

Unified Tests for a Dynamic Predictive Regression

BINGDUO YANG¹, XIAOHUI LIU², LIANG PENG³ AND ZONGWU CAI⁴

September 25, 2018

Abstract: Testing for predictability of asset returns has been a long history in economics and finance. Recently, based on a simple predictive regression, Kostakis, Magdalinos and Stamatogiannis (2015, *Review of Financial Studies*) derived a Wald type test based on the context of the extended instrumental variable (IVX) methodology for testing predictability of stock returns and Demetrescu (2014) showed that the local power of the standard IVX-based test could be improved in some cases when a lagged predicted variable is added to the predictive regression on purpose, which poses a general important question on whether a lagged predicted variable should be included in the model or not. This paper proposes novel robust procedures for testing both the existence of a lagged predicted variable and the predictability of asset returns in a predictive regression regardless of regressors being stationary or nearly integrated or unit root and the AR model for regressors with or without intercept. A simulation study confirms the good finite sample performance of the proposed tests before applying the proposed tests to some real datasets in finance to illustrate their practical usefulness.

Key words and phrases: Autoregressive errors, empirical likelihood, predictive regression, weighted score.

JEL classifications: C12, C22.

¹School of Finance, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China. E-mail: bdyang2006@sina.com.

²School of Statistics, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China. E-mail: csuliuah912@gmail.com.

³Department of Risk Management and Insurance, Georgia State University, Atlanta, GA 30303, USA. E-mail: lpeng@gsu.edu.

⁴Department of Economics, University of Kansas, Lawrence, KS 66045, USA. E-mail: caiz@ku.edu, and The Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian, 316005, China.

1 Introduction

The introduction to the 2013 Nobel for Economic Sciences states: *“There is no way to predict whether the prices of stocks and bonds will go up or down over the next few days or weeks. But it is quite possible to foresee the broad course of the prices of these assets over longer time periods, such as the next three to five years...”*. Testing for predictability of asset returns has been a long history and is of importance in economics and finance, and such a test is often built on a simple linear structural regression model between a predicted variable and some regressors; see, for example, the excellent survey papers by Campbell (2008) and Phillips (2015). Typically, some predicted variables employed in the literature are low frequency data, such as the annual, quarterly and monthly CRSP value-weighted index in Campbell and Yogo (2006); the monthly S&P 500 excess returns in Cai and Wang (2014) and in Kostakis et al. (2015). Some commonly employed regressors (financial predictors or predicting variables) are dividend payout ratio, long-term yield, dividend yield, dividend-price yield, T-bill rate, earnings-price ratio, book-to-market value ratio, default yield spread, net equity expansion and term spread; see Kostakis et al. (2015) for a detailed description of these variables.

Since empirical studies suggest that those regressors may be persistent, such as near unit root (nearly integrated) or unit root (integrated), classical tests for predictability built upon a linear regression model are no longer valid. For example, Campbell and Yogo (2006) and Demetrescu and Rodrigues (2016) pointed out that the usual asymptotic approximation of the t-test statistic by employing the (standard) normal distribution performs particularly bad when regressors are persistent with the largest autoregressive roots of the typical regressor candidate being usually smaller than one but close to one. Therefore, one may employ the nearly integrated asymptotics as an alternative framework for statistical inference. However, in the context of nearly integrated regressors, as addressed by Demetrescu and Rodrigues (2016), the limiting distribution of the slope parameter estimator is not centered at zero, and this bias depends on the mean reversion parameter of the nearly integrated regressor. Although nearly integrated asymptotics approximates the finite sample behavior of the t-statistic for no predictability considerably better when regressors are persistent, the exact degree of persistence of a given regressor, and thus the correct critical values for a predictability test, are unknown in practice. To overcome these difficulties, a number of alternative (robust) approaches have been proposed in the literature to test predictability without characterizing the stochastic properties of regressors (i.e., whether they are stationary or nearly integrated or unit root); see, for instance, Cavanagh et al. (1995),

Campbell and Yogo (2006), Jansson and Moreira (2006), Phillips and Lee (2013), Cai and Wang (2014), Breitung and Demetrescu (2015), Kostakis, et al. (2015), Demetrescu and Rodrigues (2016), and references therein.

Consider the following simple predictive regression model:

$$Y_t = \alpha + \beta X_{t-1} + U_t, \quad X_t = \theta + \phi X_{t-1} + V_t. \quad (1)$$

Due to the dependence between U_t and V_t , researchers have found that the least squares estimator for β based on the first equation in (1) is biased in finite samples when the regressor $\{X_t\}$ is nearly integrated (see Stambaugh, 1999), and some bias-corrected inferences have been proposed in the literature such as the linear projection method in Amihud and Hurvich (2004) and Chen and Deo (2009). A comprehensive summary of research for model (1) can be found in Phillips and Lee (2013). Using the linear projection of U_t onto V_t as $U_t = \rho_0 V_t + \eta_t$, Cai and Wang (2014) derived the asymptotic distribution of an estimator for β when X_t is nearly integrated, which depends on whether θ is zero or nonzero. Also, the asymptotic nonnormal distribution depends on the degree of persistence, which can not be estimated consistently, if X_t is nearly integrated. When $(U_t, V_t)^T$ has a bivariate normal distribution with A^T denoting the transpose of the matrix or vector of A throughout, Campbell and Yogo (2006) proposed a Bonferroni Q-test, based on the infeasible uniform most powerful test, and showed that this new test is more powerful than the Bonferroni t-test of Cavanagh et al. (1995) in the sense of Pitman efficiency. Implementing this Bonferroni Q-test is nontrivial at all as it requires additional estimators and tables in an unpublished technic report written by them. Under the normality assumption, Chen et al. (2013) proposed a weighted least squares approximated likelihood inference with a limit depending on whether regressors are stationary or nearly integrated or unit root. Without the normality assumption, Zhu et al. (2014) proposed a robust empirical likelihood inference for β with a chi-squared limit regardless of $\{X_t\}$ being stationary or nearly integrated or unit root, and Choi et al. (2016) proposed a unified test based on a so-called Cauchy estimation regardless of $\{X_t\}$ being nearly integrated or unit root. Without using the information on the persistence level of the predicting variable, the key idea in Zhu et al. (2014) is to employ the property that $|X_t| \xrightarrow{P} \infty$ as $t \rightarrow \infty$ when $\{X_t\}$ is either nearly integrated or unit root. Therefore, the unified method in Zhu et al. (2014) is robust with respect to the stochastic properties of the predicting variable.

On the other hand, it is known in the econometrics literature that an extended instrumental variable (dubbed as IVX) based inference is attractive in handling the dependence between U_t and V_t and avoiding a nonstandard asymptotic limit; see Phillips and Magdalinos (2007) for details. In particular, the IVX estimation approach proposed by Magdalinos and Phillips (2009) is becoming increasingly popular in predictive regressions because the relevant test statistic has the same limiting distribution in both stationary and nonstationary cases. The key idea behind this method is to construct instrumental variables (IV) by explicitly controlling the degree of the regressor's persistence in the case of near integration. As illustrated by Kostakis et al. (2015), the IVX methodology offers a good balance between size control and power loss. The power of the proposed test depends on some tuning parameters in constructing the IVX instruments with a sacrifice for the nonstationary case; see the parameters $C_z < 0$ and $\beta \in (0, 1)$ defined in (4) and the rates of convergence in Kostakis et al. (2015). Based on some Monte Carlo simulation studies, Kostakis et al. (2015) recommended taking $C_z = -I$ and $\beta \in (0.9, 0.95)$ ¹. As Kostakis et al. (2015) assumed zero intercept in modeling regressors, i.e., no θ in the second equation of (1), and it is known that the divergent rate of a nearly integrated regressor depends on whether a nonzero intercept exists in the AR model for the regressor, it remains open whether the IVX method could unify the cases of zero and nonzero intercept θ . In summary, the IVX method in Kostakis et al. (2015) has a difficulty in choosing tuning parameters, sacrifices the test power in the nonstationary case, and may not be able to unify the cases of zero and nonzero intercept.

To improve the local power of the IVX based tests, Demetrescu (2014) proposed adding the lagged predicted variable into the model so that the model becomes dynamic. Specifically, Demetrescu (2014) considered the following dynamic model with $\gamma = 0$ but the restriction is not imposed in estimating parameters (termed as variable addition approach):

$$Y_t = \alpha + \gamma Y_{t-1} + \beta X_{t-1} + U_t, \quad X_t = \theta + \phi X_{t-1} + V_t; \quad (2)$$

see Demetrescu (2014) and Breitung and Demetrescu (2015) for more details on this model and the variable addition approach. Hence, an interesting question is whether the lagged variables are really econometrically needed in real applications, i.e., how to test the existence of a lagged predicted variable in a predictive regression, which apparently has not been formally addressed in predictive regressions when regressors may be nearly integrated. This paper addresses this

¹Note that the β in Kostakis et al. (2015) is different from the β in models (1) above and (2) below.

issue by proposing novel robust procedures for testing the existence of the lagged predicted variables ($H_0 : \gamma = 0$) in a predictive regression in addition to testing predictability ($H_0 : \beta = 0$) regardless of regressors being stationary or nearly integrated or unit root and the AR model for the regressors with or without intercept.

Although many tests for predictability have been proposed in the literature, conclusions on predictability are unfortunately quite contradictory for different data sets, data periods and methods. For example, Kostakis et al. (2015) reported significant predictability with respect to dividend yield, dividend-price ratio, T-bill rate, earnings-price ratio, book-to-market value ratio, default yield spread, net equity expansion for the period 1/1927–12/1994, which is in line with the findings in Campbell and Yogo (2006), and reported predictability only with respect to term spread for the period 1/1952–12/2008 while the method in Campbell and Yogo (2006) showed predictability for dividend payout ratio, dividend yield, T-bill rate and term spread; see Table 6 in Kostakis et al. (2015). Implementing these tests assumes that U_t 's in (1) are uncorrelated errors, but this assumption has not been examined in both Campbell and Yogo (2006) and Kostakis et al. (2015). To this end, we plot the autocorrelation function (ACF) of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}X_{t-1}$ from model (1) with $\hat{\alpha}$ and $\hat{\beta}$ being the least squares estimators and Y_t being the CRSP value-weighted excess return in Figures 1 – 3 for the periods 1/1927–12/1994, 1/1952–12/2015, and 1/1982–12/2015, respectively. Clearly, Figures 1 – 3 indicate that the assumption of uncorrelated U_t 's is doubtful for the period 1/1927–12/1994, may be fine for the period 1/1952–12/2015, and is quite reasonable for the period 1/1982–12/2015. Therefore, conclusions on predictability in the literature for the period from 1/1927 to 12/1994 may be misleading due to the violation of the model assumption of uncorrelated errors.

When we plot the ACF of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}X_{t-1}$ for Y_t being the S&P 500 excess return in Figures 4 – 6, it is easy to conclude that the assumption of uncorrelated U_t 's does not hold for either of the aforementioned three periods. On the other hand, the ACF of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\gamma}Y_{t-1} - \hat{\beta}X_{t-1}$ from (2) in Figure 7 suggests that the assumption of uncorrelated U_t 's is reasonable for the period 1/1982–12/2015 with Y_t being the S&P 500 excess return. Similarly, Figure 8 suggests that the assumption of uncorrelated U_t 's in (2) is valid for the period 1/1982–12/2015 with Y_t being the CRSP value-weighted excess return too. Therefore, our conducted data analyses will be focused on the period 1/1982–12/2015 as suggested by the above ACF analyses and detailed findings will be reported in Section 3.

The main contribution of this paper is to propose novel procedures for testing $H_0 : \gamma_0 = 0$

and/or $H_0 : \beta_0 = 0$ without characterizing the stochastic properties of the regressor under model (2). Specifically, we investigate the possibility of applying the idea of the robust empirical likelihood inference in Zhu et al. (2014). Readers are referred to Owen (2001) for an overview on empirical likelihood method, which has been proved to be quite effective in interval estimations and hypothesis tests.

The rest of this paper is organized as follows. Section 2 presents the methodologies and main asymptotic results. A simulation study and real data analysis are given in Section 3. Some concluding remarks are depicted in Section 4. All proofs are relegated to the Appendix.

2 Methodologies and Main Asymptotic Results

In order to model the regressor reasonably well, we fit an ARMA(1,15) to the listed regressors in the introduction for the period 1/1982–12/2015 and plot the ACF in Figure 9, which shows evidently that the fitting is good. Hence, motivated by the aforementioned real examples, we consider the following general dynamic predictive regression model

$$Y_t = \alpha + \gamma Y_{t-1} + \beta X_{t-1} + U_t, \quad X_t = \theta + \phi X_{t-1} + \sum_{j=0}^{\infty} \psi_j V_{t-j}, \quad 1 \leq t \leq n, \quad (3)$$

where $\{\sum_{j=0}^{\infty} \psi_j V_{t-j}\}$ is a strictly stationary sequence and $\{(U_t, V_t)^T\}$ is a sequence of independent and identically distributed (iid) random vectors with zero means and finite variances. Of our interest is to test $H_0 : \gamma_0 = 0$ and/or $H_0 : \beta_0 = 0$ regardless of $\{X_t\}$ being stationary (i.e., $|\phi_0| < 1$) or nearly integrated (i.e., $\phi_0 = 1 - \rho/n$ with $\rho \neq 0$) or unit root (i.e., $\phi_0 = 1$).

2.1 Model with a Known Intercept

To better appreciate the methodology, we first consider the case by assuming that $\alpha = \alpha_0$ is known, which may have an independent interest too. In this case, to find the least squares estimator for $(\gamma, \beta)^T$ based on the first equation in (3), one shall solve the following score equations

$$\sum_{t=1}^n (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1}) Y_{t-1} = 0 \quad \text{and} \quad \sum_{t=1}^n (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1}) X_{t-1} = 0,$$

which are equivalent to

$$\begin{cases} \sum_{t=1}^n (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1})(Y_{t-1} - \beta X_{t-1}) = 0, \\ \sum_{t=1}^n (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1}) X_{t-1} = 0. \end{cases} \quad (4)$$

The reason to use $Y_{t-1} - \beta X_{t-1}$ instead of Y_{t-1} is that $\{Y_{t-1} - \beta X_{t-1}\}$ becomes stationary when $\{X_t\}$ is a unit root process. To make an inference about γ and/or β , one may directly apply the empirical likelihood method based on estimating equations in Qin and Lawless (1994) to (4) but it is easy to show that this does not lead to a chi-squared limit in case of nearly integrated $\{X_t\}$, that is, the Wilks theorem² does not hold; see Zhu et al. (2014) for details. To fix this issue, following the idea in Zhu et al. (2014), we replace the second equation in (4) by the following weighted score equation

$$\sum_{t=1}^n (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1}) \frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} = 0. \quad (5)$$

The purpose of adding a weight into (5) is to ensure that

$$\frac{\left\{ \sum_{t=1}^n (Y_t - \alpha_0 - \gamma_0 Y_{t-1} - \beta_0 X_{t-1}) \frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} \right\}^2}{\sum_{t=1}^n (Y_t - \alpha_0 - \gamma_0 Y_{t-1} - \beta_0 X_{t-1})^2 \frac{X_{t-1}^2}{1 + X_{t-1}^2}} \xrightarrow{d} \chi^2(1)$$

as $n \rightarrow \infty$ by noting that $|X_{t-1}|/\sqrt{1 + X_{t-1}^2} \xrightarrow{p} 1$ as $t \rightarrow \infty$ when $\{X_t\}$ is a nearly integrated or unit root process.

To describe the proposed empirical likelihood tests, we introduce the following notation. For $t = 1, 2, \dots, n$, define

$$\begin{cases} Z_{t1}(\gamma, \beta) = (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1})(Y_{t-1} - \beta X_{t-1}), \\ Z_{t2}(\gamma, \beta) = (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1}) \frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}}. \end{cases}$$

Based on $\{\mathbf{Z}_t(\gamma, \beta)\}_{t=1}^n$ with $\mathbf{Z}_t(\gamma, \beta) = (Z_{t1}(\gamma, \beta), Z_{t2}(\gamma, \beta))^T$, the empirical likelihood function

²The Wilks theorem says that the asymptotic limit is independent of the true parameters; see Bickel and Doksum (2001) for details.

for γ and β is given by

$$L(\gamma, \beta) = \sup \left\{ \prod_{t=1}^n (np_t) : p_1 \geq 0, \dots, p_n \geq 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t \mathbf{Z}_t(\gamma, \beta) = 0 \right\}.$$

Then it follows from the Lagrange multiplier technique that

$$-2 \log L(\gamma, \beta) = 2 \sum_{t=1}^n \log \{1 + \boldsymbol{\lambda}^T \mathbf{Z}_t(\gamma, \beta)\},$$

where $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\gamma, \beta)$ satisfies the following equation

$$\sum_{t=1}^n \frac{\mathbf{Z}_t(\gamma, \beta)}{1 + \boldsymbol{\lambda}^T \mathbf{Z}_t(\gamma, \beta)} = 0.$$

If we are interested in testing $H_0 : \gamma_0 = 0$, then we consider the profile empirical likelihood function $L^{P1}(\gamma) = \max_{\beta} L(\gamma, \beta)$. On the other hand, if the interest is in testing $H_0 : \beta_0 = 0$, then one considers the profile empirical likelihood function $L^{P2}(\beta) = \max_{\gamma} L(\gamma, \beta)$. The following theorem shows that the Wilks theorem holds for the above proposed empirical likelihood method.

Theorem 1. *Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$, and $\alpha = \alpha_0$ is known. Further assume either (i) $|\phi_0| < 1$ independent of n (stationary case), or (ii) $\phi_0 = 1 - \rho/n$ for some $\rho \neq 0$ (nearly integrated case), or (iii) $\phi_0 = 1$ (unit root case). Then, as $n \rightarrow \infty$, $-2 \log L^{P1}(0) \xrightarrow{d} \chi^2(1)$ under $H_0 : \gamma_0 = 0$, $-2 \log L^{P2}(0) \xrightarrow{d} \chi^2(1)$ under $H_0 : \beta_0 = 0$, and $-2 \log L(0, 0) \xrightarrow{d} \chi^2(2)$ under $H_0 : \gamma_0 = 0$ & $\beta_0 = 0$.*

Based on the above theorem, a robust empirical likelihood test for testing $H_0 : \gamma_0 = 0$ or $H_0 : \beta_0 = 0$ or $H_0 : \gamma_0 = 0$ & $\beta_0 = 0$ at level ξ is to reject H_0 if $-2 \log L^{P1}(0) > \chi_{1,1-\xi}^2$ or $-2 \log L^{P2}(0) > \chi_{1,1-\xi}^2$ or $-2 \log L(0, 0) > \chi_{2,1-\xi}^2$, respectively, where $\chi_{1,1-\xi}^2$ and $\chi_{2,1-\xi}^2$ denote the $(1 - \xi)$ -th quantile of a chi-squared limit with one degree of freedom and with two degrees of freedom, respectively. Clearly the proposed robust tests above do not need a prior on whether $\{X_t\}$ is stationary or nearly integrated or unit root, and whether $\theta_0 = 0$ or $\theta_0 \neq 0$.

