JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2021, VOL. 39, NO. 3, 684–699
https://doi.org/10.1080/07350015.2020.1714632

Unified Tests for a Dynamic Predictive Regression

Bingduo Yang[a], Xiaohui Liu[b], Liang Peng[c], and Zongwu Cai[d]

[a]Lingnan (University) College, Sun Yat-sen University, Guangzhou, Guangdong, China; [b]School of Statistics, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China; [c]Department of Risk Management and Insurance, Georgia State University, Atlanta, GA; [d]Department of Economics, University of Kansas, Lawrence, KS

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 derived a Wald type test based on the context of the extended instrumental variable (IVX) methodology for testing predictability of stock returns, and Demetrescu showed that the local power of the standard IVX-based test could be improved for some range of alternative hypotheses and the tuning parameter when a lagged predicted variable is added to the predictive regression on purpose, which poses an important question on whether the predictive model should include a lagged predicted variable. This article proposes novel robust procedures for testing both the existence of a lagged predicted variable and the predictability of asset returns regardless of regressors being stationary or nearly integrated or unit root and the AR model for regressors with or without an intercept. A simulation study confirms the good finite sample performance of the proposed tests before illustrating their practical usefulness in analyzing real financial datasets.

ARTICLE HISTORY
Received September 2018
Accepted January 2020

KEYWORDS
Autoregressive errors; Empirical likelihood; IVX test procedure; Predictive regression; Weighted score

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 articles by Campbell (2008) and Phillips (2015). Typically, 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), and the monthly S&P 500 excess returns in Cai and Wang (2014) and Kostakis, Magdalinos, and Stamatogiannis (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, Magdalinos, and Stamatogiannis (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. Instead, one may prefer 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 value for a predictability test, is unknown in practice. To overcome these difficulties, several 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, Elliott, and Stock (1995), Campbell and Yogo (2006), Jansson and Moreira (2006), Phillips and Lee (2013), Cai and Wang (2014), Breitung and Demetrescu (2015), Kostakis, Magdalinos, and Stamatogiannis (2015), Demetrescu and Rodrigues (2016), and references therein.

To better appreciate the proposed study in this article, we start with summarizing existing results and methods for the following simple predictive regression model:

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

CONTACT Xiaohui Liu csuliuxh912@gmail.com School of Statistics, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, China.

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 the 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 cannot 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, Elliott, and Stock (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, Deo, and Yi (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, Cai, and Peng (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, Jacewitz, and Park (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, Cai, and Peng (2014) is to employ the property that $|X_t| \xrightarrow{p} \infty$ as $t \to \infty$ when $\{X_t\}$ is either nearly integrated or unit root. Therefore, the unified method in Zhu, Cai, and Peng (2014) is robust concerning 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 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 less persistent instrumental variables (IV) to achieve a unified normal limit. As illustrated by Kostakis, Magdalinos, and Stamatogiannis (2015), the IVX methodology offers a good balance between size control and power loss, and the power depends on some tuning parameters in constructing the IVX instruments with a sacrifice on the rate of convergence 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, Magdalinos, and Stamatogiannis (2015). Based on some Monte Carlo simulation studies, Kostakis, Magdalinos, and Stamatogiannis (2015) recommended taking $C_z = -I$ and $\beta \in (0.9, 0.95)$.[1] As Kostakis, Magdalinos, and Stamatogiannis (2015) assumed zero intercept in modeling regressors, that is, 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, we conjecture that the IVX method fails to unify the cases of zero and nonzero intercept θ, which is confirmed by our simulation study presented in Section 3.1. In summary, the IVX method in Kostakis, Magdalinos, and Stamatogiannis (2015) has difficulty in choosing tuning parameters, sacrifices the test power in the nonstationary case, and fails 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 on purpose and found that IVX based tests should have a better power for some range of alternative hypotheses and the tuning parameter in the IVX method. 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):

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

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 econometrically needed in real applications, that is, how to test the existence of a lagged predicted variable in a predictive regression, which has not been formally addressed in predictive regressions when regressors may be nearly integrated. This article addresses this issue by proposing novel testing procedures for the existence of the lagged predicted variables ($H_0 : \gamma = 0$) and the predictability ($H_0 : \beta = 0$). The proposed tests are robust as they work uniformly regardless of regressors being stationary or nearly integrated or unit root and the AR model for the regressors with or without intercept. Therefore, the proposed study in this article assumes γ is unknown and tests for either $H_0 : \gamma = 0$ or $H_0 : \beta = 0$ without requiring $\gamma = 0$, while Demetrescu (2014) assumed $\gamma = 0$ and only tested for $H_0 : \beta = 0$.

