Press, New York.
Jaworski, P., Durante, F., Härdle, W.K. (Eds.), 2013. Copulae in Mathematical and Quantitative Finance. In: Lecture Notes in Statistics-Proceedings, Springer-Verlag Berlin Heidelberg.
Jaworski, P., Durante, F., Härdle, W.K., Rychlik, T. (Eds.), 2010. Copula Theory and Its Application. In: Lecture Notes in Statistics-Proceedings, Springer-Verlag Berlin Heidelberg.
Joe, H., 1997. Dependence Modeling with Copulas. Chapman & Hall/CRC.
Juhl, T., 2014. A nonparametric test for the predictive regression model. J. Bus. Econom. Statist. 32, 387–394.
Kasparis, I., Andreou, E., Phillips, P.C.B., 2015. Nonparametric predictive regression. J. Econometrics 185, 468–494.
Kostakis, A., Magdalinos, T., Stamatogiannis, M.P., 2015. Robust econometric inference for stock return predictability. Rev. Financ. Stud. 28, 1506–1553.
Li, C., Li, D., Peng, L., 2017. Uniform test for predictive regression with AR errors. J. Bus. Econom. Statist. 35, 29–39.
Owen, A., 1990. Empirical likelihood ratio confidence regions. Ann. Statist. 18, 90–120.
Owen, A., 2001. EMpirical Likelihood. Chapman and Hall, New York.
Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74, 535–547.
Phillips, P.C.M., Lee, J.H., 2013. Predictive regression under various degrees of persistence and robust long-horizon regression. J. Econometrics 177, 250–264.
Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating functions. Ann. Statist. 22, 300–325.
Stambaugh, R.F., 1999. Predictive regressions. J. Financ. Econ. 54, 375–421.
Zhu, F., Cai, Z., Peng, L., 2014. Predictive regressions for macroeconomic data. Ann. Appl. Stat. 8, 577–594.