Remark 1. *When $\{U_t\}$ follows an autoregressive model rather than independent random variables, the above theorem does not hold. Instead one should take the error structure into account like the studies in Xiao et al. (2003) and Liu et al. (2010) for nonparametric regression models. Here, for the purpose of unifying the cases of stationary, nearly integrated and unit root, one can follow the idea in Li et al. (2017) to take the model structure of $\{U_t\}$ into account by employing*

either empirical likelihood method or jackknife empirical likelihood method in Jing et al. (2009).

The following theorems analyze the test power of the above empirical likelihood test separately for the cases of $\{X_t\}$ being stationary, nonstationary with zero intercept and nonstationary with nonzero intercept.

Theorem 2. *Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$, and $\alpha = \alpha_0$ is known. Further assume $|\phi_0| < 1$ independent of n .*

(i) *Under $H_a : \gamma_0 = d_1/\sqrt{n}$ for some $d_1 \in \mathbb{R}$ and $\beta_0 = d_2/\sqrt{n}$ for some $d_2 \in \mathbb{R}$, we have*

$$-2 \log L(0, 0) = (\mathbf{W}_1 + \mathbf{D}_1)^T \Sigma_1^{-1} (\mathbf{W}_1 + \mathbf{D}_1) + o_p(1),$$

which has a non-central chi-squared limit with two degrees of freedom and non-centrality parameter $\mathbf{D}_1^T \Sigma_1^{-1} \mathbf{D}_1 > 0$ when $d_1^2 + d_2^2 > 0$, where $\mathbf{W}_1 \sim N(0, \Sigma_1)$,

$$\mathbf{D}_1 = \begin{pmatrix} d_1 \{E(U_1^2) + \alpha_0^2\} + d_2 E\{X_1(U_1 + \alpha_0)\} \\ d_1 E\left(\frac{(\alpha_0 + U_1)X_1}{\sqrt{1+X_1^2}}\right) + d_2 E\left(\frac{X_1^2}{\sqrt{1+X_1^2}}\right) \end{pmatrix}, \quad \Sigma_1 = E(U_1^2) \begin{pmatrix} E(U_1^2) + \alpha_0^2 & E\left(\frac{(U_1 + \alpha_0)X_1}{\sqrt{1+X_1^2}}\right) \\ E\left(\frac{(U_1 + \alpha_0)X_1}{\sqrt{1+X_1^2}}\right) & E\left(\frac{X_1^2}{1+X_1^2}\right) \end{pmatrix}.$$

(ii) *Under $H_a : \gamma_0 = d_1/\sqrt{n}$ for some $d_1 \in \mathbb{R}$ and β_0 is a nonzero constant, we have*

$$-2 \log L^{P1}(0) = (\mathbf{W}_2 + \mathbf{D}_2)^T \Sigma_2^{-1} (\mathbf{W}_2 + \mathbf{D}_2) + o_p(1),$$

which has a non-central chi-squared limit with one degree of freedom and non-centrality parameter $\mathbf{D}_2^T \Sigma_2^{-1} \mathbf{D}_2 > 0$ when $d_1 \neq 0$, where $\mathbf{W}_2 \sim N(0, \Sigma_2)$,

$$\mathbf{D}_2 = \begin{pmatrix} d_1 E\{(\alpha_0 + \beta_0 X_1 + U_2)(\alpha_0 + U_2 + \beta_0 X_1 - \beta_0 X_2)\} \\ d_1 E\left(\frac{(\alpha_0 + \beta_0 X_1 + U_2)X_2}{\sqrt{1+X_2^2}}\right) \end{pmatrix},$$

$$\Sigma_2 = E(U_1^2) \begin{pmatrix} E(\alpha_0 + U_2 + \beta_0 X_1 - \beta_0 X_2)^2 & E\left(\frac{(\alpha_0 + U_2 + \beta_0 X_1 - \beta_0 X_2)X_2}{\sqrt{1+X_2^2}}\right) \\ E\left(\frac{(\alpha_0 + U_2 + \beta_0 X_1 - \beta_0 X_2)X_2}{\sqrt{1+X_2^2}}\right) & E\left(\frac{X_2^2}{1+X_2^2}\right) \end{pmatrix}.$$

(iii) *Under $H_a : \beta_0 = d_2/\sqrt{n}$ for some $d_2 \in \mathbb{R}$ and γ_0 is a nonzero constant, we have*

$$-2 \log L^{P2}(0) = (\mathbf{W}_3 + \mathbf{D}_3)^T \Sigma_3^{-1} (\mathbf{W}_3 + \mathbf{D}_3) + o_p(1),$$

which has a non-central chi-squared limit with one degree of freedom and non-centrality parameter

$\mathbf{D}_3^T \Sigma_3^{-1} \mathbf{D}_3 > 0$ when $d_2 \neq 0$, where $\mathbf{W}_3 \sim N(0, \Sigma_3)$,

$$\mathbf{D}_3 = \lim_{t \rightarrow \infty} \begin{pmatrix} d_2 E\{X_{t-1}(\frac{\alpha_0}{1-\gamma_0} + \sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j)\} \\ d_2 E(\frac{X_1^2}{\sqrt{1+X_1^2}}) \end{pmatrix},$$

$$\Sigma_3 = E(U_1^2) \lim_{t \rightarrow \infty} \begin{pmatrix} E(\frac{\alpha_0}{1-\gamma_0} + \sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j)^2 & E\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}(\frac{\alpha_0}{1-\gamma_0} + \sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j)\} \\ E\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}(\frac{\alpha_0}{1-\gamma_0} + \sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j)\} & E(\frac{X_1^2}{1+X_1^2}) \end{pmatrix}.$$

Theorem 3. Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$, and $\alpha = \alpha_0$ is known. Further assume $\phi_0 = 1 - \rho/n$ for some $\rho \in \mathbb{R}$ and $\theta_0 = 0$.

(i) Under $H_a : \gamma_0 = d_1/\sqrt{n}$ for some $d_1 \in \mathbb{R}$ and $\beta_0 = d_2/n$ for some $d_2 \in \mathbb{R}$, we have

$$-2 \log L(0, 0) = (\tilde{\mathbf{W}}_1 + \tilde{\mathbf{D}}_1)^T \tilde{\Sigma}_1^{-1} (\tilde{\mathbf{W}}_1 + \tilde{\mathbf{D}}_1) + o_p(1),$$

which has a non-central chi-squared limit with two degrees of freedom and non-centrality parameter $\tilde{\mathbf{D}}_1^T \tilde{\Sigma}_1^{-1} \tilde{\mathbf{D}}_1 > 0$ when $d_1^2 + d_2^2 > 0$, where $\tilde{\mathbf{W}}_1 \sim N(0, \tilde{\Sigma}_1)$,

$$\tilde{\mathbf{D}}_1 = \begin{pmatrix} d_1 \{E(U_1^2) + \alpha_0^2\} + d_2 \alpha_0 \int_0^1 J_{V,\rho}(s) ds \\ d_1 \alpha_0 + d_2 \int_0^1 J_{V,\rho}(s) ds \end{pmatrix}, \quad \tilde{\Sigma}_1 = E(U_1^2) \begin{pmatrix} E(U_1^2) + \alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix},$$

$J_{V,\rho}(r) = \int_0^r e^{-(r-s)\rho} dW_V(s)$ and $W_V(s) = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \sum_{j=0}^{\infty} \psi_j V_{t-j}$ for $s \in [0, 1]$.

(ii) Under $H_a : \gamma_0 = d_1/n$ for some $d_1 \in \mathbb{R}$ and β_0 is a nonzero constant, we have

$$-2 \log L^{P1}(0) = (\tilde{\mathbf{W}}_2 + \tilde{\mathbf{D}}_2)^T \left\{ \tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1} \tilde{S}_2 \tilde{S}_2^T \tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2} \right\} (\tilde{\mathbf{W}}_2 + \tilde{\mathbf{D}}_2) + o_p(1),$$

which has a central chi-squared limit with one degree of freedom even when $d_1 \neq 0$, where $\tilde{S}_2 = -(\alpha_0, 1)^T$, $\tilde{\mathbf{W}}_2 \sim N(0, \tilde{\Sigma}_2)$,

$$\tilde{\mathbf{D}}_2 = -\tilde{S}_2 d_1 \beta_0 \int_0^1 J_{V,\rho}(s) ds, \quad \tilde{\Sigma}_2 = E(U_1^2) \begin{pmatrix} E(U_1 - \beta_0 \sum_{j=0}^{\infty} \psi_j V_{1-j})^2 + \alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix}.$$

(iii) Under $H_a : \beta_0 = d_2/n$ for some $d_2 \in \mathbb{R}$ and γ_0 is a nonzero constant, we have

$$-2 \log L^{P2}(0) = (\tilde{\mathbf{W}}_3 + \tilde{\mathbf{D}}_3)^T \left\{ \tilde{\Sigma}_3^{-1} - \frac{\tilde{\Sigma}_3^{-1} \tilde{S}_3 \tilde{S}_3^T \tilde{\Sigma}_3^{-1}}{\tilde{S}_3^T \tilde{\Sigma}_3^{-1} \tilde{S}_3} \right\} (\tilde{\mathbf{W}}_3 + \tilde{\mathbf{D}}_3) + o_p(1),$$

which has a non-central chi-squared limit with one degree of freedom and non-central parameter

$\tilde{\mathbf{D}}_3 \{ \tilde{\Sigma}_3^{-1} - \frac{\tilde{\Sigma}_3^{-1} \tilde{S}_3 \tilde{S}_3^T \tilde{\Sigma}_3^{-1}}{\tilde{S}_3^T \tilde{\Sigma}_3^{-1} \tilde{S}_3} \} \tilde{\mathbf{D}}_3 > 0$ when $d_2 \neq 0$, where

$$\tilde{S}_3 = - \left(\lim_{t \rightarrow \infty} E \left(\sum_{j=1}^t \gamma_0^{t-j} U_j \right)^2 + \left(\frac{\alpha_0}{1-\gamma_0} \right)^2, \quad \frac{\alpha_0}{1-\gamma_0} \right)^T, \quad \tilde{\mathbf{W}}_3 \sim N(0, \tilde{\Sigma}_3),$$

$$\tilde{\mathbf{D}}_3 = \begin{pmatrix} d_2 \frac{\alpha_0}{1-\gamma_0} \int_0^1 J_{V,\rho}(s) ds \\ d_2 \int_0^1 J_{V,\rho}(s) ds \end{pmatrix}, \quad \tilde{\Sigma}_3 = E(U_1^2) \lim_{t \rightarrow \infty} \begin{pmatrix} E(\sum_{j=1}^t \gamma_0^{t-j} U_j)^2 + \left(\frac{\alpha_0}{1-\gamma_0} \right)^2 & \frac{\alpha_0}{1-\gamma_0} \\ \frac{\alpha_0}{1-\gamma_0} & 1 \end{pmatrix}.$$

Theorem 4. Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$, and $\alpha = \alpha_0$ is known. Further assume $\phi_0 = 1 - \rho/n$ for some $\rho \in \mathbb{R}$ and $\theta_0 \neq 0$.

(i) Under $H_a : \gamma_0 = d_1/\sqrt{n}$ for some $d_1 \in \mathbb{R}$ and $\beta_0 = d_2/n^{3/2}$ for some $d_2 \in \mathbb{R}$, we have

$$-2 \log L(0, 0) = (\bar{\mathbf{W}}_1 + \bar{\mathbf{D}}_1)^T \bar{\Sigma}_1^{-1} (\bar{\mathbf{W}}_1 + \bar{\mathbf{D}}_1) + o_p(1),$$

which has a non-central chi-squared limit with two degrees of freedom and non-central parameter

$\bar{\mathbf{D}}_1 \bar{\Sigma}_1^{-1} \bar{\mathbf{D}}_1 > 0$ when $d_1^2 + d_2^2 > 0$, where $\bar{\mathbf{W}}_1 \sim N(0, \bar{\Sigma}_1)$,

$$\bar{\mathbf{D}}_1 = \begin{pmatrix} d_1 \{ E(U_1^2) + \alpha_0^2 \} + d_2 \alpha_0 \theta_0 \int_0^1 \frac{1-e^{-\rho s}}{\rho} ds \\ d_1 \alpha_0 \operatorname{sgn}(\theta_0) + d_2 |\theta_0| \int_0^1 \frac{1-e^{-\rho s}}{\rho} ds \end{pmatrix}, \quad \bar{\Sigma}_1 = E(U_1^2) \begin{pmatrix} E(U_1^2) + \alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix},$$

and $\operatorname{sgn}(x)$ denotes the sign function.

(ii) Under $H_a : \gamma_0 = d_1/n^{3/2}$ for some $d_1 \in \mathbb{R}$ and β_0 is a nonzero constant, we have

$$-2 \log L^{P1}(0) = (\bar{\mathbf{W}}_2 + \bar{\mathbf{D}}_2)^T \left\{ \bar{\Sigma}_2^{-1} - \frac{\bar{\Sigma}_2^{-1} \bar{S}_2 \bar{S}_2^T \bar{\Sigma}_2^{-1}}{\bar{S}_2^T \bar{\Sigma}_2^{-1} \bar{S}_2} \right\} (\bar{\mathbf{W}}_2 + \bar{\mathbf{D}}_2) + o_p(1),$$

which has a central chi-squared limit with one degree of freedom even when $d_1 \neq 0$, where

$\bar{S}_2 = -(\theta_0(\alpha_0 - \beta_0 \theta_0), |\theta_0|)^T \int_0^1 \frac{1-e^{-\rho s}}{\rho} ds$, $\bar{\mathbf{W}}_2 \sim N(0, \bar{\Sigma}_2)$, $\bar{\mathbf{D}}_2 = -\bar{S}_2 d_1 \beta_0$,

$$\bar{\Sigma}_2 = E(U_1^2) \begin{pmatrix} E(\alpha_0 + U_1 - \beta_0 \theta_0 - \beta_0 \sum_{j=0}^{\infty} \psi_j V_{1-j})^2 & (\alpha_0 - \beta_0 \theta_0) \operatorname{sgn}(\theta_0) \\ (\alpha_0 - \beta_0 \theta_0) \operatorname{sgn}(\theta_0) & 1 \end{pmatrix}.$$

(iii) Under $H_a : \beta_0 = d_2/n^{3/2}$ for some $d_2 \in \mathbb{R}$ and γ_0 is a nonzero constant, we have

$$-2 \log L^{P2}(0) = (\bar{\mathbf{W}}_3 + \bar{\mathbf{D}}_3)^T \left\{ \bar{\Sigma}_3^{-1} - \frac{\bar{\Sigma}_3^{-1} \bar{S}_3 \bar{S}_3^T \bar{\Sigma}_3^{-1}}{\bar{S}_3^T \bar{\Sigma}_3^{-1} \bar{S}_3} \right\} (\bar{\mathbf{W}}_3 + \bar{\mathbf{D}}_3) + o_p(1),$$

which has a non-central chi-squared limit with one degree of freedom and non-central parameter

$\bar{D}_3\{\bar{\Sigma}_3^{-1} - \frac{\bar{\Sigma}_3^{-1}\bar{S}_3\bar{S}_3^T\bar{\Sigma}_3^{-1}}{\bar{S}_3^T\bar{\Sigma}_3^{-1}\bar{S}_3}\}\bar{D}_3 > 0$ when $d_2 \neq 0$, where

$$\bar{S}_3 = - \left(\lim_{t \rightarrow \infty} E \left(\sum_{j=1}^t \gamma_0^{t-j} U_j \right)^2 + \left(\frac{\alpha_0}{1-\gamma_0} \right)^2, \quad \frac{\alpha_0 \text{sgn}(\theta_0)}{1-\gamma_0} \right)^T, \quad \bar{W}_3 \sim N(0, \bar{\Sigma}_3),$$

$$\bar{D}_3 = \begin{pmatrix} d_2 \theta_0 \frac{\alpha_0}{1-\gamma_0} \int_0^1 \frac{1-e^{-\rho s}}{\rho} ds \\ d_2 |\theta_0| \int_0^1 \frac{1-e^{-\rho s}}{\rho} ds \end{pmatrix}, \quad \bar{\Sigma}_3 = E(U_1^2) \lim_{t \rightarrow \infty} \begin{pmatrix} E(\sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j)^2 + \left(\frac{\alpha_0}{1-\gamma_0} \right)^2 & \frac{\alpha_0}{1-\gamma_0} \text{sgn}(\theta_0) \\ \frac{\alpha_0}{1-\gamma_0} \text{sgn}(\theta_0) & 1 \end{pmatrix}.$$

Remark 2. Theorems 3(ii) and 4(ii) show that testing for $H_0 : \beta_0 = 0$ is more powerful than that for $H_0 : \gamma_0 = 0$ when $\{X_t\}$ is a nearly integrated or unit root process. The reason is that \tilde{D}_2 and \bar{D}_2 become a multiplier of \tilde{S}_2 and \bar{S}_2 , respectively in these two cases.

2.2 Model with an Unknown Intercept

Next, we consider the case that α in model (3) is unknown. Again, our interest is to test $H_0 : \gamma_0 = 0$ and/or $H_0 : \beta_0 = 0$ without knowing whether $\{X_t\}$ is stationary or nearly integrated or unit root.

As before, one may simply apply the empirical likelihood method to the following weighted score equations:

$$\begin{cases} \sum_{t=1}^n \{Y_t - \alpha - \gamma Y_{t-1} - \beta X_{t-1}\} = 0 \\ \sum_{t=1}^n \{Y_t - \alpha - \gamma Y_{t-1} - \beta X_{t-1}\} \{Y_{t-1} - \beta X_{t-1}\} = 0 \\ \sum_{t=1}^n \{Y_t - \alpha - \gamma Y_{t-1} - \beta X_{t-1}\} \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} = 0. \end{cases} \quad (6)$$

However, this does not work by noting that the joint normalized limit of the first and third equations in (6) is degenerate in the near unit root and unit root cases. Moreover, by re-expressing the first two equations in (6) as

$$\begin{cases} \sum_{t=1}^n (Y_t - \alpha - \gamma Y_{t-1} - \beta X_{t-1}) = \sum_{t=1}^n (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1}) - n(\alpha - \alpha_0), \\ \sum_{t=1}^n (Y_t - \alpha - \gamma Y_{t-1} - \beta X_{t-1})(Y_{t-1} - \beta X_{t-1}) = \sum_{t=1}^n (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1}) \times \\ (Y_{t-1} - \beta X_{t-1}) - (\alpha - \alpha_0) \sum_{t=1}^n (Y_{t-1} - \beta X_{t-1}), \end{cases}$$

it is clear that the term $(\alpha - \alpha_0) \sum_{t=1}^n (Y_{t-1} - \beta_0 X_{t-1})$ in the above second equation becomes a smaller order than the term $n(\alpha - \alpha_0)$ in the above first equation when $\theta_0 = 0$ and $\alpha_0 = 0$ by noting that $\frac{1}{n} \sum_{t=1}^n (Y_{t-1} - \beta_0 X_{t-1}) \xrightarrow{P} 0$ in this case. Hence, when we profile the nuisance parameter α later, a solution for α in the first equation of (6) may not be a solution for α in the second equation of (6), which causes a problem in the minimization. These problems are not surprising since it is known in the literature that an inference for a unit root or nearly integrated process has a different asymptotic distribution when the model has zero or nonzero intercept, and unifying the cases with and without intercept requires a careful effort.