Although many tests for predictability have been proposed in the literature, conclusions on predictability are unfortunately quite contradictory for different datasets, data periods, and methods. For example, Kostakis, Magdalinos, and Stamatogiannis (2015) reported significant predictability for 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 01/1927–12/1994, which is in line with the findings in Campbell and Yogo (2006), and reported predictability only for term spread for the period 01/1952–12/2008 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, Magdalinos, and Stamatogiannis (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, Magdalinos, and Stamatogiannis (2015). After plotting the autocorrelation

[1] Note that the β in Kostakis, Magdalinos, and Stamatogiannis (2015) is different from the β in models (1) and (2).

function (ACF) of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}X_{t-1}$ from the model (1) with $\hat{\alpha}$ and $\hat{\beta}$ being the least squares estimators and Y_t being the CRSP value-weighted excess return for the periods 01/1927–12/1994, 01/1952–12/2015, and 01/1982–12/2015, respectively, it is clear that the assumption of uncorrelated U_t's is doubtful for the period 01/1927–12/1994, may be fine for the period 01/1952–12/2015, and is quite reasonable for the period 01/1982–12/2015. That is, conclusions on predictability in the literature for the period from 01/1927 to 12/1994 may be misleading due to the violation of the model assumption of uncorrelated errors. On the other hand, after plotting the ACF of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\beta}X_{t-1}$ for Y_t being the S&P 500 excess returns, we conclude that the assumption of uncorrelated U_t's does not hold for either of the three periods above. However, the ACF plots of $\hat{U}_t = Y_t - \hat{\alpha} - \hat{\gamma}Y_{t-1} - \hat{\beta}X_{t-1}$ from (2) with Y_t being either S&P 500 excess returns or CRSP value-weighted excess returns suggest that the assumption of uncorrelated U_t's is reasonable for the period 01/1982–12/2015, which will be our focused time window in Section 3. To save space, we do not include these ACF plots in the article, which are available upon request.

The main contribution of this article is to propose novel procedures for testing $H_0 : \gamma_0 = 0$ and $H_0 : \beta_0 = 0$ 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, Cai, and Peng (2014). Readers are referred to Owen (2001) for an overview of the empirical likelihood method, which has been proved to be quite effective in interval estimations and hypothesis tests. The developed new methodologies and theoretical results in this article are different from existing ones in several folds. First, the method is different from that in Zhu, Cai, and Peng (2014) as we have to deal with the lagged predicted variable carefully, and the developed power analysis is not addressed in Zhu, Cai, and Peng (2014). Second, the proposed unified inference has a faster rate of convergence than the IVX methods in Demetrescu (2014) and Kostakis, Magdalinos, and Stamatogiannis (2015) in the nonstationary case. Finally, the new methods work well for all cases, while the IVX tests in these two articles are severely undersized in testing $H_0 : \beta = 0$, especially when the intercept in modeling predicting variables is nonzero; see the simulation study in Section 3.1.

The rest of this article is organized as follows. Section 2 presents the methodologies and the 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

We consider the following general dynamic predictive regression model

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

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 $H_0 : \beta_0 = 0$ 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., $\phi_0 = 1$). Throughout, we use $\alpha_0, \gamma_0, \beta_0, \theta_0, \phi_0$ to denote the corresponding true values of parameters.

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 has an independent interest too. When the capital asset pricing model is applicable, it is common to assume $\alpha = 0$. The literature of mutual funds shows that around 80% of U.S. actively managed mutual funds have zero α in factor models; see, for example, Jensen (1968), Kosowski et al. (2006), and Fama and French (2010).

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

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

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 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 theorem[2] does not hold; see Zhu, Cai, and Peng (2014) for details. To fix this issue, following the idea in Zhu, Cai, and Peng (2014), we replace the second equation in (4) by the following weighted score equation

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

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

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

To describe the proposed empirical likelihood tests, we introduce the following notation. For $t = 1, 2, \ldots, n$, define $Z_{t1}(\gamma, \beta) = (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1})(Y_{t-1} - \beta X_{t-1})$ and $Z_{t2}(\gamma, \beta) = (Y_t - \alpha_0 - \gamma Y_{t-1} - \beta X_{t-1})\, X_{t-1}/\sqrt{1 + X_{t-1}^2}$. Based on $\{\boldsymbol{Z}_t(\gamma, \beta)\}_{t=1}^n$ with $\boldsymbol{Z}_t(\gamma, \beta) = (Z_{t1}(\gamma, \beta), Z_{t2}(\gamma, \beta))^T$, the empirical likelihood function for γ and β is given by

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

Then, it follows from the Lagrange multiplier technique that

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

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

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

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 $n \to \infty$, $-2\log L^{P1}(0) \xrightarrow{d} \chi^2(1)$ 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 $-2\log L^{P1}(0) > \chi^2_{1,1-\xi}$ or $-2\log L^{P2}(0) > \chi^2_{1,1-\xi}$ or $-2\log L(0,0) > \chi^2_{2,1-\xi}$, respectively, where $\chi^2_{1,1-\xi}$ and $\chi^2_{2,1-\xi}$ denote the $(1-\xi)$th quantile of a chi-squared limit with one degree of freedom and with two degrees of freedom, respectively. 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 $\theta_0 \neq 0$. Finally, note that the above method is different from Zhu, Cai, and Peng (2014) because of the term γY_{t-1}. Moreover, this extra term complicates the power analysis given below, which is not provided in Zhu, Cai, and Peng (2014).

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, Chen, and Yao (2010) for nonparametric regression models. Here, to unify the cases of stationary, nearly integrated and unit root, one can follow the idea in Li, Li, and Peng (2017) to take the model structure of $\{U_t\}$ into account by employing either empirical likelihood method or jackknife empirical likelihood method in Jing, Yuan, and Zhou (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

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

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

$$\boldsymbol{D}_1 = \begin{pmatrix} d_1\{E(U_1^2) + \alpha_0^2\} + d_2 E\{X_1(U_1 + \alpha_0)\} \\ d_1 E(\frac{(\alpha_0 + U_1)X_1}{\sqrt{1+X_1^2}}) + d_2 E(\frac{X_1^2}{\sqrt{1+X_1^2}}) \end{pmatrix},$$

$$\Sigma_1 = E(U_1^2)\begin{pmatrix} E(U_1^2) + \alpha_0^2 & E(\frac{(U_1+\alpha_0)X_1}{\sqrt{1+X_1^2}}) \\ E(\frac{(U_1+\alpha_0)X_1}{\sqrt{1+X_1^2}}) & E(\frac{X_1^2}{1+X_1^2}) \end{pmatrix}.$$

(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

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

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

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

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

(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

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

which has a noncentral chi-squared limit with one degree of freedom and noncentrality parameter $\boldsymbol{D}_3^T \Sigma_3^{-1} \boldsymbol{D}_3 > 0$ when

$d_2 \neq 0$, where $\boldsymbol{W}_3 \sim N(0, \Sigma_3)$,

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

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

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

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

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

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

$$\tilde{\Sigma}_1 = E(U_1^2) \begin{pmatrix} E(U_1^2) + \alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix},$$

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

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

where $\tilde{S}_2 = -(\alpha_0, 1)^T$, $\tilde{\boldsymbol{W}}_2 \sim N(0, \tilde{\Sigma}_2)$,

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

(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

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

where

$$\tilde{S}_3 = -\left(\lim_{t\to\infty} E(\sum_{j=1}^{t} \gamma_0^{t-j} U_j)^2 + (\frac{\alpha_0}{1-\gamma_0})^2, \quad \frac{\alpha_0}{1-\gamma_0}\right)^T,$$

$$\tilde{\boldsymbol{W}}_3 \sim N(0, \tilde{\Sigma}_3),$$

$$\tilde{\boldsymbol{D}}_3 = \begin{pmatrix} d_2 \frac{\alpha_0}{1-\gamma_0} \int_0^1 J_{V,\rho}(s)\, ds \\ d_2 \int_0^1 J_{V,\rho}(s)\, ds \end{pmatrix}, \text{ and } \tilde{\Sigma}_3$$

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

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

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

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

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

$$\bar{\Sigma}_1 = E(U_1^2) \begin{pmatrix} E(U_1^2) + \alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix},$$

and $\text{sgn}(x)$ 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

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

which has a central chi-squared limit with one degree of freedom even when $d_1 \neq 0$, where $\bar{S}_2 = -\big(\theta_0(\alpha_0 - \beta_0\theta_0), \quad |\theta_0|\big)^T \int_0^1 \frac{1-e^{-\rho s}}{\rho}\, ds$, $\bar{\boldsymbol{W}}_2 \sim N(0, \bar{\Sigma}_2)$, $\bar{\boldsymbol{D}}_2 = -\bar{S}_2 d_1 \beta_0$,

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

(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

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

which has a noncentral chi-squared limit with one degree of freedom and noncentral parameter $\bar{\boldsymbol{D}}_3\{\bar{\Sigma}_3^{-1} - \frac{\bar{\Sigma}_3^{-1}\bar{S}_3\bar{S}_3^T\bar{\Sigma}_3^{-1}}{\bar{S}_3^T\bar{\Sigma}_3^{-1}\bar{S}_3}\}\bar{\boldsymbol{D}}_3 > 0$ when $d_2 \neq 0$, where

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

$$\bar{\boldsymbol{W}}_3 \sim N(0, \bar{\Sigma}_3),$$

$$\bar{\boldsymbol{D}}_3 = \begin{pmatrix} d_2\theta_0 \frac{\alpha_0}{1-\gamma_0} \int_0^1 \frac{1-e^{-\rho s}}{\rho}\, ds \\ d_2|\theta_0| \int_0^1 \frac{1-e^{-\rho s}}{\rho}\, ds \end{pmatrix},$$

$$\bar{\Sigma}_3 = E(U_1^2) \lim_{t\to\infty} \begin{pmatrix} E(\sum_{j=1}^{t-1} \gamma_0^{t-1-j} U_j)^2 + (\frac{\alpha_0}{1-\gamma_0})^2 & \frac{\alpha_0}{1-\gamma_0}\text{sgn}(\theta_0) \\ \frac{\alpha_0}{1-\gamma_0}\text{sgn}(\theta_0) & 1 \end{pmatrix}.$$

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 $\tilde{\boldsymbol{D}}_2$ and $\bar{\boldsymbol{D}}_2$ become a multiplier of $\tilde{S}_2$ and $\bar{S}_2$, respectively, in these two cases.

Remark 3. When the predictive regression has a d-dimensional predictor $\boldsymbol{X}_t=(X_{t,1},\ldots,X_{t,d})^T$ and the number of nonstationary variables is less than or equal to one, we can develop a similar empirical likelihood test by using weights $\{\frac{X_{t-1,i}}{\sqrt{1+X_{t-1,i}^2}}\}_{i=1}^d$ in the corresponding score equations. However, this is not true when the predictive regression has a d-dimensional predictor with more than one nonstationary variable or has more than one lagged predicted variable as some score equations become asymptotically equivalent.