To overcome these issues and unify the cases with and without intercept, we follow the idea in Zhu et al. (2014) by splitting the data into two parts and using the difference with a big lag to get rid of the intercept first so as to still keep the differences of regressor as a nonstationary process. This is important as inference with nonstationarity has a faster rate of convergence than that in the stationary case. More specifically, put $m = [n/2]$ with $[\cdot]$ denoting the ceiling function, and define $\tilde{X}_t = X_{t+m} - X_t$, $\tilde{Y}_t = Y_{t+m} - Y_t$, $\tilde{U}_t = U_{t+m} - U_t$ and $\tilde{V}_t = V_{t+m} - V_t$ for $t = 1, \dots, m$. Then, model (3) implies the following model

$$\tilde{Y}_t = \gamma \tilde{Y}_{t-1} + \beta \tilde{X}_{t-1} + \tilde{U}_t$$

without intercept, which is the same as model (3) with known $\alpha = 0$. Clearly, if $\{U_t\}$ is independent, then is $\{\tilde{U}_t\}$. Furthermore, $|\tilde{X}_t| \xrightarrow{P} \infty$ and $|\tilde{X}_t|/\sqrt{1 + \tilde{X}_t^2} \xrightarrow{P} 1$ as $t \rightarrow \infty$ when $\{X_t\}$ is nearly integrated or unit root. As discussed before, this property is the key to ensure that the Wilks theorem holds for the proposed empirical likelihood test.

Therefore, similar to the model with known intercept in (3), we define

$$\tilde{\mathbf{Z}}_t(\gamma, \beta) = (\tilde{Z}_{t1}(\gamma, \beta), \tilde{Z}_{t2}(\gamma, \beta))^T,$$

where

$$\begin{cases} \tilde{Z}_{t1}(\gamma, \beta) = (\tilde{Y}_t - \gamma \tilde{Y}_{t-1} - \beta \tilde{X}_{t-1})(\tilde{Y}_{t-1} - \beta \tilde{X}_{t-1}), \\ \tilde{Z}_{t2}(\gamma, \beta) = (\tilde{Y}_t - \gamma \tilde{Y}_{t-1} - \beta \tilde{X}_{t-1}) \frac{\tilde{X}_{t-1}}{\sqrt{1 + \tilde{X}_{t-1}^2}}. \end{cases}$$

Then, based on $\{\tilde{\mathbf{Z}}_t(\gamma, \beta)\}_{t=1}^m$, the empirical likelihood function for γ and β is defined as

$$\tilde{L}(\gamma, \beta) = \sup \left\{ \prod_{t=1}^m (mp_t) : p_1 \geq 0, \dots, p_m \geq 0, \sum_{t=1}^m p_t = 1, \sum_{t=1}^m p_t \tilde{\mathbf{Z}}_t(\gamma, \beta) = 0 \right\}.$$

If we are interested in testing $H_0 : \gamma_0 = 0$, then we consider the profile empirical likelihood function $\tilde{L}^{P1}(\gamma) = \max_{\beta} \tilde{L}(\gamma, \beta)$. On the other hand, if the interest is in testing $H_0 : \beta_0 = 0$, then one considers the profile empirical likelihood function $\tilde{L}^{P2}(\beta) = \max_{\gamma} \tilde{L}(\gamma, \beta)$. The following theorem shows that the Wilks theorem holds for this empirical likelihood test.

Theorem 5. *Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$. Further assume either (i) $|\phi_0| < 1$ independent of n , or (ii) $\phi_0 = 1 - \rho/n$ for some $\rho \neq 0$, or (iii) $\phi_0 = 1$. Then, as $n \rightarrow \infty$, $-2 \log \tilde{L}^{P1}(0) \xrightarrow{d} \chi^2(1)$ under $H_0 : \gamma_0 = 0$, $-2 \log \tilde{L}^{P2}(0) \xrightarrow{d} \chi^2(1)$ under $H_0 : \beta_0 = 0$, and $-2 \log \tilde{L}(0, 0) \xrightarrow{d} \chi^2(2)$ under $H_0 : \gamma_0 = 0 \ \& \ \beta_0 = 0$.*

Again, based on the above theorem, a robust empirical likelihood test for $H_0 : \gamma_0 = 0$ or $H_0 : \beta_0 = 0$ or $H_0 : \gamma_0 = 0 \ \& \ \beta_0 = 0$ under model (3) is to reject H_0 at level ξ whenever $-2 \log \tilde{L}^{P1}(0) > \chi_{1,1-\xi}^2$ or $-2 \log \tilde{L}^{P2}(0) > \chi_{1,1-\xi}^2$ or $-2 \log \tilde{L}(0, 0) > \chi_{2,1-\xi}^2$, respectively. These tests are robust without knowing whether $\{X_t\}$ is stationary or nearly integrated or unit root and has a zero or nonzero intercept. Similar power analyses like Theorems 2–4 can be done.

3 Finite Sample Analysis

3.1 Monte Carlo Simulation Study

In this subsection, we investigate the performance of the proposed robust tests in terms of size and power. Note that all simulations are implemented in the statistical software R.

Consider model (3) with $\alpha_0 = 1$, $\theta_0 = 0$, $\psi_0 = 1$, $\psi_j = 0$ for $j \geq 1$, $\phi = 0.9$ or $1 - 2/n$ or 1, where the sample size $n = 200$ or 2000 or 5000. For testing $H_0 : \gamma_0 = 0$, we take $\beta_0 = 0.5$ and consider $\gamma_0 = 0, \pm \frac{1}{\sqrt{n}}, \pm \frac{2}{\sqrt{n}}, \pm \frac{3}{\sqrt{n}}, \pm \frac{4}{\sqrt{n}}$. Hence, the results for $\gamma_0 = 0$ represent the size of the proposed test and the results for nonzero γ 's represent the power of the proposed test. The reason for considering these scaled γ 's with the order of $1/\sqrt{n}$ instead of $1/n$ even for the nonstationary case is given in Remark 2. For testing $H_0 : \beta_0 = 0$, we take $\gamma_0 = 0.5$ and consider $\beta_0 = 0, \pm \frac{1}{\sqrt{n}}, \pm \frac{2}{\sqrt{n}}, \pm \frac{3}{\sqrt{n}}, \pm \frac{4}{\sqrt{n}}$ in case of $\phi_0 = 0.9$, and $\beta_0 = 0, \pm \frac{2}{n}, \pm \frac{4}{n}, \pm \frac{6}{n}, \pm \frac{8}{n}$ in case of $\phi_0 = 1 - 2/n$ and $\phi_0 = 1$.

We generate 1,000 random samples from model (3) with the above settings and $(U_t, V_t)^T$ having a normal copula with correlation $\rho = -0.5$ and marginal distributions $t(5)$ and $t(4)$. We compute the sizes and powers of the proposed tests with the significance levels 5% and 10% based on either Theorem 1 (i.e., α is assumed to be known) or Theorem 5 (i.e., α is unknown)

by employing the package *'emplik'* in the R software. Empirical sizes and powers are reported in Tables 1–4, respectively, for testing $H_0 : \gamma_0 = 0$ with known α or unknown α , for testing $H_0 : \beta_0 = 0$ with known α or unknown α .

The results for $\gamma_0 = 0$ in Tables 1 and 2 show that the size of the proposed test with known α is slightly more accurate than that with unknown α for $n = 200$, but both have a quite accurate size for a larger n . The results for $\gamma_0 \neq 0$ in Tables 1 and 2 show that the proposed test for $H_0 : \gamma_0 = 0$ has a good power whenever α is known or unknown, the power increases as γ_0 deviates more from zero, and the test with known α is more powerful than that with unknown α since the latter requires splitting the data into two parts, which reduces the effective sample size in the inference.

The results for $\beta_0 = 0$ in Tables 3 and 4 show that the proposed test for $H_0 : \beta_0 = 0$ has a good size whenever α is known or unknown. Also, the results for $\beta_0 \neq 0$ in Tables 3 and 4 show that the proposed test has a good power and the test with known α is more powerful than that with unknown α , which is due to the technique of splitting data.

In conclusion, Tables 1–4 show that the proposed robust tests for $H_0 : \gamma_0 = 0$ and/or $H_0 : \beta_0 = 0$ with a lagged predicted variable in a predictive regression perform pretty well regardless of $\{X_t\}$ being a stationary process, or a nearly integrated process or a unit root process. The test with known α is more powerful than the test with unknown α , which can be explained by the employed technique of splitting data in order to unify the cases with and without intercept in the predictive regression.

3.2 Real Data Analyses

This subsection demonstrates the practical usefulness of the proposed tests by applying them to test the predictability of stock returns in the U.S. market regardless of the financial variable (regressor) being stationary or nearly integrated or unit root.

We revisit the data analysis in Kostakis et al. (2015) and Cai and Wang (2014) by focusing on the period 1/1982–12/2015 for one of the two predicted variables: the CRSP value-weighted excess return and the S&P 500 excess return, and one of the ten financial predictors: dividend payout ratio, long-term yield, dividend yield, dividend-price ratio, T-bill rate, earnings-price ratio, book-to-market value ratio, default yield spread, net equity expansion and term spread. The reason to focus on this time period is due to the ACF analyses in Figures 1 – 9 supporting the assumption of uncorrelated errors in model (3).

In Tables 5 and 6, we report the P-values of the proposed robust empirical likelihood tests for testing $H_0 : \gamma_0 = 0$ and $H_0 : \beta_0 = 0$. When we say a known α in model (3), it means α is set to be the least squares estimator, that is, α minimizes the least squares distance $\sum_{t=1}^n \{Y_t - \alpha - \gamma Y_{t-1} - \beta X_{t-1}\}^2$. Some findings are summarized as follows by comparing the obtained P-values with the significance level 10%.

- When the predicted variable is the CRSP value-weighted excess return, the null hypothesis $H_0 : \gamma_0 = 0$ can not be rejected for all considered regressors and the predictability (i.e., $\beta_0 \neq 0$) exists for four regressors whenever α is treated as a known or an unknown parameter.
- When the predicted variable is the S&P 500 excess return, the null hypothesis $H_0 : \gamma_0 = 0$ is rejected for all cases whenever α is known or unknown, and the predictability (i.e., $\beta_0 \neq 0$) exists for four regressors when α is known, but there exists no predictability for all regressors when α is unknown.

As argued in the introduction, the existing literature on testing predictability often ignores checking uncorrelated errors, while the employed tests for predictability heavily rely on this assumption. In other words, the existing findings on predictability may not be methodologically rigorous. Based on a different time period, where the assumption of uncorrelated errors is checked to be reasonable, our proposed robust test does not reject $H_0 : \gamma_0 = 0$ for the predicted variable CRSP value-weighted excess return, which means the method in Demetrescu (2014) by adding a lagged predicted variable on purpose is valid for this predicted variable. However, our proposed robust test does reject $H_0 : \gamma_0 = 0$ for the predicted variable S&P 500 excess return, which means the method in Demetrescu (2014) for improving the test power of IVX-based tests is invalid for this predicted variable. Unlike the results in Campbell and Yogo (2006) and Kostakis et al. (2015) for the period after 1/1952, where the assumption of uncorrelated errors may be reasonable, our tests clearly reject predictability for the term spread. The predictability for dividend yield and no predictability for default yield spread are consistent with that in Campbell and Yogo (2006) for the period 1/1952–12/2008. In comparison with findings in Cai and Wang (2014) and Kostakis et al. (2015), the proposed robust tests clearly indicate that it is necessary to include a lagged predicted variable into the predictive regression for the predicted variable S&P 500 excess return.

4 Conclusions

Without characterizing the stochastic properties of regressors, this paper introduces new test statistics based on empirical likelihood with weighted score equations in a predictive regression to test both the existence of the lagged variables and the predictability, and reexamines the empirical evidence on the predictability of stock returns of Kostakis et al. (2015) and Cai and Wang (2014) using the proposed new robust tests. The Wilks theorem is proved for the proposed empirical likelihood tests regardless of regressors being stationary or nearly integrated or unit root. Hence the proposed new tests are easy to implement without any ad hoc method such as bootstrap method for obtaining critical values.

The Monte Carlo simulation study shows that the finite sample performance of the proposed tests is good. The proposed robust procedure for testing $H_0 : \gamma_0 = 0$ can be employed to check whether the method in Demetrescu (2014) is applicable. The empirical analysis shows that adding a lagged predicted variable for the S&P 500 excess return is necessary while there is no need to add the lag for the CRSP value-weighted excess return.

References

- Amihud, Y. and Hurvich, C.M. (2004). Predictive regressions: a reduced-bias estimation method. *Journal of Financial and Quantitative Analysis* **39**, 813–841.
- Bickel, P. and Doksum, K.A. (2001). Mathematical Statistics I. 2nd Edition. *Prentice Hall*, New Jersey.
- Billingsley, P. (1999). Convergence of Probability Measures. 2nd. *Wiley*, New York.
- Breitung, J. and Demetrescu, M. (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics* **187**, 358–375.
- Cai, Z. and Wang, Y. (2014). Testing predictive regression models with nonstationary regressors. *Journal of Econometrics* **178**, 4–14.
- Campbell, J.Y. (2008). Viewpoint: Estimating the equity premium. *Canadian Journal of Economics/Revue canadienne d'économie* **41**, 1–21.
- Campbell, J.Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics* **81**, 27–60.
- Cavanagh, C.L., Elliott, G. and Stock, J.H. (1995). Inference in models with nearly integrated regressors. *Econometric Theory* **11**, 1131–1147.
- Chen, W.W. and Deo, R.S. (2009). Bias reduction and likelihood-based almost exactly sized hypothesis testing in predictive regressions using the restricted likelihood. *Econometric Theory* **25**, 1143–1179.

- Chen, W.W., Deo, R.S. and Yi, Y. (2013). Uniform inference in predictive regression models. *Journal of Business and Economic Statistics* **31**, 525–533.
- Choi, Y., Jacewitz, S. and Park, J.Y. (2016). A reexamination of stock return predictability. *Journal of Econometrics* **192**, 168–189.
- Demetrescu, M. (2014). Enhancing the local power of IVX based tests in predictive regressions. *Economics Letters* **124**, 269–273.
- Demetrescu, M. and Rodrigues, P.M.M. (2016). Residual-augmented IVX predictive regression. *Working Paper*, Economics and Research Department, Banco de Portugal.
- Jansson, M. and Moreira, M.J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica* **74**, 681–714.
- Jing, B.Y., Yuan, J. and Zhou, W. (2009). Jackknife empirical likelihood. *Journal of American Statistical Association* **104**, 1224–1232.
- Kostakis, A., Magdalinos, T. and Stamatogiannis, M.P. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies* **28**, 1506–1553.
- Li, C., Li, D. and Peng, L. (2017). Uniform test for predictive regression with AR errors. *Journal of Business and Economic Statistics* **35**, 29–39.
- Liu, J.M., Chen, R. and Yao, Q. (2010). Nonparametric transfer function models. *Journal of Econometrics* **157**, 151–164.
- Magdalinos, T. and Phillips, P.C.B. (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory* **25**, 482–526.
- 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.B. (2015). Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics* **13**, 521–555.
- Phillips, P.C.B. and Lee, J.H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics* **177**, 250–264.
- Phillips, P.C.B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics* **136**, 115–130.
- Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating functions. *Annals of Statistics* **22**, 300–325.
- Stambaugh, R.F. (1999). Predictive regressions. *Journal of Financial Economics* **54**, 375–421.
- Xiao, Z., Linton, O.B., Carroll, R.J. and Mammen, E. (2003). More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *Journal of American Statistical Association* **98**, 980–992.
- Zhu, F., Cai, Z. and Peng, L. (2014). Predictive regressions for macroeconomic data. *Annals of Applied Statistics* **8**, 577–594.

Appendix: Proofs of Theorems

Before proving theorems, we need some lemmas.

Lemma 1. *Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$, and $\alpha = \alpha_0$ is known. Further assume $|\phi_0| < 1$ independent of n .*

(i) *If $\gamma_0 = d_1/\sqrt{n}$ for some $d_1 \in \mathbb{R}$ and $\beta_0 = d_2/\sqrt{n}$ for some $d_2 \in \mathbb{R}$, then*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{Z}_t(0, 0) = \mathbf{W}_1 + \begin{pmatrix} d_1 \{E(U_1^2) + \alpha_0^2\} + d_2 E\{X_1(\alpha_0 + U_1)\} \\ d_1 E\left(\frac{(\alpha_0 + U_1)X_1}{\sqrt{1+X_1^2}}\right) + d_2 E\left(\frac{X_1^2}{\sqrt{1+X_1^2}}\right) \end{pmatrix} + o_p(1), \quad (7)$$

$$\frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t(0, 0) \mathbf{Z}_t^T(0, 0) = E(U_1^2) \begin{pmatrix} E(U_1^2) + \alpha_0^2 & E\left(\frac{(\alpha_0 + U_1)X_1}{\sqrt{1+X_1^2}}\right) \\ E\left(\frac{(\alpha_0 + U_1)X_1}{\sqrt{1+X_1^2}}\right) & E\left(\frac{X_1^2}{1+X_1^2}\right) \end{pmatrix} + o_p(1) := \Sigma_1 + o_p(1), \quad (8)$$

$$\text{and } \max_{1 \leq t \leq n} \|\mathbf{Z}_t(0, 0)\| = o_p(n^{1/2}), \quad (9)$$

where $\mathbf{W}_1 \sim N(0, \Sigma_1)$.

(ii) *If $\gamma_0 = d_1/\sqrt{n}$ for some $d_1 \in \mathbb{R}$ and β_0 is a nonzero constant, then*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{Z}_t(0, \beta_0) = \mathbf{W}_2 + \begin{pmatrix} d_1 E\{(\alpha_0 + \beta_0 X_1 + U_2)(\alpha_0 + U_2 + \beta_0 X_1 - \beta_0 X_2)\} \\ d_1 E\left(\frac{(\alpha_0 + \beta_0 X_1 + U_2)X_2}{\sqrt{1+X_2^2}}\right) \end{pmatrix} + o_p(1),$$

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t(0, \beta_0) \mathbf{Z}_t^T(0, \beta_0) \\ &= E(U_1^2) \begin{pmatrix} E(\alpha_0 + U_2 + \beta_0 X_1 - \beta_0 X_2)^2 & E\left(\frac{(\alpha_0 + U_2 + \beta_0 X_1 - \beta_0 X_2)X_2}{\sqrt{1+X_2^2}}\right) \\ E\left(\frac{(\alpha_0 + U_2 + \beta_0 X_1 - \beta_0 X_2)X_2}{\sqrt{1+X_2^2}}\right) & E\left(\frac{X_2^2}{1+X_2^2}\right) \end{pmatrix} + o_p(1) \\ &:= \Sigma_2 + o_p(1), \end{aligned}$$

$$\text{and } \max_{1 \leq t \leq n} \|\mathbf{Z}_t(0, \beta_0)\| = o_p(n^{1/2}),$$

where $\mathbf{W}_2 \sim N(0, \Sigma_2)$.