2.2. Model With an Unknown Intercept

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

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

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

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.

To unify all cases including zero or nonzero intercept θ, we follow the idea in Zhu, Cai, and Peng (2014) by splitting the data into two parts and using the difference with a big lag to get rid of the intercept first to 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. Again, the study is more involved than that in Zhu, Cai, and Peng (2014) as the extra term γY_{t-1} may be stationary or nonstationary. 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 $t=1,\ldots,m$. Then, the model (3) implies the following model

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

without an intercept, which is the same as the 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 $t\to\infty$ 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

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

where

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

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

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

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 $n\to\infty$, $-2\log\tilde{L}^{P1}(0)\xrightarrow{d}\chi^2(1)$ 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 $-2\log\tilde{L}^{P1}(0)>\chi^2_{1,1-\xi}$ or $-2\log\tilde{L}^{P2}(0)>\chi^2_{1,1-\xi}$ or $-2\log\tilde{L}(0,0)>\chi^2_{2,1-\xi}$, 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, and we skip this theoretical analysis.

Remark 4. It is known that the IVX tests in Demetrescu (2014) and Kostakis, Magdalinos, and Stamatogiannis (2015) employed a less persistent instrumental variable to achieve a unified normal limit so that the rate of convergence is slower than the standard rate in the nonstationary case with $\theta=0$. That is, the IVX tests for $H_0:\beta=0$ are less powerful than the proposed unified empirical likelihood tests, which have the standard rate given in Theorem 4(iii). Moreover, these IVX tests do not work for the nonstationary case with $\theta\neq 0$.

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 $\alpha=0$, $\theta=0$ or 0.2, $\psi_0=1$, $\psi_j=0$ for $j\geq 1$, and $\phi=0.2$ or 0.95 or 1. We generate 10,000 random samples with sample size $n=200$ or 400 or 1000

Table 1. Empirical sizes for testing the null hypothesis $H_0 : \beta = 0$.

ϕ	$\gamma = 0$ EL	EL1	KMS	Demes	$\gamma = 0.2$ EL	EL1	KMS	Demes	$\gamma = 0.5$ EL	EL1	KMS	Demes
Panel 1: $\theta = 0$												
$n = 200$												
0.2	0.0557	0.0556	0.0370	0.0471	0.0545	0.0595	0.0388	0.0491	0.0579	0.0586	0.0408	0.0363
0.95	0.0588	0.0585	0.0314	0.0508	0.0562	0.0572	0.0346	0.0443	0.0555	0.0595	0.0343	0.0276
1	0.0533	0.0562	0.0298	0.0537	0.0537	0.0582	0.0311	0.0449	0.0567	0.0585	0.0310	0.0250
$n = 400$												
0.2	0.0547	0.0553	0.0379	0.0502	0.0532	0.0551	0.0382	0.0474	0.0515	0.0524	0.0350	0.0310
0.95	0.0516	0.0480	0.0327	0.0476	0.0503	0.0524	0.0342	0.0415	0.0532	0.0600	0.0359	0.0284
1	0.0515	0.0563	0.0301	0.0522	0.0492	0.0577	0.0293	0.0419	0.0526	0.0530	0.0314	0.0241
$n = 1000$												
0.2	0.0522	0.0509	0.0463	0.0507	0.0469	0.0469	0.0545	0.0485	0.0510	0.0537	0.0527	0.0330
0.95	0.0532	0.0534	0.0499	0.0472	0.0503	0.0534	0.0463	0.0511	0.0521	0.0506	0.0513	0.0310
1	0.0488	0.0553	0.0251	0.0483	0.0536	0.0546	0.0207	0.0404	0.0486	0.0528	0.0218	0.0264
Panel 2: $\theta = 0.2$												
$n = 200$												
0.2	0.0545	0.0556	0.0380	0.0504	0.0524	0.0595	0.0394	0.0488	0.0581	0.0586	0.0377	0.0328
0.95	0.0591	0.0585	0.0326	0.0446	0.0575	0.0572	0.0329	0.0446	0.0552	0.0597	0.0381	0.0300
1	0.0506	0.0580	0.0097	0.0251	0.0543	0.0556	0.0088	0.0218	0.0517	0.0653	0.0096	0.0080
$n = 400$												
0.2	0.0559	0.0553	0.0386	0.0487	0.0492	0.0551	0.0389	0.0505	0.0499	0.0524	0.0401	0.0325
0.95	0.0486	0.0480	0.0326	0.0433	0.0492	0.0524	0.0323	0.0395	0.0543	0.0600	0.0353	0.0262
1	0.0524	0.0516	0.0025	0.0190	0.0480	0.0549	0.0034	0.0137	0.0545	0.0549	0.0019	0.0024
$n = 1000$												
0.2	0.0513	0.0509	0.0372	0.0510	0.0500	0.0469	0.0403	0.0450	0.0521	0.0537	0.0363	0.0331
0.95	0.0546	0.0534	0.0364	0.0519	0.0513	0.0534	0.0325	0.0465	0.0499	0.0506	0.0317	0.0342
1	0.0476	0.0513	0.0004	0.0019	0.0486	0.0514	0.0000	0.0000	0.0523	0.0528	0.0001	0.0001

NOTE: This table documents the empirical sizes for testing the null hypothesis $H_0 : \beta = 0$ versus the alternative $H_1 : \beta \neq 0$ in Equation (3) under the 5% nominal size. EL and EL1, respectively, represent the rejection rate for the Wald statistic calculated by the models with known α in Section 2.1 and unknown α in Section 2.2. KMS and Demes, respectively, denote the rejection rate for the Wald statistic calculated by the models in Kostakis, Magdalinos, and Stamatogiannis (2015) and Demetrescu (2014). Panel 1 refers to the case that $\theta = 0$, while Panel 2 refers to the case that $\theta = 0.2$. For each panel, the rejection rate is calculated through 10,000 repetitions with $\gamma \in \{0, 0.2, 0.5\}$ and $n \in \{200, 400, 1000\}$.

from model (3) with the above settings and $(U_t, V_t)^T$ having a bivariate Gaussian copula $C(F_1(U_t), F_2(V_t); \varrho)$ with correlation $\varrho = -0.5$ and marginal distributions being $t(5)$ and $t(4)$. We compute the empirical sizes and powers of the proposed tests for $H_0 : \beta = 0$ and $H_0 : \gamma = 0$ at 5% level based on Theorem 1 (i.e., α is assumed to be known) and Theorem 5 (i.e., α is unknown) by employing the R package "emplik."

Results in Table 1 show that the size of the proposed test for $H_0 : \beta = 0$ with known α is slightly more accurate than that with unknown α because the latter splits the data into two parts and hence reduces the effective sample size in the inference, but both have a quite accurate size for a larger n. In contrast, when X_t has zero intercept, the IVX test in Kostakis, Magdalinos, and Stamatogiannis (2015) tends to be undersized for all γ and that in Demetrescu (2014) is undersized clearly for $\gamma = 0.5$. Both become severely undersized when the predicting variable X_t is a unit root with a nonzero drift, that is, $\phi = 1$ and $\theta = 0.2$. These observations suggest that these IVX tests are not able to unify zero and nonzero intercept in modeling predicting variables, while the proposed empirical likelihood tests do unify all cases, including zero or nonzero intercept θ.

Results in Table 2 show that the size of the proposed tests for $H_0 : \gamma = 0$ is accurate whenever α is known or unknown. In contrast, the IVX test in Kostakis, Magdalinos, and Stamatogiannis (2015) is severely undersized when θ is nonzero and generally is less accurate than the proposed empirical likelihood tests when $\theta = 0$. We do not report the empirical sizes by the IVX method in Demetrescu (2014) because it is unavailable as $\gamma = 0$ is assumed in the model. Once again, these results suggest that the IVX test fails to unify the zero and nonzero θ.

We also compute the empirical powers for testing $H_0 : \beta = 0$ and $H_0 : \gamma = 0$. To save space, we only report results for testing $H_0 : \beta = 0$ when $\theta = 0$ and sample size $n = 200$. Note that the IVX test is severely undersized when $\theta \neq 0$. Figure 1 shows that the proposed empirical likelihood test with known α is much more powerful than the IVX tests in Demetrescu (2014) and Kostakis, Magdalinos, and Stamatogiannis (2015) when the predicting variable X_t is a unit root, which confirms that the IVX tests sacrifice the rate of convergence for achieving a unified normal limit in the nonstationary case. The proposed empirical likelihood test with unknown α is less powerful, but when β is close to zero, it has better local power than the IVX tests in the unit root case. Theoretically, the proposed test with unknown α should be more powerful than the IVX tests in the nonstationary case, but the technique of splitting data does impact the finite sample performance.

In conclusion, the proposed robust methods for testing $H_0 : \gamma = 0$ and $H_0 : \beta = 0$ provide an accurate size for all cases, while the IVX tests in Demetrescu (2014) and Kostakis, Magdalinos, and Stamatogiannis (2015) fail to unify zero and nonzero θ. The IVX method for testing $H_0 : \beta = 0$ in Kostakis, Magdalinos, and Stamatogiannis (2015) is severely undersized for all considered cases and that in Demetrescu (2014) suffers a

Table 2. Empirical sizes for testing the null hypothesis $H_0 : \gamma = 0$.