(iii) *If $\beta_0 = d_2/\sqrt{n}$ for some $d_2 \in \mathbb{R}$ and γ_0 is a nonzero constant, then*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{Z}_t(\gamma_0, 0) = \mathbf{W}_3 + \lim_{t \rightarrow \infty} \begin{pmatrix} d_2 E\{X_{t-1}(\frac{\alpha_0}{1-\gamma_0} + \sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j)\} \\ d_2 E\left(\frac{X_1^2}{\sqrt{1+X_1^2}}\right) \end{pmatrix} + o_p(1),$$

$$\begin{aligned}
& \frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t(\gamma_0, 0) \mathbf{Z}_t^T(\gamma_0, 0) = \\
& \lim_{t \rightarrow \infty} \left(\begin{array}{cc} E\left(\frac{\alpha_0}{1-\gamma_0} + \sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j\right)^2 & E\left\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} \left(\frac{\alpha_0}{1-\gamma_0} + \sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j\right)\right\} \\ E\left\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} \left(\frac{\alpha_0}{1-\gamma_0} + \sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j\right)\right\} & E\left(\frac{X_1^2}{1+X_1^2}\right) \end{array} \right) \times \\
& E(U_1^2) + o_p(1) \\
& := \Sigma_3 + o_p(1),
\end{aligned}$$

$$\text{and } \max_{1 \leq t \leq n} \|\mathbf{Z}_t(\gamma_0, 0)\| = o_p(n^{1/2}),$$

where $\mathbf{W}_3 \sim N(0, \Sigma_3)$.

Proof. (i) Since

$$Y_t = \alpha_0 \frac{1 - \gamma_0^t}{1 - \gamma_0} + \gamma_0^t Y_0 + \sum_{j=0}^{t-1} \gamma_0^{t-1-j} \beta_0 X_j + \sum_{j=1}^t \gamma_0^{t-j} U_j, \quad (10)$$

we have $Y_t - \alpha_0 - U_t = o_p(1)$ as $t \rightarrow \infty$, which is used to show that

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t1}(0, 0) \\
& = \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t Y_{t-1} + \frac{d_1}{n} \sum_{t=1}^n Y_{t-1} Y_{t-1} + \frac{d_2}{n} \sum_{t=1}^n X_{t-1} Y_{t-1} \\
& = \frac{1}{\sqrt{n}} \sum_{t=3}^n U_t \left\{ \alpha_0 \frac{1 - (d_1/\sqrt{n})^{t-1}}{1 - d_1/\sqrt{n}} + \left(\frac{d_1}{\sqrt{n}}\right)^{t-1} Y_0 + \sum_{j=0}^{t-2} \left(\frac{d_1}{\sqrt{n}}\right)^{t-2-j} \frac{d_2}{\sqrt{n}} X_j \right. \\
& \quad \left. + \sum_{j=1}^{t-2} \left(\frac{d_1}{\sqrt{n}}\right)^{t-1-j} U_j + U_{t-1} \right\} + d_1 \{E(U_1^2) + \alpha_0^2\} + d_2 E\{X_1(\alpha_0 + U_1)\} + o_p(1) \\
& = \frac{1}{\sqrt{n}} \sum_{t=3}^n U_t (\alpha_0 + U_{t-1}) + d_1 \{E(U_1^2) + \alpha_0^2\} + d_2 E\{X_1(\alpha_0 + U_1)\} + o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t2}(0, 0) \\
& = \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \frac{d_1}{n} \sum_{t=1}^n \frac{Y_{t-1} X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \frac{d_2}{n} \sum_{t=1}^n \frac{X_{t-1}^2}{\sqrt{1+X_{t-1}^2}} \\
& = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{U_t X_{t-1}}{\sqrt{1+X_{t-1}^2}} + d_1 E\left(\frac{(\alpha_0 + U_1) X_1}{\sqrt{1+X_1^2}}\right) + d_2 E\left(\frac{X_1^2}{\sqrt{1+X_1^2}}\right) + o_p(1),
\end{aligned}$$

which imply (7). Similarly we can prove (8) and (9).

(ii) By noting that $Y_t - \alpha_0 - \beta_0 X_{t-1} - U_t = o_p(1)$ as $t \rightarrow \infty$, results can be shown in a way similar to the proof of (i).

(iii) By noting that $Y_t - \frac{\alpha_0}{1-\gamma_0} - \sum_{j=1}^t \gamma_0^{t-j} U_j = o_p(1)$ as $t \rightarrow \infty$, results follow from similar arguments in proving (i). \square

Lemma 2. Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$, and $\alpha = \alpha_0$ is known. Further assume $\phi_0 = 1 - \rho/n$ for some $\rho \in \mathbb{R}$ and $\theta_0 = 0$. Put

$$\mathbf{Z}_t^*(\gamma, \beta) = (Z_{t1}(\gamma, \beta), \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} Z_{t2}(\gamma, \beta))^T.$$

(i) If $\gamma_0 = d_1/\sqrt{n}$ for some $d_1 \in \mathbb{R}$ and $\beta_0 = d_2/n$ for some $d_2 \in \mathbb{R}$, then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{Z}_t^*(0, 0) = \tilde{\mathbf{W}}_1 + \begin{pmatrix} d_1 \{E(U_1^2) + \alpha_0^2\} + d_2 \alpha_0 \int_0^1 J_{V,\rho}(s) ds \\ d_1 \alpha_0 + d_2 \int_0^1 J_{V,\rho}(s) ds \end{pmatrix} + o_p(1), \quad (11)$$

$$\frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t^*(0, 0) \mathbf{Z}_t^{*T}(0, 0) = E(U_1^2) \begin{pmatrix} E(U_1^2) + \alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix} + o_p(1) := \tilde{\Sigma}_1 + o_p(1), \quad (12)$$

$$\text{and } \max_{1 \leq t \leq n} \|\mathbf{Z}_t^*(0, 0)\| = o_p(n^{1/2}), \quad (13)$$

where $\tilde{\mathbf{W}}_1 \sim N(0, \tilde{\Sigma}_1)$.

(ii) If $\gamma_0 = d_1/n$ for some $d_1 \in \mathbb{R}$ and β_0 is a nonzero constant, then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{Z}_t^*(0, \beta_0) = \tilde{\mathbf{W}}_2 + \begin{pmatrix} d_1 \alpha_0 \beta_0 \int_0^1 J_{V,\rho}(s) ds \\ d_1 \beta_0 \int_0^1 J_{V,\rho}(s) ds \end{pmatrix} + o_p(1), \quad (14)$$

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t^*(0, \beta_0) \mathbf{Z}_t^{*T}(0, \beta_0) \\ &= E(U_1^2) \begin{pmatrix} E(U_1 - \beta_0 \sum_{j=0}^{\infty} \psi_j V_{1-j})^2 + \alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix} + o_p(1) \\ &:= \tilde{\Sigma}_2 + o_p(1), \end{aligned} \quad (15)$$

$$\text{and } \max_{1 \leq t \leq n} \|\mathbf{Z}_t^*(0, \beta_0)\| = o_p(n^{1/2}), \quad (16)$$

where $\tilde{\mathbf{W}}_2 \sim N(0, \tilde{\Sigma}_2)$.

(iii) If $\beta_0 = d_2/n$ for some $d_2 \in \mathbb{R}$ and γ_0 is a nonzero constant, then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{Z}_t^*(\gamma_0, 0) = \tilde{\mathbf{W}}_3 + \begin{pmatrix} d_2 \frac{\alpha_0}{1-\gamma_0} \int_0^1 J_{V,\rho}(s) ds \\ d_2 \int_0^1 J_{V,\rho}(s) ds \end{pmatrix} + o_p(1), \quad (17)$$

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t^*(\gamma_0, 0) \mathbf{Z}_t^{*T}(\gamma_0, 0) \\ &= E(U_1^2) \lim_{t \rightarrow \infty} \begin{pmatrix} E(\sum_{j=1}^t \gamma_0^{t-j} U_j)^2 + (\frac{\alpha_0}{1-\gamma_0})^2 & \frac{\alpha_0}{1-\gamma_0} \\ \frac{\alpha_0}{1-\gamma_0} & 1 \end{pmatrix} + o_p(1) \\ &:= \tilde{\Sigma}_3 + o_p(1), \end{aligned} \quad (18)$$

$$\text{and } \max_{1 \leq t \leq n} \|\mathbf{Z}_t^*(\gamma_0, 0)\| = o_p(n^{1/2}), \quad (19)$$

where $\tilde{\mathbf{W}}_3 \sim N(0, \tilde{\Sigma}_3)$.

Proof. (i) It follows from Phillips (1987) that

$$\frac{1}{\sqrt{n}}X_{[nr]} \xrightarrow{D} J_{V,\rho}(r) \quad \text{in the space } D[0,1], \quad (20)$$

where $D[0,1]$ is the collection of real-valued functions on $[0,1]$ which are right continuous with left limits; see Billingsley (1999). By (10) and (20), we have $Y_t - \alpha_0 - U_t = o_p(1)$ as $t \rightarrow \infty$. Hence,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t1}(0,0) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t Y_{t-1} + \frac{d_1}{n} \sum_{t=1}^n Y_{t-1} Y_{t-1} + \frac{d_2}{n^{3/2}} \sum_{t=1}^n X_{t-1} Y_{t-1} \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t (\alpha_0 + U_{t-1}) + d_1 \{E(U_1^2) + \alpha_0^2\} + d_2 E(Y_1) \int_0^1 J_{V,\rho}(s) ds + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t (\alpha_0 + U_{t-1}) + d_1 \{E(U_1^2) + \alpha_0^2\} + d_2 \alpha_0 \int_0^1 J_{V,\rho}(s) ds + o_p(1) \end{aligned} \quad (21)$$

and

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t2}(0,0) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \frac{d_1}{n} \sum_{t=1}^n \frac{Y_{t-1} X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \frac{d_2}{n^{3/2}} \sum_{t=1}^n \frac{X_{t-1}^2}{\sqrt{1+X_{t-1}^2}} \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \frac{d_1 \alpha_0}{n} \sum_{t=1}^n \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \frac{d_1}{n} \sum_{t=1}^n U_t \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} \\ & \quad + \frac{d_2}{n^{3/2}} \sum_{t=1}^n \frac{X_{t-1}^2}{\sqrt{1+X_{t-1}^2}} + o_p(1). \end{aligned} \quad (22)$$

Put $S_0 = 0$ and $S_t = \sum_{j=1}^t U_j$ for $t = 1, \dots, n$. Then

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n (S_t - S_{t-1}) \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} \\ &= \frac{1}{\sqrt{n}} S_n \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} + \frac{1}{\sqrt{n}} \sum_{t=1}^n S_t \left\{ \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} - \frac{X_t}{\sqrt{1+X_t^2}} \right\}. \end{aligned} \quad (23)$$

It follows from Taylor expansion that

$$\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} - \frac{X_t}{\sqrt{1+X_t^2}} = (1 + \xi_t^2)^{-3/2} (X_{t-1} - X_t), \quad (24)$$

where ξ_t lies between X_{t-1} and X_t . By (20), we have $|X_{t-1}|/t^a \xrightarrow{p} \infty$, $|X_t|/t^a \xrightarrow{p} \infty$ and $|X_{t-1} - X_t|/t^a \xrightarrow{p} 0$ for any $a \in (0, 1/2)$ as $t \rightarrow \infty$, which imply that

$$|\xi_t|/t^a \xrightarrow{p} \infty \quad \text{for any } a \in (0, 1/2) \quad \text{as } t \rightarrow \infty. \quad (25)$$

It follows from (24) and (25) that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n S_t \left(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} - \frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}} \right) = o_p(1). \quad (26)$$

By (23) and (26), we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n U_t \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} = \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t + o_p(1). \quad (27)$$

Similarly we have

$$\begin{cases} \frac{1}{n} \sum_{t=1}^n \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} = \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} + o_p(1) \\ \frac{1}{n^{3/2}} \sum_{t=1}^n \frac{X_{t-1}^2}{\sqrt{1+X_{t-1}^2}} = \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} \frac{1}{n} \sum_{t=1}^n \frac{X_{t-1}}{n^{1/2}} + o_p(1). \end{cases} \quad (28)$$

Hence, (11) follows from (20), (21), (22), (27) and (28).

(ii) It follows from (10) and (20) that

$$\frac{1}{\sqrt{n}} Y_{[nr]} = \frac{\beta_0}{\sqrt{n}} X_{[nr]} + o_p(1) \xrightarrow{D} \beta_0 J_{V,\rho}(r) \quad \text{in the space } D[0,1]. \quad (29)$$

Hence,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t1}(0, \beta_0) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t (Y_{t-1} - \beta_0 X_{t-1}) + \frac{d_1}{n^{3/2}} \sum_{t=1}^n Y_{t-1} (Y_{t-1} - \beta_0 X_{t-1}) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t (\alpha_0 + U_{t-1} - \beta_0 \sum_{j=0}^{\infty} \psi_j V_{t-1-j}) + d_1 \alpha_0 \beta_0 \int_0^1 J_{V,\rho}(s) ds + o_p(1) \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t2}(0, \beta_0) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \frac{d_1}{n^{3/2}} \sum_{t=1}^n Y_{t-1} \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} \\ &= \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t + \frac{d_1 \beta_0}{n^{3/2}} \sum_{t=1}^n \frac{X_{t-1}^2}{\sqrt{1+X_{t-1}^2}} + o_p(1) \\ &= \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t + \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} d_1 \beta_0 \int_0^1 J_{V,\rho}(s) ds + o_p(1), \end{aligned}$$

which imply (14). Similarly we can prove (15) and (16).

(iii) By noting that $Y_t - \frac{\alpha_0}{1-\gamma_0} - \sum_{j=1}^t \gamma_0^{t-j} U_j = o_p(1)$ as $t \rightarrow \infty$, we have

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t1}(\gamma_0, 0) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t Y_{t-1} + \frac{d_2}{n^{3/2}} \sum_{t=1}^n X_{t-1} Y_{t-1} \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t \left(\frac{\alpha_0}{1-\gamma_0} + \sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j \right) + d_2 \frac{\alpha_0}{1-\gamma_0} \int_0^1 J_{V,\rho}(s) ds + o_p(1) \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t2}(\gamma_0, 0) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \frac{d_2}{n^{3/2}} \sum_{t=1}^n \frac{X_{t-1}^2}{\sqrt{1+X_{t-1}^2}} \\ &= \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} \frac{1}{\sqrt{n}} \sum_{t=1}^n U_t + \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} d_2 \int_0^1 J_{V,\rho}(s) ds + o_p(1), \end{aligned}$$

which imply (17). Similarly we can prove (18) and (19). □

Lemma 3. Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$, and $\alpha = \alpha_0$ is known. Further assume $\phi_0 = 1 - \rho/n$ for some $\rho \in \mathbb{R}$ and $\theta_0 \neq 0$.

(i) If $\gamma_0 = d_1/\sqrt{n}$ for some $d_1 \in \mathbb{R}$ and $\beta_0 = d_2/n^{3/2}$ for some $d_2 \in \mathbb{R}$, then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{Z}_t(0, 0) = \bar{\mathbf{W}}_1 + \begin{pmatrix} d_1 \{E(U_1^2) + \alpha_0^2\} + d_2 \alpha_0 \theta_0 \int_0^1 \frac{1-e^{-\rho s}}{s} ds \\ d_1 \alpha_0 \operatorname{sgn}(\theta_0) + d_2 |\theta_0| \int_0^1 \frac{1-e^{-\rho s}}{\rho} ds \end{pmatrix} + o_p(1), \quad (30)$$

$$\frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t(0, 0) \mathbf{Z}_t^T(0, 0) = E(U_1^2) \begin{pmatrix} E(U_1^2) + \alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix} + o_p(1) := \bar{\Sigma}_1 + o_p(1), \quad (31)$$

$$\text{and } \max_{1 \leq t \leq n} \|\mathbf{Z}_t(0, 0)\| = o_p(n^{1/2}), \quad (32)$$

where $\bar{\mathbf{W}}_1 \sim N(0, \bar{\Sigma}_1)$.

(ii) If $\gamma_0 = d_1/n^{3/2}$ for some $d_1 \in \mathbb{R}$ and β_0 is a nonzero constant, then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{Z}_t(0, \beta_0) = \bar{\mathbf{W}}_2 + \begin{pmatrix} d_1 \beta_0 \theta_0 (\alpha_0 - \beta_0 \theta_0) \int_0^1 \frac{1-e^{-\rho s}}{\rho} ds \\ d_1 \beta_0 |\theta_0| \int_0^1 \frac{1-e^{-\rho s}}{\rho} ds \end{pmatrix} + o_p(1), \quad (33)$$

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t(0, \beta_0) \mathbf{Z}_t^T(0, \beta_0) \\ &= E(U_1^2) \begin{pmatrix} E(\alpha_0 + U_1 - \beta_0 \theta_0 - \beta_0 \sum_{j=0}^{\infty} \psi_j V_{1-j})^2 & (\alpha_0 - \beta_0 \theta_0) \operatorname{sgn}(\theta_0) \\ (\alpha_0 - \beta_0 \theta_0) \operatorname{sgn}(\theta_0) & 1 \end{pmatrix} + o_p(1) \\ &:= \bar{\Sigma}_2 + o_p(1), \end{aligned} \quad (34)$$

$$\text{and } \max_{1 \leq t \leq n} \|\mathbf{Z}_t(0, \beta_0)\| = o_p(n^{1/2}), \quad (35)$$

where $\bar{\mathbf{W}}_2 \sim N(0, \bar{\Sigma}_2)$.

(iii) If $\beta_0 = d_2/n^{3/2}$ for some $d_2 \in \mathbb{R}$ and γ_0 is a nonzero constant, then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{Z}_t(\gamma_0, 0) = \bar{\mathbf{W}}_3 + \begin{pmatrix} d_2 \theta_0 \frac{\alpha_0}{1-\gamma_0} \int_0^1 \frac{1-e^{-\rho s}}{s} ds \\ d_2 |\theta_0| \int_0^1 \frac{1-e^{-\rho s}}{\rho} ds \end{pmatrix} + o_p(1), \quad (36)$$

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t(\gamma_0, 0) \mathbf{Z}_t^T(\gamma_0, 0) \\ &= E(U_1^2) \lim_{t \rightarrow \infty} \begin{pmatrix} E(\sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j)^2 + (\frac{\alpha_0}{1-\gamma_0})^2 & \frac{\alpha_0}{1-\gamma_0} \text{sgn}(\theta_0) \\ \frac{\alpha_0}{1-\gamma_0} \text{sgn}(\theta_0) & 1 \end{pmatrix} + o_p(1) \\ &:= \bar{\Sigma}_3 + o_p(1), \end{aligned} \quad (37)$$

$$\text{and } \max_{1 \leq t \leq n} \|\mathbf{Z}_t(\gamma_0, 0)\| = o_p(n^{1/2}), \quad (38)$$

where $\bar{\mathbf{W}}_3 \sim N(0, \bar{\Sigma}_3)$.

Proof. (i) By noting that $X_{[ns]}/n \xrightarrow{p} \theta_0 \frac{1-e^{-\rho s}}{\rho}$ for $s \in [0, 1]$ and $Y_t - \alpha_0 - U_t = o_p(1)$ as $t \rightarrow \infty$, results follow from the same arguments in proving Lemma 2(i).

(ii) By noting that $Y_{[ns]}/n = \beta_0 X_{[ns]}/n + o_p(1) \xrightarrow{p} \beta_0 \theta_0 \frac{1-e^{-\rho s}}{\rho}$ for $s \in [0, 1]$, results follow from the same arguments in proving Lemma 2(ii).