	$\beta = 0$				$\beta = 0.2$				$\beta = 0.5$			
ϕ	EL	EL1	KMS	Demes	EL	EL1	KMS	Demes	EL	EL1	KMS	Demes
Panel 1: $\theta = 0$												
$n = 200$												
0.2	0.0598	0.0647	0.0339	–	0.0620	0.0654	0.0346	–	0.0576	0.0647	0.0367	–
0.95	0.0670	0.0658	0.0358	–	0.0603	0.0673	0.0424	–	0.0596	0.0633	0.0447	–
1	0.0599	0.0658	0.0338	–	0.0604	0.0661	0.0461	–	0.0598	0.0613	0.0437	–
$n = 400$												
0.2	0.0585	0.0568	0.0380	–	0.0537	0.0593	0.0362	–	0.0594	0.0578	0.0349	–
0.95	0.0559	0.0557	0.0318	–	0.0567	0.0559	0.0420	–	0.0532	0.0591	0.0470	–
1	0.0538	0.0568	0.0299	–	0.0570	0.0559	0.0437	–	0.0575	0.0605	0.0387	–
$n = 1000$												
0.2	0.0551	0.0462	0.0528	–	0.0525	0.0542	0.0500	–	0.0483	0.0552	0.0519	–
0.95	0.0470	0.0510	0.0506	–	0.0502	0.0505	0.0550	–	0.0546	0.0517	0.0515	–
1	0.0530	0.0493	0.0519	–	0.0540	0.0504	0.0506	–	0.0563	0.0554	0.0430	–
Panel 2: $\theta = 0.2$												
$n = 200$												
0.2	0.0596	0.0647	0.0338	–	0.0609	0.0654	0.0389	–	0.0573	0.0647	0.0358	–
0.95	0.0673	0.0658	0.0360	–	0.0621	0.0673	0.0396	–	0.0591	0.0633	0.0438	–
1	0.0603	0.0648	0.0273	–	0.0600	0.0666	0.0442	–	0.0621	0.0620	0.0411	–
$n = 400$												
0.2	0.0581	0.0598	0.0382	–	0.0532	0.0620	0.0362	–	0.0579	0.0554	0.0369	–
0.95	0.0560	0.0554	0.0356	–	0.0556	0.0594	0.0433	–	0.0536	0.0587	0.0414	–
1	0.0544	0.0555	0.0258	–	0.0561	0.0630	0.0405	–	0.0563	0.0575	0.0366	–
$n = 1000$												
0.2	0.0545	0.0516	0.0354	–	0.0521	0.0526	0.0338	–	0.0505	0.0508	0.0332	–
0.95	0.0485	0.0517	0.0372	–	0.0506	0.0514	0.0390	–	0.0522	0.0491	0.0423	–
1	0.0534	0.0517	0.0287	–	0.0538	0.0509	0.0363	–	0.0557	0.0528	0.0373	–

This table documents the empirical sizes for testing the null hypothesis $H_0 : \gamma = 0$ versus the alternative $H_1 : \gamma \neq 0$ in Equation (3) under the 5% nominal size. EL and EL1, respectively, represent the rejection rate for the Wald statistic calculated by the models with known α in Section 2.1 and unknown α in Section 2.2. KMS and Demes, respectively, denote the rejection rate for the Wald statistic calculated by the models in Kostakis, Magdalinos, and Stamatogiannis (2015) and Demetrescu (2014). Panel 1 refers to the case that $\theta = 0$, while Panel 2 refers to the case that $\theta = 0.2$. For each panel, the rejection rate is calculated through 10,000 repetitions with $\beta \in \{0, 0.2, 0.5\}$ and $n \in \{200, 400, 1000\}$.

serious size distortion when γ is larger or $\{X_t\}$ is nonstationary. The proposed test with known α is most powerful in the nonstationary case. The technique of splitting data for the proposed test with unknown α does impact the finite sample performance, although theoretically, it is better than the IVX method in the nonstationary case.

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, Magdalinos, and Stamatogiannis (2015) and Cai and Wang (2014) by focusing on the period 01/1982–12/2015 for one of the two predicted variables, the CRSP value-weighted excess returns and the S&P 500 excess returns, and one of the ten financial predictors as mentioned in Section 1. These predictors are confirmed to be highly persistent by Kostakis, Magdalinos, and Stamatogiannis (2015).

In Table 3, 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 the 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 of 10%.

- When the predicted variable is the CRSP value-weighted excess return, the null hypothesis $H_0 : \gamma_0 = 0$ cannot be rejected for all considered regressors and the predictability (i.e., $\beta_0 \neq 0$) exists for dividend yield, dividend price ratio, and earnings price ratio 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 long-term yield, dividend yield, dividend price ratio, and earnings price ratio when α is known, but there exists no predictability for all regressors when α is unknown.

As argued in Section 1, the existing literature on testing predictability often ignores checking uncorrelated errors, while the employed tests for predictability heavily rely on this assumption. For example, under the assumption of uncorrelated errors, Demetrescu (2014) examined the IVX-based test by adding one period lagged predicted variable into the model on purpose, and found out that this improves the local power of the IVX-based test when the tuning parameter η in the IVX-test is less than 1/3. After checking the reasonable assumption of uncorrelated errors for the considered period, our proposed robust tests do not reject $H_0 : \gamma_0 = 0$ for the predicted variable

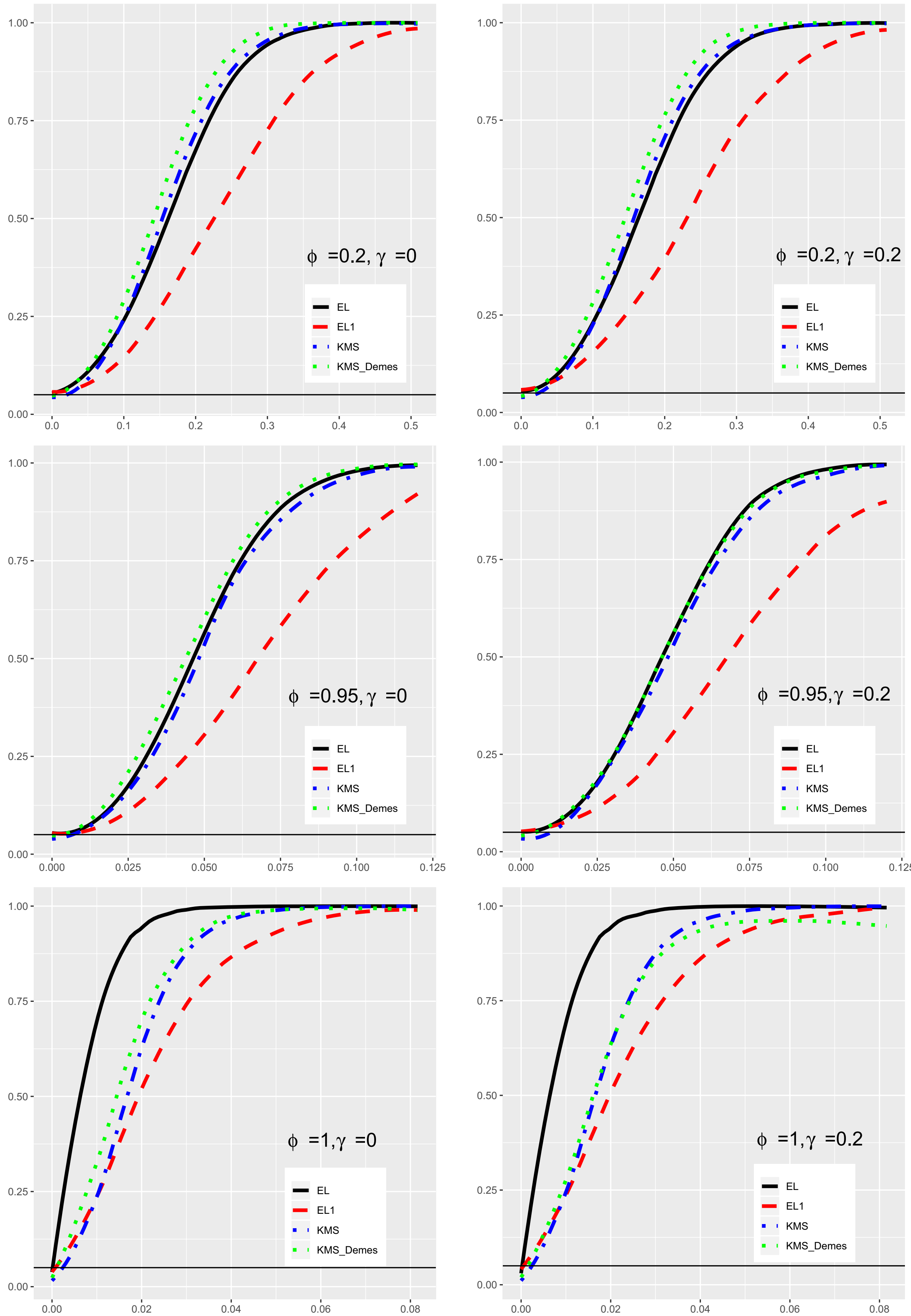

Figure 1. Empirical powers for testing $H_0 : \beta = 0$ when $\theta = 0$. Rejection rates for the proposed empirical likelihood tests and the IVX tests at level 5% are plotted against $\beta = 0.3 * (j-1)/\sqrt{200}, 0.68 * (j-1)/200, (j-1)/200, j \in \{0, 1, \ldots, 25\}$ with respect to $\phi = 0.2, 0.95, 1$, respectively. The solid curve and dashed curve, respectively, represent the proposed empirical likelihood tests with known α and unknown α, the dotted-dash curve and dotted curve, respectively, represent the IVX test in Kostakis, Magdalinos, and Stamatogiannis (2015) and Demetrescu (2014). The six panels correspond to the cases when $\phi \in \{0.2, 0.95, 1\}$ and $\gamma \in \{0, 0.2\}$. For each panel, the rejection rates are calculated through 10,000 repetitions with sample size $n = 200$.

Table 3. Testing results for dynamic predictive regression.

Regressor	Model with known α $H_0 : \gamma_0 = 0$	Model with known α $H_0 : \beta_0 = 0$	Model with unknown α $H_0 : \gamma_0 = 0$	Model with unknown α $H_0 : \beta_0 = 0$
Panel 1: Predicted variable: the CRSP value-weighted excess return				
Dividend payout ratio	0.505	0.500	0.288	0.748
Long-term yield	0.509	0.992	0.356	0.114
Dividend yield	0.604	0.000	0.418	0.066
Dividend-price ratio	0.498	0.000	0.331	0.065
T-bill rate	0.503	0.922	0.354	0.124
Earnings-price ratio	0.503	0.000	0.295	0.041
Book-to-market value ratio	0.494	0.010	0.328	0.053
Default yield spread	0.513	0.723	0.280	0.548
Net equity expansion	0.489	0.781	0.278	0.914
Term spread	0.511	0.740	0.276	0.929
Panel 2: Predicted variable: the S&P500 value-weighted excess return				
Dividend payout ratio	0.000	0.465	0.000	0.843
Long-term yield	0.000	0.082	0.000	0.316
Dividend yield	0.000	0.000	0.000	0.124
Dividend-price ratio	0.000	0.000	0.000	0.267
T-bill rate	0.000	0.248	0.000	0.334
Earnings-price ratio	0.000	0.001	0.000	0.227
Book-to-market value ratio	0.000	0.382	0.000	0.209
Default yield spread	0.000	0.309	0.000	0.506
Net equity expansion	0.000	0.807	0.000	0.843
Term spread	0.000	0.763	0.000	0.885

p-values are reported for testing the null hypothesis $H_0 : \gamma_0 = 0$ and $H_0 : \beta_0 = 0$ with known and unknown α under model (3).