(iii) By noting that $Y_t - \frac{\alpha_0}{1-\gamma_0} - \sum_{j=1}^t \gamma_0^{t-j} U_j = o_p(1)$ as $t \rightarrow \infty$, results follow from the same arguments in proving Lemma 2(iii). \square

Lemma 4. *Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$, and $\alpha = \alpha_0$ is known. Further assume $|\phi_0| < 1$ independent of n .*

(i) *Under $H_0 : \gamma_0 = 0$, with probability tending to one, $L(0, \beta)$ attains its maximum value at some point β^* in the interior of the ball $|\beta - \beta_0| \leq n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2 + \delta)$ as $n \rightarrow \infty$, and β^* and $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^*(\beta^*)$ satisfy $Q_{1n}(\beta^*, \boldsymbol{\lambda}^*) = 0$ and $Q_{2n}(\beta^*, \boldsymbol{\lambda}^*) = 0$, where*

$$Q_{1n}(\beta, \boldsymbol{\lambda}) := \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{Z}_t(0, \beta)}{1 + \boldsymbol{\lambda}^T \mathbf{Z}_t(0, \beta)}, \text{ and } Q_{2n}(\beta, \boldsymbol{\lambda}) = \frac{1}{n} \sum_{t=1}^n \frac{1}{1 + \boldsymbol{\lambda}^T \mathbf{Z}_t(0, \beta)} \left(\frac{\partial \mathbf{Z}_t(0, \beta)}{\partial \beta} \right)^T \boldsymbol{\lambda}.$$

(ii) *Under $H_0 : \beta_0 = 0$, with probability tending to one, $L(\gamma, 0)$ attains its maximum value at some point γ^* in the interior of the ball $|\gamma - \gamma_0| \leq n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2 + \delta)$ as $n \rightarrow \infty$,*

and γ^* and $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^*(\gamma^*)$ satisfy $Q_{3n}(\gamma^*, \boldsymbol{\lambda}^*) = 0$ and $Q_{4n}(\gamma^*, \boldsymbol{\lambda}^*) = 0$, where

$$Q_{3n}(\gamma, \boldsymbol{\lambda}) := \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{Z}_t(\gamma, 0)}{1 + \boldsymbol{\lambda}^T \mathbf{Z}_t(\gamma, 0)}, \text{ and } Q_{4n}(\gamma, \boldsymbol{\lambda}) = \frac{1}{n} \sum_{t=1}^n \frac{1}{1 + \boldsymbol{\lambda}^T \mathbf{Z}_t(\gamma, 0)} \left(\frac{\partial \mathbf{Z}_t(\gamma, 0)}{\partial \gamma} \right)^T \boldsymbol{\lambda}.$$

Proof. Using Lemma 1, this lemma follows from the arguments in the proof of Lemma 1 of Qin and Lawless (1994). \square

Lemma 5. *Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$, and $\alpha = \alpha_0$ is known. Further assume $\phi_0 = 1 - \rho/n$ for some $\rho \in \mathbb{R}$ and $\theta_0 = 0$.*

(i) *Put $\bar{\beta} = \beta\sqrt{n}$, $\bar{\beta}_0 = \beta_0\sqrt{n}$, $\bar{\mathbf{Z}}_t^*(\gamma, \bar{\beta}) = \mathbf{Z}_t^*(\gamma, \beta)$ defined in Lemma 2 and $\bar{L}(\gamma, \bar{\beta}) = L(\gamma, \beta)$. Under $H_0 : \gamma_0 = 0$, with probability tending to one, $\bar{L}(0, \bar{\beta})$ attains its maximum value at some point $\bar{\beta}^*$ in the interior of the ball $|\bar{\beta} - \bar{\beta}_0| \leq n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2 + \delta)$ as $n \rightarrow \infty$, and $\bar{\beta}^*$ and $\bar{\boldsymbol{\lambda}}^* = \bar{\boldsymbol{\lambda}}^*(\bar{\beta}^*)$ satisfy $\tilde{Q}_{1n}(\bar{\beta}^*, \bar{\boldsymbol{\lambda}}^*) = 0$ and $\tilde{Q}_{2n}(\bar{\beta}^*, \bar{\boldsymbol{\lambda}}^*) = 0$, where*

$$\tilde{Q}_{1n}(\bar{\beta}, \bar{\boldsymbol{\lambda}}) := \frac{1}{n} \sum_{t=1}^n \frac{\bar{\mathbf{Z}}_t^*(0, \bar{\beta})}{1 + \bar{\boldsymbol{\lambda}}^T \bar{\mathbf{Z}}_t^*(0, \bar{\beta})}, \text{ and } \tilde{Q}_{2n}(\bar{\beta}, \bar{\boldsymbol{\lambda}}) = \frac{1}{n} \sum_{t=1}^n \frac{1}{1 + \bar{\boldsymbol{\lambda}}^T \bar{\mathbf{Z}}_t^*(0, \bar{\beta})} \left(\frac{\partial \bar{\mathbf{Z}}_t^*(0, \bar{\beta})}{\partial \bar{\beta}} \right)^T \bar{\boldsymbol{\lambda}}.$$

(ii) *Under $H_0 : \beta_0 = 0$, with probability tending to one, $L(\gamma, 0)$ attains its maximum value at some point γ^* in the interior of the ball $|\gamma - \gamma_0| \leq n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2 + \delta)$ as $n \rightarrow \infty$, and γ^* and $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^*(\gamma^*)$ satisfy $\tilde{Q}_{3n}(\gamma^*, \boldsymbol{\lambda}^*) = 0$ and $\tilde{Q}_{4n}(\gamma^*, \boldsymbol{\lambda}^*) = 0$, where*

$$\tilde{Q}_{3n}(\gamma, \boldsymbol{\lambda}) := \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{Z}_t^*(\gamma, 0)}{1 + \boldsymbol{\lambda}^T \mathbf{Z}_t^*(\gamma, 0)}, \text{ and } \tilde{Q}_{4n}(\gamma, \boldsymbol{\lambda}) = \frac{1}{n} \sum_{t=1}^n \frac{1}{1 + \boldsymbol{\lambda}^T \mathbf{Z}_t^*(\gamma, 0)} \left(\frac{\partial \mathbf{Z}_t^*(\gamma, 0)}{\partial \gamma} \right)^T \boldsymbol{\lambda}.$$

Proof. Using Lemma 2, this lemma follows from the arguments in the proof of Lemma 1 of Qin and Lawless (1994). \square

Lemma 6. *Suppose model (3) holds with $|\gamma_0| < 1$ and $E\{|U_t|^{2+\delta} + |V_t|^{2+\delta}\} < \infty$ for some $\delta > 0$, and $\alpha = \alpha_0$ is known. Further assume $\phi_0 = 1 - \rho/n$ for some $\rho \in \mathbb{R}$ and $\theta_0 \neq 0$.*

(i) *Put $\bar{\beta} = \beta n$, $\bar{\beta}_0 = \beta_0 n$, $\bar{\mathbf{Z}}_t(\gamma, \bar{\beta}) = \mathbf{Z}_t(\gamma, \beta)$ and $\bar{L}(\gamma, \bar{\beta}) = L(\gamma, \beta)$. Under $H_0 : \gamma_0 = 0$, with probability tending to one, $\bar{L}(0, \bar{\beta})$ attains its maximum value at some point $\bar{\beta}^*$ in the interior of the ball $|\bar{\beta} - \bar{\beta}_0| \leq n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2 + \delta)$ as $n \rightarrow \infty$, and $\bar{\beta}^*$ and $\bar{\boldsymbol{\lambda}}^* = \bar{\boldsymbol{\lambda}}^*(\bar{\beta}^*)$ satisfy $\bar{Q}_{1n}(\bar{\beta}^*, \bar{\boldsymbol{\lambda}}^*) = 0$ and $\bar{Q}_{2n}(\bar{\beta}^*, \bar{\boldsymbol{\lambda}}^*) = 0$, where*

$$\bar{Q}_{1n}(\bar{\beta}, \bar{\boldsymbol{\lambda}}) := \frac{1}{n} \sum_{t=1}^n \frac{\bar{\mathbf{Z}}_t(0, \bar{\beta})}{1 + \bar{\boldsymbol{\lambda}}^T \bar{\mathbf{Z}}_t(0, \bar{\beta})}, \text{ and } \bar{Q}_{2n}(\bar{\beta}, \bar{\boldsymbol{\lambda}}) = \frac{1}{n} \sum_{t=1}^n \frac{1}{1 + \bar{\boldsymbol{\lambda}}^T \bar{\mathbf{Z}}_t(0, \bar{\beta})} \left(\frac{\partial \bar{\mathbf{Z}}_t(0, \bar{\beta})}{\partial \bar{\beta}} \right)^T \bar{\boldsymbol{\lambda}}.$$

(ii) Under $H_0 : \beta_0 = 0$, with probability tending to one, $L(\gamma, 0)$ attains its maximum value at some point γ^* in the interior of the ball $|\gamma - \gamma_0| \leq n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2 + \delta)$ as $n \rightarrow \infty$, and γ^* and $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^*(\gamma^*)$ satisfy $\bar{Q}_{3n}(\gamma^*, \boldsymbol{\lambda}^*) = 0$ and $\bar{Q}_{4n}(\gamma^*, \boldsymbol{\lambda}^*) = 0$, where

$$\bar{Q}_{3n}(\gamma, \boldsymbol{\lambda}) := \frac{1}{n} \sum_{t=1}^n \frac{\mathbf{Z}_t(\gamma, 0)}{1 + \boldsymbol{\lambda}^T \mathbf{Z}_t(\gamma, 0)}, \text{ and } \bar{Q}_{4n}(\gamma, \boldsymbol{\lambda}) = \frac{1}{n} \sum_{t=1}^n \frac{1}{1 + \boldsymbol{\lambda}^T \mathbf{Z}_t(\gamma, 0)} \left(\frac{\partial \mathbf{Z}_t(\gamma, 0)}{\partial \gamma} \right)^T \boldsymbol{\lambda}.$$

Proof. Using Lemma 3, this lemma follows from the arguments in the proof of Lemma 1 of Qin and Lawless (1994). \square

Proof of Theorem 1. Case A1) Assume $\phi_0 = 1 - \rho/n$, $\theta_0 = 0$ and $H_0 : \gamma_0 = 0$ & $\beta_0 = 0$. Then it follows from Lemma 2(i) with $d_1 = d_2 = 0$ and standard arguments in empirical likelihood method (see Owen (2001)) that

$$-2 \log L(0, 0) = \tilde{\mathbf{W}}_1^T \tilde{\Sigma}_1^{-1} \tilde{\mathbf{W}}_1 + o_p(1) \xrightarrow{d} \chi^2(2) \text{ as } n \rightarrow \infty.$$

Case A2) Assume $\phi_0 = 1 - \rho/n$, $\theta_0 = 0$ and $H_0 : \gamma_0 = 0$. Using notations in Lemma 5(i), it follows from (20) that

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n \frac{\partial \bar{Z}_{t1}(0, \bar{\beta}_0)}{\partial \beta} \\ &= -\frac{1}{n} \sum_{t=1}^n \frac{X_{t-1}}{\sqrt{n}} (Y_{t-1} - \beta_0 X_{t-1}) - \frac{1}{n} \sum_{t=1}^n U_t \frac{X_{t-1}}{\sqrt{n}} \\ &= -\frac{1}{n} \sum_{t=1}^n \frac{X_{t-1}}{\sqrt{n}} \{ \alpha_0 + U_{t-1} - \beta_0 \sum_{j=0}^{\infty} \psi_j V_{t-1-j} \} + o_p(1) \\ &= -\alpha_0 \int_0^1 J_{V, \rho}(s) ds + o_p(1) \end{aligned}$$

and

$$\frac{1}{n} \sum_{t=1}^n \frac{\partial \bar{Z}_{t2}^*(0, \bar{\beta}_0)}{\partial \beta} = -\frac{X_{n-1}}{\sqrt{1 + X_{n-1}^2}} \frac{1}{n} \sum_{t=1}^n \frac{X_{t-1}}{\sqrt{n}} \frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} = -\int_0^1 J_{v, \rho}(s) ds + o_p(1),$$

which imply that

$$\frac{\partial \tilde{Q}_{1n}(\bar{\beta}_0, 0)}{\partial \bar{\beta}} = \begin{pmatrix} -\alpha_0 \int_0^1 J_{V, \rho}(s) ds \\ -\int_0^1 J_{v, \rho}(s) ds \end{pmatrix} + o_p(1) =: \tilde{S}_2^* + o_p(1). \quad (39)$$

By Lemma 2(ii) with $d_1 = 0$, we can show that

$$\left\{ \begin{array}{l} \frac{\partial \tilde{Q}_{1n}(\bar{\beta}_0, 0)}{\partial \boldsymbol{\lambda}} = -\tilde{\Sigma}_2 + o_p(1), \\ \tilde{Q}_{1n}(\bar{\beta}_0, 0) = O_p(n^{-1/2}), \quad \frac{\partial \tilde{Q}_{1n}(\bar{\beta}_0, 0)}{\partial \bar{\beta}} = O_p(1), \\ \frac{\partial \tilde{Q}_{2n}(\bar{\beta}_0, 0)}{\partial \bar{\beta}} = 0, \quad \frac{\partial \tilde{Q}_{2n}(\bar{\beta}_0, 0)}{\partial \boldsymbol{\lambda}} = \tilde{S}_2^* + o_p(1) = O_p(1), \end{array} \right. \quad (40)$$

where $\tilde{\Sigma}_2$ is defined in Lemma 2(ii). By (39) and (40), expanding $\tilde{Q}_{1n}(\bar{\beta}^*, \bar{\boldsymbol{\lambda}}^*)$ and $\tilde{Q}_{2n}(\bar{\beta}^*, \bar{\boldsymbol{\lambda}}^*)$ around $(\bar{\beta}_0, 0)^T$ yields

$$\begin{aligned} 0 &= \tilde{Q}_{1n}(\bar{\beta}_0, 0) + \frac{\partial \tilde{Q}_{1n}(\bar{\beta}_0, 0)}{\partial \bar{\beta}}(\bar{\beta}^* - \bar{\beta}_0) + \frac{\partial \tilde{Q}_{1n}(\bar{\beta}_0, 0)}{\partial \bar{\boldsymbol{\lambda}}^T} \bar{\boldsymbol{\lambda}}^* + o_p(\|\bar{\boldsymbol{\lambda}}^*\| + |\bar{\beta}^* - \bar{\beta}_0|) \\ &= \tilde{Q}_{1n}(\bar{\beta}_0, 0) + \tilde{S}_2^*(\bar{\beta}^* - \bar{\beta}_0) - \tilde{\Sigma}_2 \bar{\boldsymbol{\lambda}}^* + o_p(\|\bar{\boldsymbol{\lambda}}^*\| + |\bar{\beta}^* - \bar{\beta}_0|) \end{aligned}$$

and

$$\begin{aligned} 0 &= \tilde{Q}_{2n}(\bar{\beta}_0, 0) + \frac{\partial \tilde{Q}_{2n}(\bar{\beta}_0, 0)}{\partial \bar{\beta}}(\bar{\beta}^* - \bar{\beta}_0) + \frac{\partial \tilde{Q}_{2n}(\bar{\beta}_0, 0)}{\partial \bar{\boldsymbol{\lambda}}^T} \bar{\boldsymbol{\lambda}}^* + o_p(\|\bar{\boldsymbol{\lambda}}^*\| + |\bar{\beta}^* - \bar{\beta}_0|) \\ &= \tilde{S}_2^{*T} \bar{\boldsymbol{\lambda}}^* + o_p(\|\bar{\boldsymbol{\lambda}}^*\| + |\bar{\beta}^* - \bar{\beta}_0|), \end{aligned}$$

which imply that

$$\tilde{S}_2^{*T} \tilde{\Sigma}_2^{-1} \tilde{S}_2^* \sqrt{n}(\bar{\beta}^* - \bar{\beta}_0) = -\tilde{S}_2^{*T} \tilde{\Sigma}_2^{-1} \sqrt{n} \tilde{Q}_{1n}(\bar{\beta}_0, 0) + o_p(1)$$

and

$$\begin{aligned} \sqrt{n} \bar{\boldsymbol{\lambda}}^* &= \tilde{\Sigma}_2^{-1} \sqrt{n} \tilde{Q}_{1n}(\bar{\beta}_0, 0) + \tilde{\Sigma}_2^{-1} \tilde{S}_2^* \sqrt{n}(\bar{\beta}^* - \bar{\beta}_0) + o_p(1) \\ &= \left\{ \tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1} \tilde{S}_2^* \tilde{S}_2^{*T} \tilde{\Sigma}_2^{-1}}{\tilde{S}_2^{*T} \tilde{\Sigma}_2^{-1} \tilde{S}_2^*} \right\} \sqrt{n} \tilde{Q}_{1n}(\bar{\beta}_0, 0) + o_p(1) \\ &= \left\{ \tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1} \tilde{S}_2 \tilde{S}_2^T \tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2} \right\} \sqrt{n} \tilde{Q}_{1n}(\bar{\beta}_0, 0) + o_p(1), \end{aligned} \quad (41)$$

where $\tilde{S}_2 = -(\alpha_0, 1)^T$. It follows from (41) and Taylor expansion that

$$\begin{aligned} &-2 \log L^{P1}(0) \\ &= -2 \log \bar{L}(0, \bar{\beta}^*) \\ &= 2 \sum_{t=1}^n \bar{\boldsymbol{\lambda}}^{*T} \bar{\mathbf{Z}}_t(0, \bar{\beta}^*) - \sum_{t=1}^n \bar{\boldsymbol{\lambda}}^{*T} \bar{\mathbf{Z}}_t(0, \bar{\beta}^*) \bar{\mathbf{Z}}_t^T(0, \bar{\beta}^*) \bar{\boldsymbol{\lambda}}^* + o_p(1) \\ &= 2n \bar{\boldsymbol{\lambda}}^{*T} \tilde{Q}_{1n}(\bar{\beta}_0, 0) + 2n \bar{\boldsymbol{\lambda}}^{*T} \frac{\partial \tilde{Q}_{1n}(\bar{\beta}_0, 0)}{\partial \bar{\beta}}(\bar{\beta}^* - \bar{\beta}_0) - n \bar{\boldsymbol{\lambda}}^{*T} \tilde{\Sigma}_2 \bar{\boldsymbol{\lambda}}^* + o_p(1) \\ &= 2\sqrt{n} \bar{\boldsymbol{\lambda}}^{*T} \left\{ \sqrt{n} \tilde{Q}_{1n}(\bar{\beta}_0, 0) + \tilde{S}_2 \sqrt{n}(\bar{\beta}^* - \bar{\beta}_0) \right\} - n \bar{\boldsymbol{\lambda}}^{*T} \tilde{\Sigma}_2 \bar{\boldsymbol{\lambda}}^* + o_p(1) \\ &= 2\sqrt{n} \bar{\boldsymbol{\lambda}}^{*T} \tilde{\Sigma}_2 \sqrt{n} \bar{\boldsymbol{\lambda}}^* - n \bar{\boldsymbol{\lambda}}^{*T} \tilde{\Sigma}_2 \bar{\boldsymbol{\lambda}}^* + o_p(1) \\ &= \left\{ \sqrt{n} \tilde{Q}_{1n}(\bar{\beta}_0, 0) \right\}^T \left\{ \tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1} \tilde{S}_2 \tilde{S}_2^T \tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2} \right\} \tilde{\Sigma}_2 \left\{ \tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1} \tilde{S}_2 \tilde{S}_2^T \tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2} \right\} \left\{ \sqrt{n} \tilde{Q}_{1n}(\bar{\beta}_0, 0) \right\} + o_p(1) \\ &= (\tilde{\Sigma}_2^{-1/2} \tilde{\mathbf{W}}_2)^T \left\{ I_{2 \times 2} - \frac{\tilde{\Sigma}_2^{-1/2} \tilde{S}_2 \tilde{S}_2^T \tilde{\Sigma}_2^{-1/2}}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2} \right\} (\tilde{\Sigma}_2^{-1/2} \tilde{\mathbf{W}}_2) + o_p(1), \end{aligned} \quad (42)$$

where $I_{2 \times 2}$ denotes the 2×2 identity matrix. Since the matrix $I_{2 \times 2} - \frac{\tilde{\Sigma}_2^{-1/2} \tilde{S}_2 \tilde{S}_2^T \tilde{\Sigma}_2^{-1/2}}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2}$ is idempotent and

$$\text{rank}(I_{2 \times 2} - \frac{\tilde{\Sigma}_2^{-1/2} \tilde{S}_2 \tilde{S}_2^T \tilde{\Sigma}_2^{-1/2}}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2}) = 2 - \text{trace}(\frac{\tilde{\Sigma}_2^{-1/2} \tilde{S}_2 \tilde{S}_2^T \tilde{\Sigma}_2^{-1/2}}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2}) = 2 - \text{trace}(\frac{\tilde{S}_2^T \tilde{\Sigma}_2^{-1/2} \tilde{\Sigma}_2^{-1/2} \tilde{S}_2}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2}) = 1,$$

it follows from Lemma 2(ii) with $d_1 = 0$ that $-2 \log L^{P1}(0) \xrightarrow{d} \chi^2(1)$ as $n \rightarrow \infty$.

Case A3) Assume $\phi_0 = 1 - \rho/n$, $\theta_0 = 0$ and $H_0 : \beta_0 = 0$. Like the proof for the case A2), it follows from Lemmas 2(iii) with $d_2 = 0$ and 5(ii) that

$$-2 \log L^{P2}(0) = (\tilde{\Sigma}_3^{-1/2} \tilde{\mathbf{W}}_3)^T \{I_{2 \times 2} - \frac{\tilde{\Sigma}_3^{-1/2} \tilde{S}_3 \tilde{S}_3^T \tilde{\Sigma}_3^{-1/2}}{\tilde{S}_3^T \tilde{\Sigma}_3^{-1} \tilde{S}_3}\} (\tilde{\Sigma}_3^{-1/2} \tilde{\mathbf{W}}_3) + o_p(1),$$

where $\tilde{\Sigma}_3$ and $\tilde{\mathbf{W}}_3$ are defined in Lemma 2(iii), and

$$\tilde{S}_3 = \lim_{n \rightarrow \infty} \frac{\partial \tilde{Q}_{3n}(\gamma_0, 0)}{\partial \gamma} = - \left(\lim_{t \rightarrow \infty} E(\sum_{j=1}^t \gamma_0^{t-j} U_j)^2 + (\frac{\alpha_0}{1 - \gamma_0})^2, \quad \frac{\alpha_0}{1 - \gamma_0} \right)^T.$$

Since the matrix $I_{2 \times 2} - \frac{\tilde{\Sigma}_3^{-1/2} \tilde{S}_3 \tilde{S}_3^T \tilde{\Sigma}_3^{-1/2}}{\tilde{S}_3^T \tilde{\Sigma}_3^{-1} \tilde{S}_3}$ is idempotent with rank one, it follows from Lemma 2(iii) with $d_2 = 0$ that $-2 \log L^{P2}(0) \xrightarrow{d} \chi^2(1)$ as $n \rightarrow \infty$.

Therefore it follows from the above cases A1)–A3) that Theorem 1 holds for the case of $\phi_0 = 1 - \rho/n$ and $\theta_0 = 0$. Similarly we can show Theorem 1 holds for the case of $|\phi_0| < 1$ by using Lemmas 1 and 4, and for the case of $\phi_0 = 1 - \rho/n$ and $\theta_0 \neq 0$ by using Lemmas 3 and 6. \square

Proofs of Theorems 2–5. They can be shown in the same way as the proof of Theorem 1 by using Lemmas 1–6. For computing the non-central parameters in Theorems 3(ii) and 4(ii), we use the facts that

$$\tilde{\mathbf{D}}_2^T \{ \tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1} \tilde{S}_2 \tilde{S}_2^T \tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2} \} \tilde{\mathbf{D}}_2 = d_1^2 \beta_0^2 \left(\int_0^1 J_{V,\rho}(s) ds \right)^2 \tilde{S}_2^T \{ \tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1} \tilde{S}_2 \tilde{S}_2^T \tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T \tilde{\Sigma}_2^{-1} \tilde{S}_2} \} \tilde{S}_2 = 0$$

and

$$\bar{\mathbf{D}}_2^T \{ \bar{\Sigma}_2^{-1} - \frac{\bar{\Sigma}_2^{-1} \bar{S}_2 \bar{S}_2^T \bar{\Sigma}_2^{-1}}{\bar{S}_2^T \bar{\Sigma}_2^{-1} \bar{S}_2} \} \bar{\mathbf{D}}_2 = d_1^2 \beta_0^2 \bar{S}_2^T \{ \bar{\Sigma}_2^{-1} - \frac{\bar{\Sigma}_2^{-1} \bar{S}_2 \bar{S}_2^T \bar{\Sigma}_2^{-1}}{\bar{S}_2^T \bar{\Sigma}_2^{-1} \bar{S}_2} \} \bar{S}_2 = 0.$$

\square

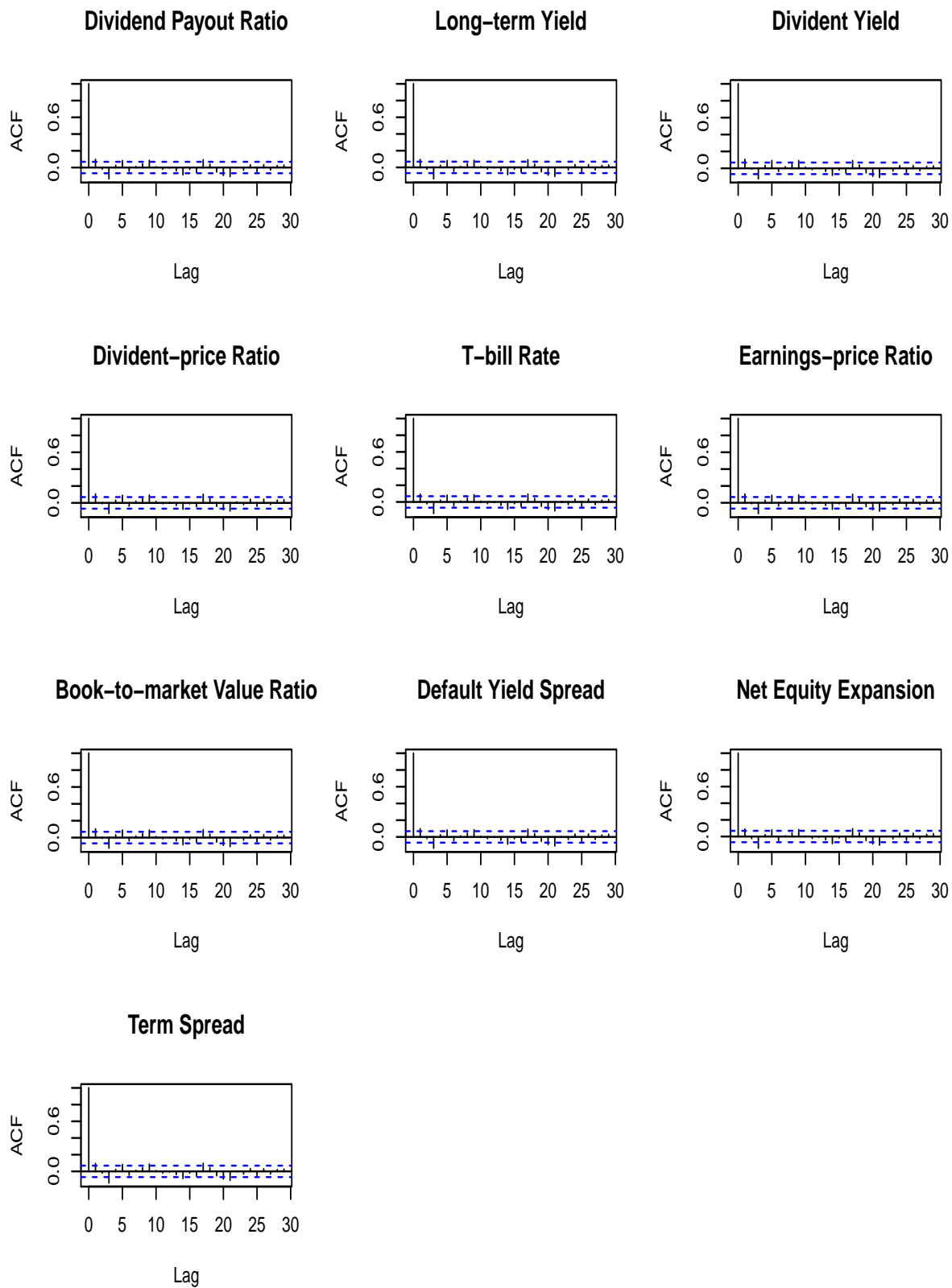


Figure 1: Autocorrelation function of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}X_{t-1}$ for the period 1/1927–12/1994 and Y_t being the CRSP value-weighted excess return.

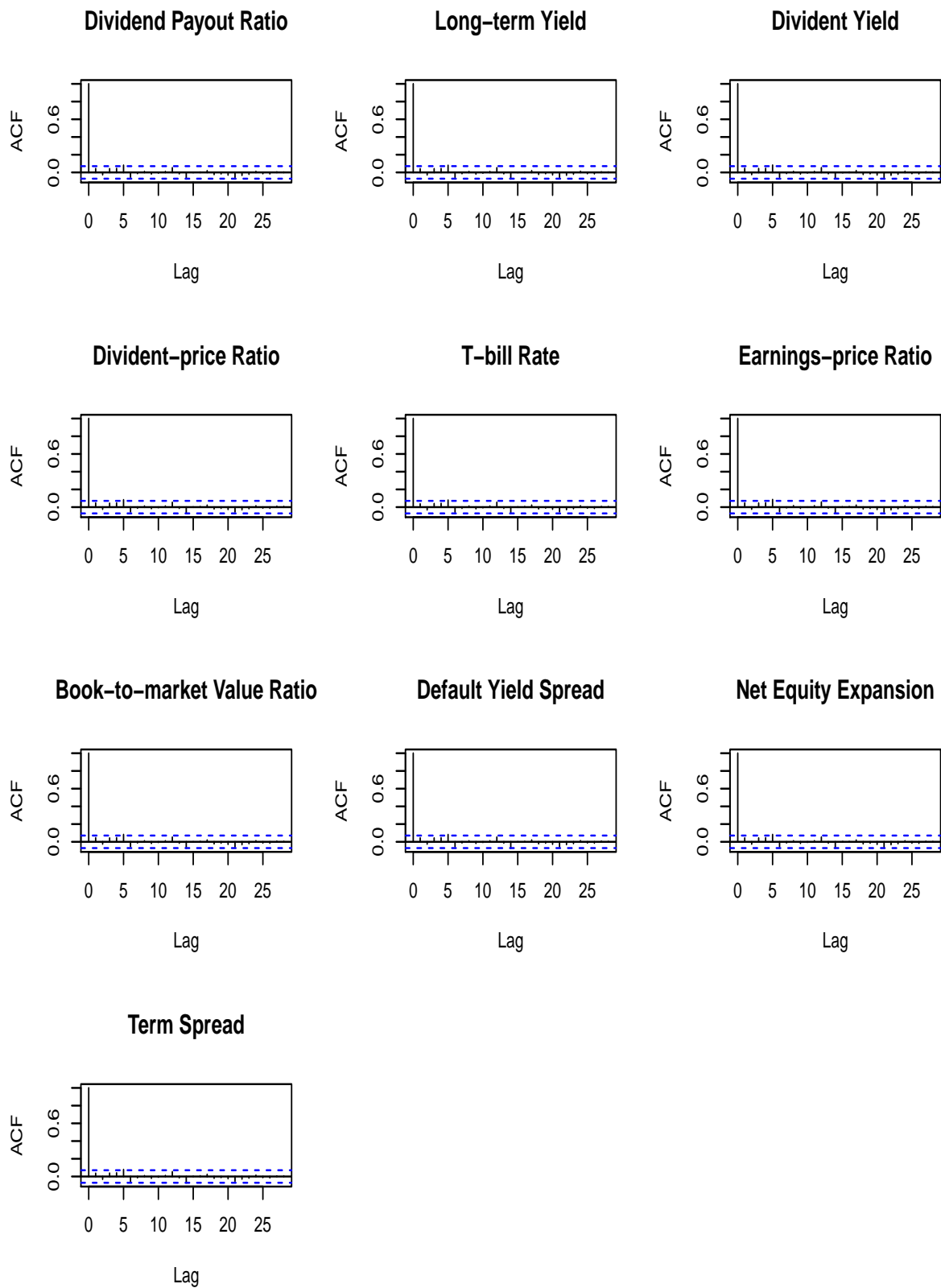


Figure 2: Autocorrelation function of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}X_{t-1}$ for the period 1/1952–12/2015 and Y_t being the CRSP value-weighted excess return.

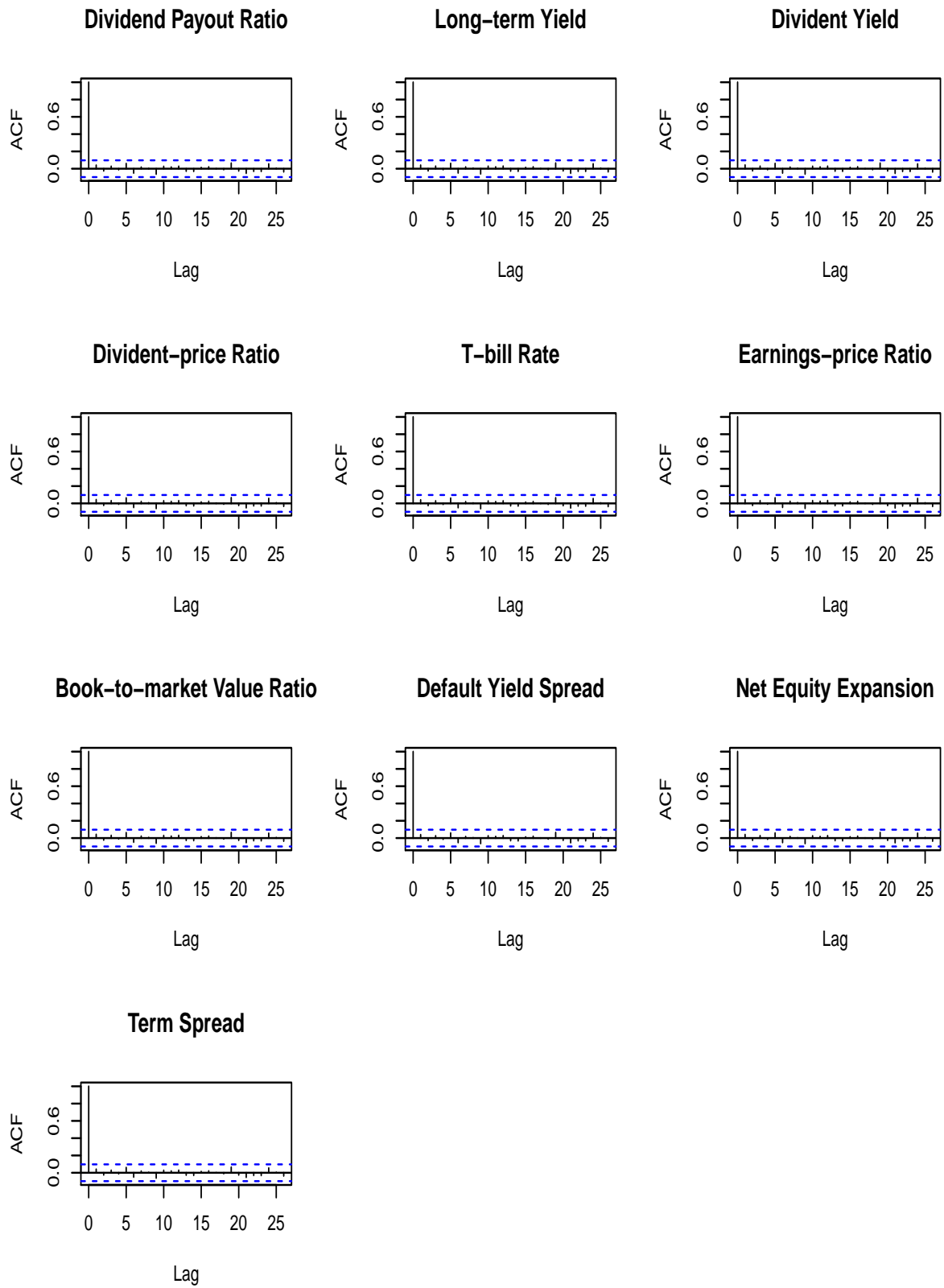


Figure 3: Autocorrelation function of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}X_{t-1}$ for the period 1/1982–12/2015 and Y_t being the CRSP value-weighted excess return.

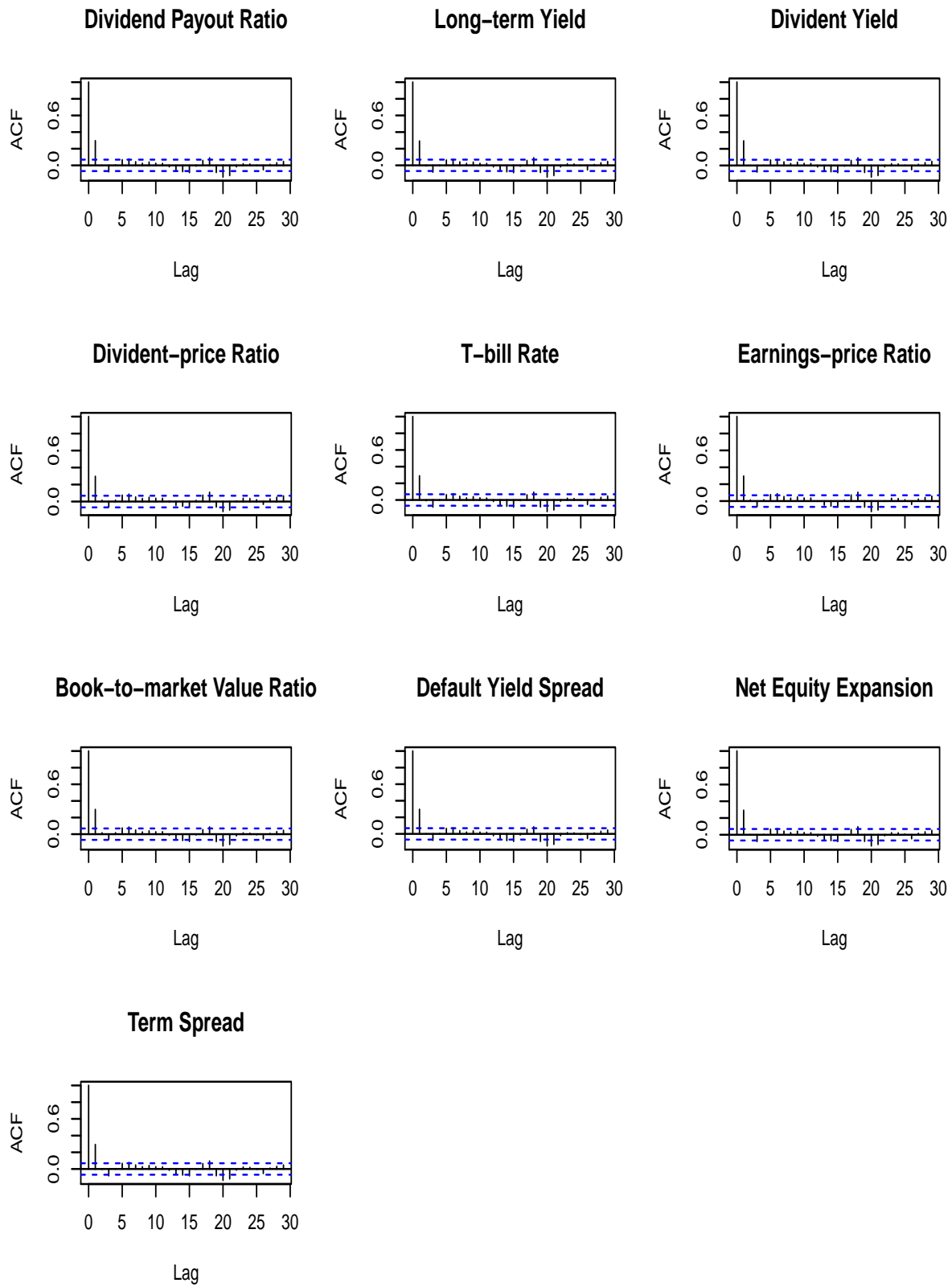


Figure 4: Autocorrelation function of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}X_{t-1}$ for the period 1/1927–12/1994 and Y_t being the S&P 500 excess return.

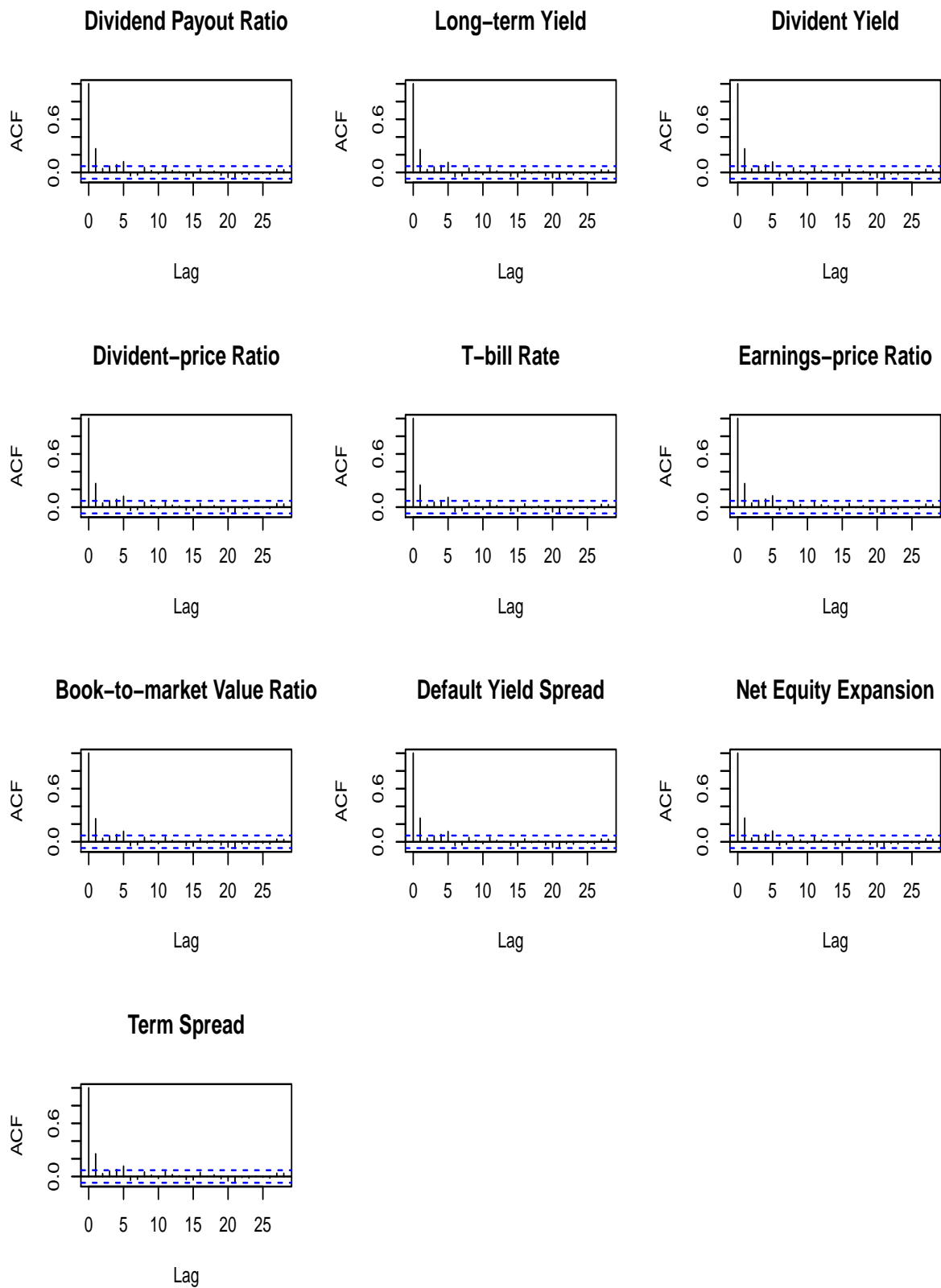


Figure 5: Autocorrelation function of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}X_{t-1}$ for the period 1/1952–12/2015 and Y_t being the S&P500 excess return.

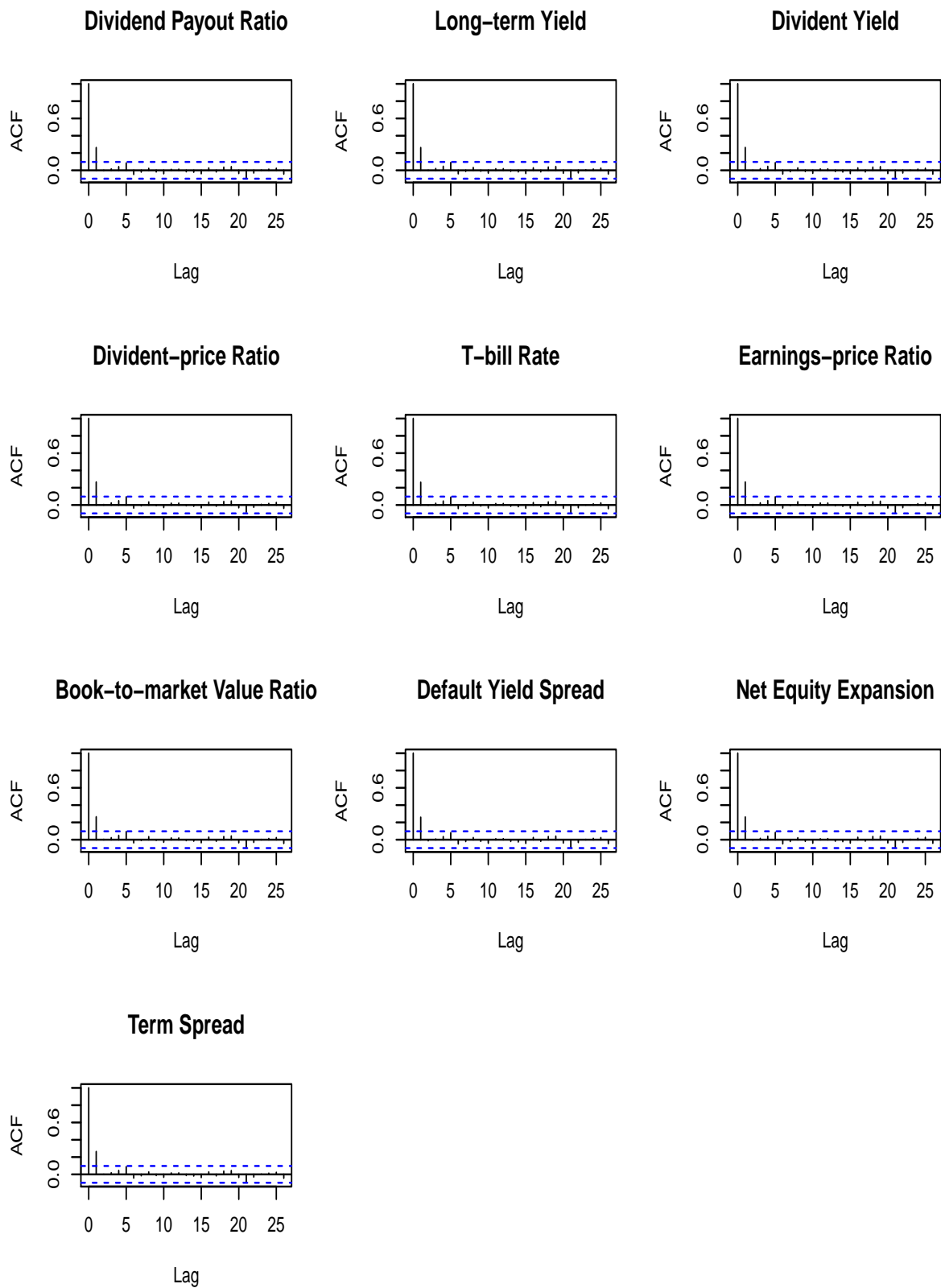


Figure 6: Autocorrelation function of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}X_{t-1}$ for the period 1/1982–12/2015 and Y_t being the S&P 500 excess return.

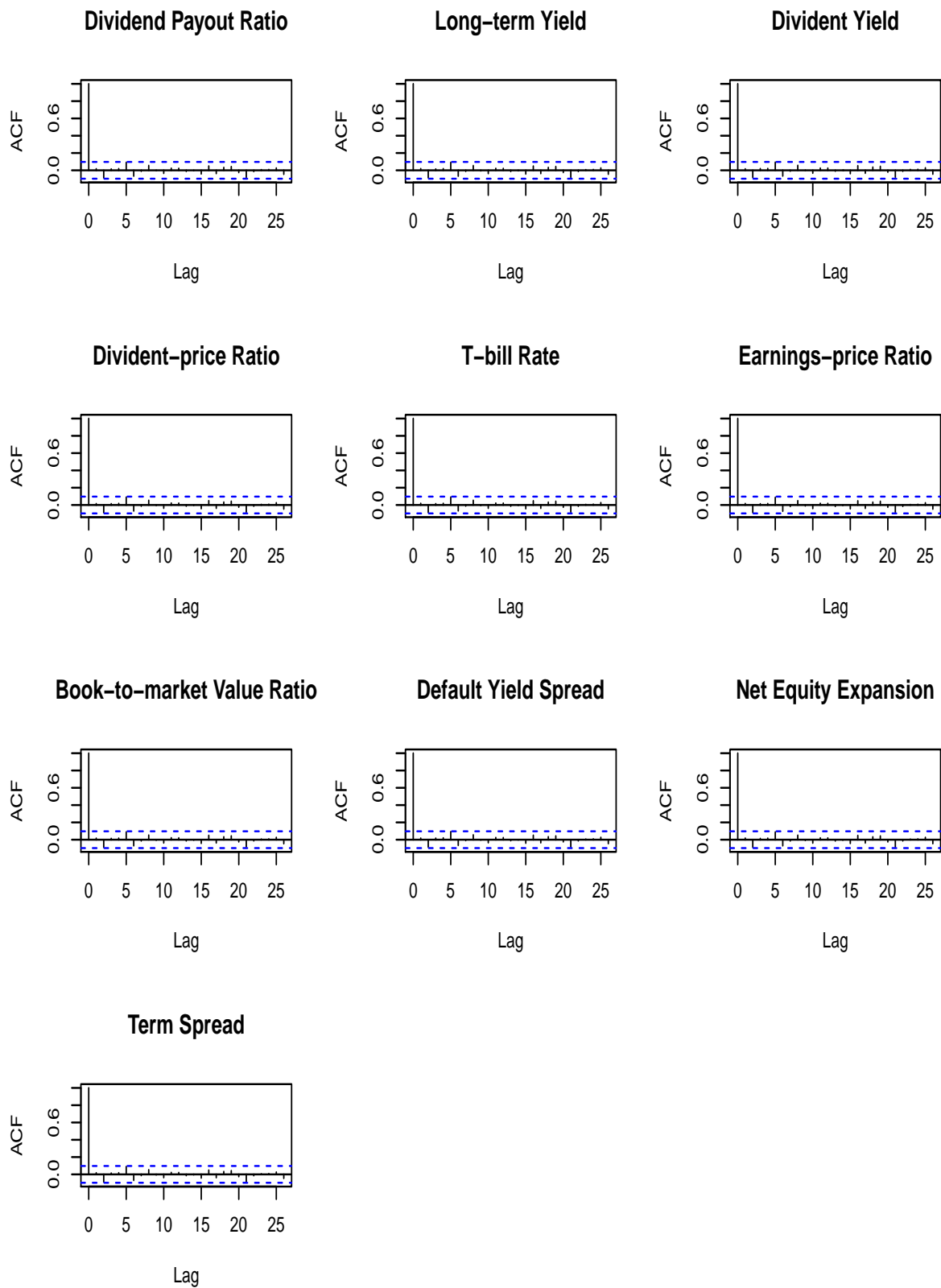


Figure 7: Autocorrelation function of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\gamma}Y_{t-1} - \hat{\beta}X_{t-1}$ for the period 1/1982–12/2015 and Y_t being the S&P 500 excess return.

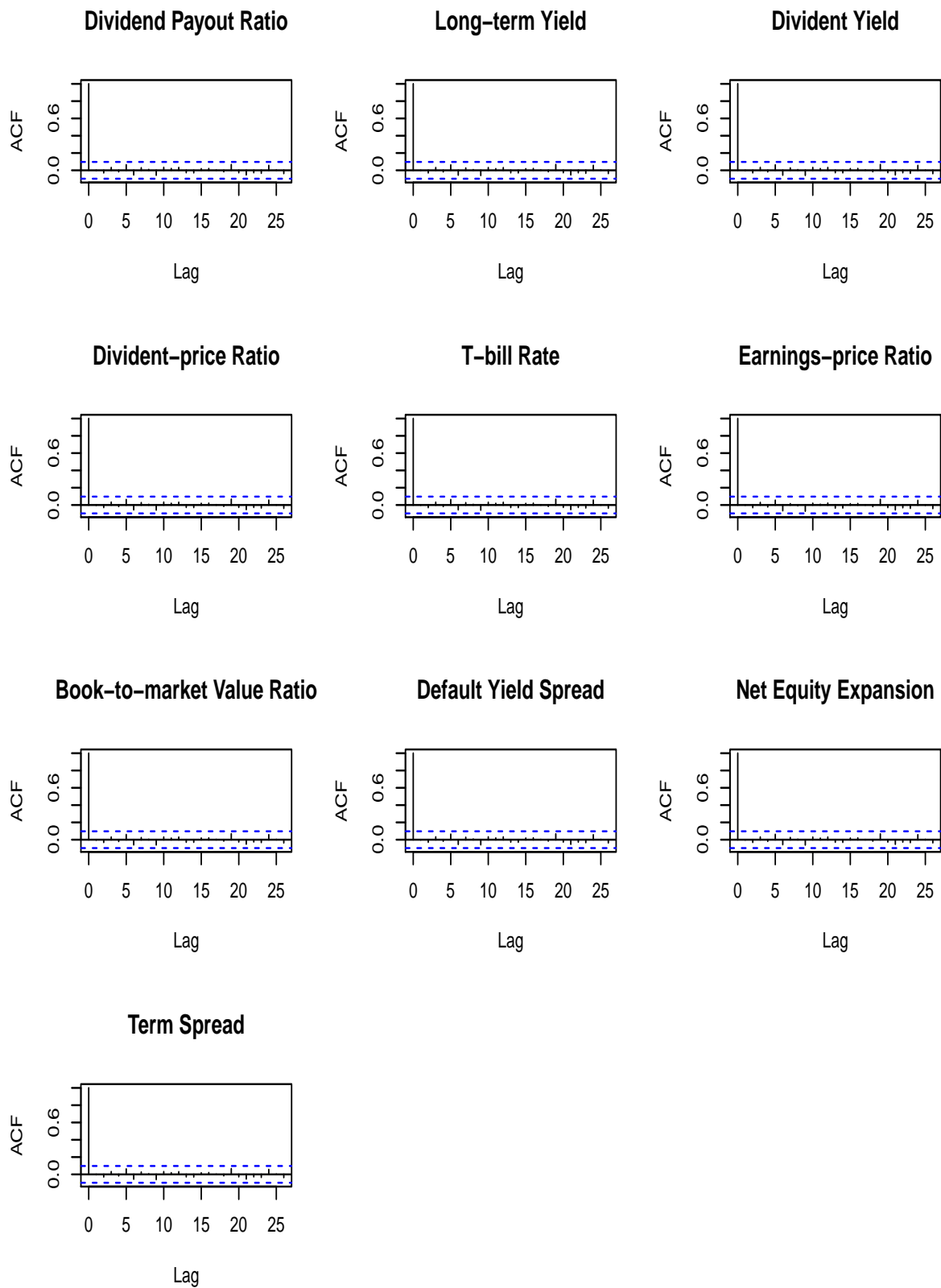


Figure 8: Autocorrelation function of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\gamma}Y_{t-1} - \hat{\beta}X_{t-1}$ for the period 1/1982–12/2015 and Y_t being the CRSP value-weighted excess return.

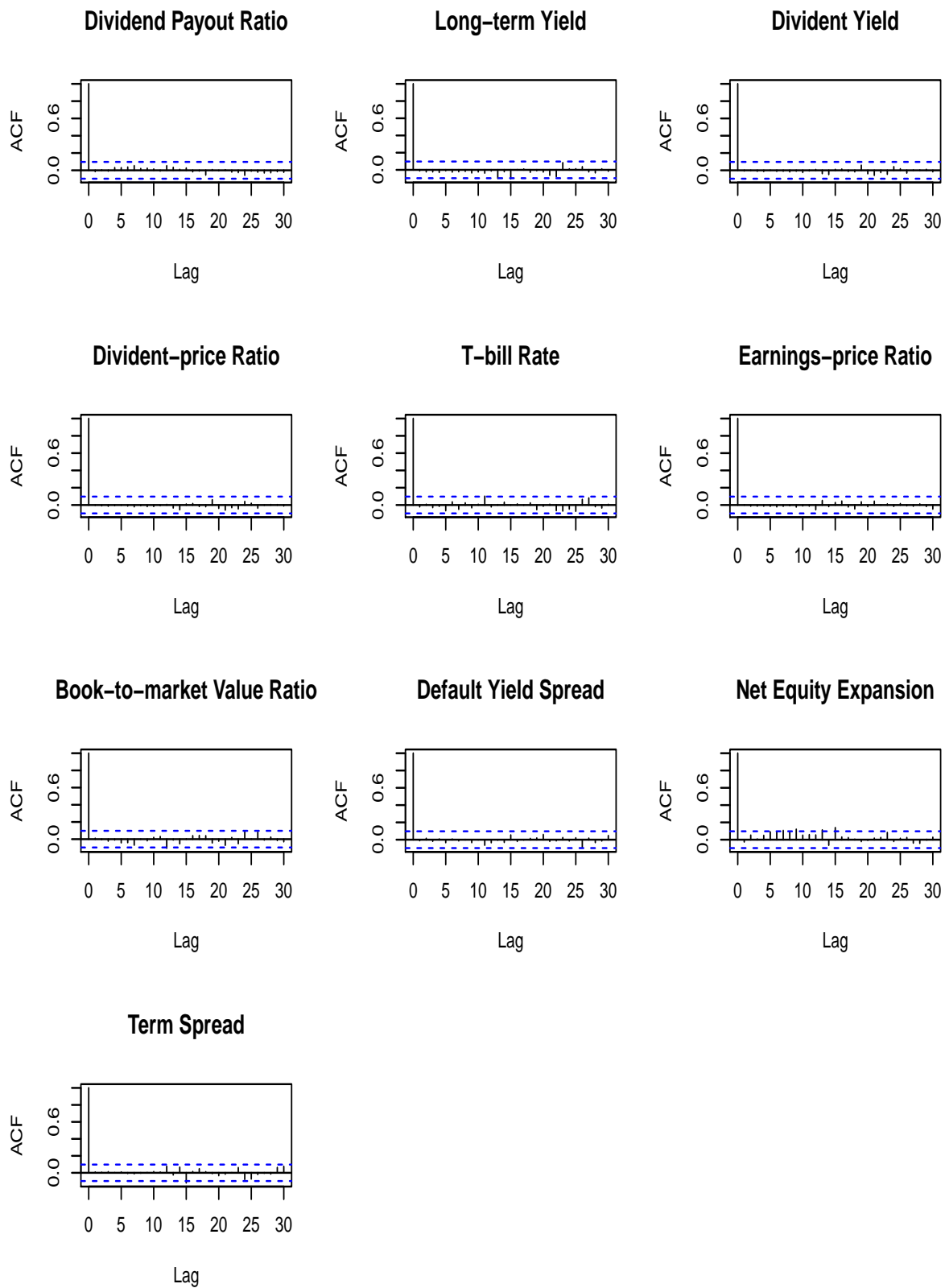


Figure 9: Autocorrelation function of the estimated V_t 's for the period 1/1982–12/2015 based on the ARMA(1,15) model, i.e., $X_t = \theta + \phi X_{t-1} + V_t - \sum_{j=1}^{15} \psi_j V_{t-j}$.

Table 1: Empirical sizes and powers are reported for the test based on Theorem 1 and model (3) with $\theta_0 = 0$ and known $\alpha = 1$ for testing $H_0 : \gamma_0 = 0$ vs $H_a : \gamma_0 \neq 0$.

(γ_0, ϕ_0)	$n = 200$		$n = 2000$		$n = 5000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.108	0.058	0.117	0.065	0.110	0.067
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.427	0.314	0.480	0.352	0.459	0.352
$(\frac{1}{\sqrt{n}}, 0.9)$	0.464	0.339	0.430	0.314	0.440	0.323
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.869	0.790	0.890	0.833	0.908	0.857
$(\frac{2}{\sqrt{n}}, 0.9)$	0.913	0.864	0.918	0.862	0.933	0.876
$(-\frac{3}{\sqrt{n}}, 0.9)$	0.988	0.973	0.999	0.997	0.997	0.991
$(\frac{3}{\sqrt{n}}, 0.9)$	0.998	0.996	0.998	0.994	1.000	0.999
$(-\frac{4}{\sqrt{n}}, 0.9)$	0.999	0.987	1.000	1.000	1.000	0.999
$(\frac{4}{\sqrt{n}}, 0.9)$	1.000	1.000	1.000	1.000	1.000	1.000
$(0, 1 - \frac{2}{n})$	0.103	0.059	0.115	0.056	0.110	0.048
$(-\frac{1}{\sqrt{n}}, 1 - \frac{2}{n})$	0.415	0.299	0.459	0.329	0.439	0.317
$(\frac{1}{\sqrt{n}}, 1 - \frac{2}{n})$	0.436	0.324	0.412	0.300	0.423	0.312
$(-\frac{2}{\sqrt{n}}, 1 - \frac{2}{n})$	0.838	0.768	0.874	0.791	0.887	0.817
$(\frac{2}{\sqrt{n}}, 1 - \frac{2}{n})$	0.908	0.826	0.903	0.816	0.894	0.832
$(-\frac{3}{\sqrt{n}}, 1 - \frac{2}{n})$	0.985	0.969	0.996	0.988	0.992	0.982
$(\frac{3}{\sqrt{n}}, 1 - \frac{2}{n})$	0.994	0.992	0.996	0.989	0.998	0.997
$(-\frac{4}{\sqrt{n}}, 1 - \frac{2}{n})$	0.998	0.995	1.000	1.000	1.000	0.999
$(\frac{4}{\sqrt{n}}, 1 - \frac{2}{n})$	1.000	1.000	1.000	1.000	1.000	1.000
$(0, 1)$	0.106	0.057	0.110	0.051	0.102	0.058
$(-\frac{1}{\sqrt{n}}, 1)$	0.404	0.301	0.442	0.323	0.425	0.306
$(\frac{1}{\sqrt{n}}, 1)$	0.402	0.281	0.398	0.294	0.405	0.298
$(-\frac{2}{\sqrt{n}}, 1)$	0.836	0.739	0.864	0.782	0.882	0.797
$(\frac{2}{\sqrt{n}}, 1)$	0.881	0.799	0.883	0.815	0.885	0.809
$(-\frac{3}{\sqrt{n}}, 1)$	0.978	0.959	0.992	0.980	0.990	0.980
$(\frac{3}{\sqrt{n}}, 1)$	0.995	0.989	0.997	0.989	0.997	0.994
$(-\frac{4}{\sqrt{n}}, 1)$	1.000	0.993	1.000	1.000	1.000	0.999
$(\frac{4}{\sqrt{n}}, 1)$	1.000	0.999	1.000	1.000	1.000	1.000

Table 2: Empirical sizes and powers are reported for the test based on Theorem 5 and model (3) with $\theta_0 = 0$ and unknown $\alpha = 1$ for testing $H_0 : \gamma_0 = 0$ vs $H_a : \gamma_0 \neq 0$.

(γ_0, ϕ_0)	$n = 200$		$n = 2000$		$n = 5000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.113	0.062	0.099	0.048	0.097	0.042
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.265	0.171	0.222	0.126	0.232	0.138
$(\frac{1}{\sqrt{n}}, 0.9)$	0.283	0.193	0.247	0.164	0.256	0.178
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.560	0.445	0.569	0.440	0.587	0.451
$(\frac{2}{\sqrt{n}}, 0.9)$	0.567	0.433	0.574	0.469	0.551	0.429
$(-\frac{3}{\sqrt{n}}, 0.9)$	0.883	0.752	0.864	0.781	0.865	0.787
$(\frac{3}{\sqrt{n}}, 0.9)$	0.876	0.808	0.850	0.780	0.894	0.810
$(-\frac{4}{\sqrt{n}}, 0.9)$	0.965	0.934	0.966	0.943	0.971	0.948
$(\frac{4}{\sqrt{n}}, 0.9)$	0.979	0.959	0.977	0.950	0.982	0.959
$(0, 1 - \frac{2}{n})$	0.1210	0.061	0.096	0.049	0.107	0.048
$(-\frac{1}{\sqrt{n}}, 1 - \frac{2}{n})$	0.262	0.172	0.229	0.140	0.247	0.160
$(\frac{1}{\sqrt{n}}, 1 - \frac{2}{n})$	0.293	0.204	0.261	0.182	0.268	0.179
$(-\frac{2}{\sqrt{n}}, 1 - \frac{2}{n})$	0.575	0.463	0.590	0.481	0.597	0.485
$(\frac{2}{\sqrt{n}}, 1 - \frac{2}{n})$	0.592	0.484	0.616	0.494	0.574	0.467
$(-\frac{3}{\sqrt{n}}, 1 - \frac{2}{n})$	0.863	0.771	0.886	0.793	0.884	0.823
$(\frac{3}{\sqrt{n}}, 1 - \frac{2}{n})$	0.909	0.825	0.898	0.815	0.910	0.835
$(-\frac{4}{\sqrt{n}}, 1 - \frac{2}{n})$	0.967	0.938	0.976	0.953	0.975	0.964
$(\frac{4}{\sqrt{n}}, 1 - \frac{2}{n})$	0.988	0.972	0.987	0.966	0.985	0.972
$(0, 1)$	0.123	0.066	0.093	0.050	0.106	0.049
$(-\frac{1}{\sqrt{n}}, 1)$	0.275	0.177	0.233	0.139	0.251	0.162
$(\frac{1}{\sqrt{n}}, 1)$	0.292	0.208	0.256	0.181	0.265	0.173
$(-\frac{2}{\sqrt{n}}, 1)$	0.587	0.475	0.589	0.483	0.597	0.489
$(\frac{2}{\sqrt{n}}, 1)$	0.598	0.492	0.617	0.484	0.576	0.464
$(-\frac{3}{\sqrt{n}}, 1)$	0.861	0.776	0.887	0.793	0.883	0.825
$(\frac{3}{\sqrt{n}}, 1)$	0.904	0.835	0.896	0.813	0.907	0.834
$(-\frac{4}{\sqrt{n}}, 1)$	0.970	0.934	0.977	0.955	0.975	0.964
$(\frac{4}{\sqrt{n}}, 1)$	0.989	0.973	0.984	0.969	0.983	0.972

Table 3: Empirical sizes and powers are reported for the test based on Theorem 1 and model (3) with $\theta_0 = 0$ and known $\alpha = 1$ for testing $H_0 : \beta_0 = 0$ vs $H_a : \beta_0 \neq 0$.

(γ_0, ϕ_0)	$n = 200$		$n = 2000$		$n = 5000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.119	0.063	0.095	0.044	0.090	0.039
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.566	0.458	0.655	0.542	0.690	0.570
$(\frac{1}{\sqrt{n}}, 0.9)$	0.719	0.590	0.692	0.583	0.705	0.577
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.936	0.890	0.990	0.977	0.992	0.987
$(\frac{2}{\sqrt{n}}, 0.9)$	0.994	0.987	0.997	0.993	0.998	0.996
$(-\frac{3}{\sqrt{n}}, 0.9)$	0.990	0.982	1.000	1.000	0.999	0.999
$(\frac{3}{\sqrt{n}}, 0.9)$	1.000	1.000	1.000	1.000	1.000	1.000
$(-\frac{4}{\sqrt{n}}, 0.9)$	0.998	0.994	1.000	1.000	1.000	1.000
$(\frac{4}{\sqrt{n}}, 0.9)$	0.999	0.999	1.000	1.000	1.000	1.000
$(0, 1 - \frac{2}{n})$	0.126	0.069	0.098	0.050	0.091	0.050
$(-\frac{2}{n}, 1 - \frac{2}{n})$	0.187	0.119	0.169	0.115	0.182	0.098
$(\frac{2}{n}, 1 - \frac{2}{n})$	0.221	0.127	0.173	0.112	0.195	0.109
$(-\frac{4}{n}, 1 - \frac{2}{n})$	0.378	0.303	0.332	0.241	0.296	0.208
$(\frac{4}{n}, 1 - \frac{2}{n})$	0.482	0.369	0.413	0.290	0.388	0.281
$(-\frac{6}{n}, 1 - \frac{2}{n})$	0.539	0.445	0.535	0.437	0.499	0.398
$(\frac{6}{n}, 1 - \frac{2}{n})$	0.738	0.613	0.678	0.575	0.652	0.529
$(-\frac{8}{n}, 1 - \frac{2}{n})$	0.666	0.595	0.648	0.576	0.645	0.572
$(\frac{8}{n}, 1 - \frac{2}{n})$	0.854	0.766	0.825	0.753	0.811	0.722
$(0, 1)$	0.128	0.071	0.103	0.051	0.111	0.057
$(-\frac{2}{n}, 1)$	0.430	0.345	0.256	0.181	0.260	0.184
$(\frac{2}{n}, 1)$	0.459	0.364	0.285	0.183	0.296	0.197
$(-\frac{4}{n}, 1)$	0.673	0.619	0.499	0.405	0.450	0.351
$(\frac{4}{n}, 1)$	0.791	0.703	0.621	0.505	0.552	0.456
$(-\frac{6}{n}, 1)$	0.775	0.725	0.678	0.606	0.645	0.568
$(\frac{6}{n}, 1)$	0.890	0.830	0.856	0.763	0.811	0.727
$(-\frac{8}{n}, 1)$	0.853	0.798	0.775	0.722	0.776	0.707
$(\frac{8}{n}, 1)$	0.934	0.898	0.941	0.888	0.926	0.864

Table 4: Empirical sizes and powers are reported for the test based on Theorem 5 and model (3) with $\theta_0 = 0$ and unknown $\alpha = 1$ for testing $H_0 : \beta_0 = 0$ vs $H_a : \beta_0 \neq 0$.

(γ_0, ϕ_0)	$n = 200$		$n = 2000$		$n = 5000$	
	10%	5%	10%	5%	10%	5%
$(0, 0.9)$	0.113	0.063	0.108	0.056	0.088	0.043
$(-\frac{1}{\sqrt{n}}, 0.9)$	0.359	0.262	0.380	0.248	0.395	0.284
$(\frac{1}{\sqrt{n}}, 0.9)$	0.474	0.329	0.422	0.287	0.423	0.287
$(-\frac{2}{\sqrt{n}}, 0.9)$	0.638	0.528	0.824	0.743	0.846	0.747
$(\frac{2}{\sqrt{n}}, 0.9)$	0.921	0.842	0.915	0.848	0.913	0.847
$(-\frac{3}{\sqrt{n}}, 0.9)$	0.810	0.680	0.980	0.949	0.986	0.968
$(\frac{3}{\sqrt{n}}, 0.9)$	0.994	0.986	0.998	0.995	1.000	0.995
$(-\frac{4}{\sqrt{n}}, 0.9)$	0.867	0.779	0.998	0.996	0.998	0.997
$(\frac{4}{\sqrt{n}}, 0.9)$	0.999	0.998	1.000	1.000	1.000	1.000
$(0, 1 - \frac{2}{n})$	0.108	0.061	0.095	0.036	0.082	0.048
$(-\frac{2}{n}, 1 - \frac{2}{n})$	0.152	0.0829	0.132	0.070	0.117	0.058
$(\frac{2}{n}, 1 - \frac{2}{n})$	0.163	0.096	0.136	0.057	0.153	0.089
$(-\frac{4}{n}, 1 - \frac{2}{n})$	0.248	0.156	0.236	0.153	0.258	0.164
$(\frac{4}{n}, 1 - \frac{2}{n})$	0.286	0.175	0.300	0.184	0.268	0.167
$(-\frac{6}{n}, 1 - \frac{2}{n})$	0.370	0.271	0.397	0.309	0.377	0.301
$(\frac{6}{n}, 1 - \frac{2}{n})$	0.490	0.367	0.469	0.359	0.440	0.317
$(-\frac{8}{n}, 1 - \frac{2}{n})$	0.474	0.366	0.508	0.419	0.510	0.419
$(\frac{8}{n}, 1 - \frac{2}{n})$	0.654	0.521	0.642	0.524	0.615	0.488
$(0, 1)$	0.116	0.065	0.1012	0.041	0.091	0.052
$(-\frac{2}{n}, 1)$	0.150	0.084	0.126	0.072	0.141	0.075
$(\frac{2}{n}, 1)$	0.227	0.141	0.212	0.140	0.231	0.150
$(-\frac{4}{n}, 1)$	0.267	0.178	0.301	0.215	0.295	0.219
$(\frac{4}{n}, 1)$	0.444	0.316	0.472	0.361	0.436	0.317
$(-\frac{6}{n}, 1)$	0.412	0.329	0.457	0.365	0.474	0.375
$(\frac{6}{n}, 1)$	0.619	0.497	0.650	0.548	0.624	0.537
$(-\frac{8}{n}, 1)$	0.561	0.449	0.607	0.527	0.551	0.474
$(\frac{8}{n}, 1)$	0.750	0.630	0.774	0.692	0.759	0.675

Table 5: P-values are reported for testing $H_0 : \gamma_0 = 0$ and $H_0 : \beta_0 = 0$ for both known and unknown α under model (3) for the CRSP value-weighted excess return.

Regressor	Model with known α		Model with unknown α	
	$H_0 : \gamma_0 = 0$	$H_0 : \beta_0 = 0$	$H_0 : \gamma_0 = 0$	$H_0 : \beta_0 = 0$
Dividend payout ratio	0.5058	0.5001	0.2885	0.7483
Long-term yield	0.5098	0.9924	0.3563	0.1145
Dividend yield	0.6041	0.0000	0.4188	0.0668
Dividend-price ratio	0.4983	0.0000	0.3318	0.0656
T-bill rate	0.5034	0.9225	0.3541	0.1246
Earnings-price ratio	0.5036	0.0000	0.2956	0.0412
Book-to-market value ratio	0.4948	0.0107	0.3289	0.0533
Default yield spread	0.5136	0.7235	0.2801	0.5484
Net equity expansion	0.4890	0.7816	0.2788	0.9143
Term spread	0.5117	0.7401	0.2768	0.9295

Table 6: P-values are reported for testing $H_0 : \gamma_0 = 0$ and $H_0 : \beta_0 = 0$ for both known and unknown α under model (3) for the S&P500 value-weighted excess return.

Regressor	Model with known α		Model with unknown α	
	$H_0 : \gamma_0 = 0$	$H_0 : \beta_0 = 0$	$H_0 : \gamma_0 = 0$	$H_0 : \beta_0 = 0$
Dividend payout ratio	0.0000	0.4658	0.0005	0.8430
Long-term yield	0.0000	0.0820	0.0006	0.3160
Dividend yield	0.0000	0.0000	0.0008	0.1247
Dividend-price ratio	0.0000	0.0000	0.0006	0.2671
T-bill rate	0.0000	0.2485	0.0006	0.3349
Earnings-price ratio	0.0000	0.0015	0.0005	0.2275
Book-to-market value ratio	0.0000	0.3828	0.0006	0.2092
Default yield spread	0.0000	0.3095	0.0005	0.5061
Net equity expansion	0.0000	0.8079	0.0005	0.8436
Term spread	0.0000	0.7632	0.0004	0.8855