CRSP value-weighted excess return, which means the method in Demetrescu (2014) by assuming no serial correlation and adding a lagged predicted variable to improve the test power for the less persistent case is valid for this predicted variable. However, our proposed robust tests 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, Magdalinos, and Stamatogiannis (2015) for the period after 01/1952, 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, Magdalinos, and Stamatogiannis (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. Furthermore, we implement the method in Zhu, Cai, and Peng (2014) for testing the predictability under the predictive regression without the lag and find that all predictors for significant at the level 1% except the long-term yield for the CRSP value-weighted excess return with p-value 0.137. That is, it is important to study the proposed model.

4. Conclusions

Without characterizing the stochastic properties of regressors, this article proposes new unified empirical likelihood tests based on 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, Magdalinos, and Stamatogiannis (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 and zero or nonzero intercept in modeling predictors. Hence, the proposed new tests are easy to implement without any ad hoc method such as a bootstrap method for obtaining critical values.

The Monte Carlo simulation study shows that the proposed tests give accurate size and are powerful while the IVX tests for $H_0 : \beta = 0$ in Demetrescu (2014) and Kostakis, Magdalinos, and Stamatogiannis (2015) are severely undersized, and the IVX test for $H_0 : \gamma = 0$ in Kostakis, Magdalinos, and Stamatogiannis (2015) suffers a serious size distortion when the predicting variable has a nonzero intercept. The proposed robust procedure for testing $H_0 : \gamma_0 = 0$ can be employed to check whether the method in Demetrescu (2014) is applicable as $\gamma = 0$ is imposed in the model. 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.

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{Z}_t(0,0) = \boldsymbol{W}_1 + \begin{pmatrix} d_1\{E(U_1^2)+\alpha_0^2\} + d_2 E\{X_1(\alpha_0+U_1)\} \\ d_1 E(\frac{(\alpha_0+U_1)X_1}{\sqrt{1+X_1^2}}) + d_2 E(\frac{X_1^2}{\sqrt{1+X_1^2}}) \end{pmatrix} + o_p(1), \tag{A.1}$$

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

$$\text{and} \quad \max_{1\le t\le n}||\boldsymbol{Z}_t(0,0)||=o_p(n^{1/2}), \quad (A.3)$$

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

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{Z}_t(0,\beta_0)=\boldsymbol{W}_2+\begin{pmatrix}d_1E\{(\alpha_0+\beta_0X_1+U_2)(\alpha_0+U_2+\beta_0X_1-\beta_0X_2)\}\\ d_1E(\frac{(\alpha_0+\beta_0X_1+U_2)X_2}{\sqrt{1+X_2^2}})\end{pmatrix}+o_p(1),$$

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{Z}_t(0,\beta_0)\boldsymbol{Z}_t^T(0,\beta_0)=E(U_1^2)\begin{pmatrix}E(\alpha_0+U_2+\beta_0X_1-\beta_0X_2)^2 & E(\frac{(\alpha_0+U_2+\beta_0X_1-\beta_0X_2)X_2}{\sqrt{1+X_2^2}})\\ E(\frac{(\alpha_0+U_2+\beta_0X_1-\beta_0X_2)X_2}{\sqrt{1+X_2^2}}) & E(\frac{X_1^2}{1+X_1^2})\end{pmatrix}+o_p(1):=\Sigma_2+o_p(1),$$

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

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

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{Z}_t(\gamma_0,0)=\boldsymbol{W}_3+\lim_{t\to\infty}\begin{pmatrix}d_2E\{X_{t-1}(\frac{\alpha_0}{1-\gamma_0}+\sum_{j=1}^{t-1}\gamma_0^{t-1-j}U_j)\}\\ d_2E(\frac{X_1^2}{\sqrt{1+X_1^2}})\end{pmatrix}+o_p(1),$$

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{Z}_t(\gamma_0,0)\boldsymbol{Z}_t^T(\gamma_0,0)=\lim_{t\to\infty}\begin{pmatrix}E(\frac{\alpha_0}{1-\gamma_0}+\sum_{j=1}^{t-1}\gamma_0^{t-1-j}U_j)^2 & E\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}(\frac{\alpha_0}{1-\gamma_0}+\sum_{j=1}^{t-1}\gamma_0^{t-1-j}U_j)\}\\ E\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}(\frac{\alpha_0}{1-\gamma_0}+\sum_{j=1}^{t-1}\gamma_0^{t-1-j}U_j)\} & E(\frac{X_1^2}{1+X_1^2})\end{pmatrix}\times E(U_1^2)+o_p(1):=\Sigma_3+o_p(1),$$

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

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

Proof. (i) Since

$$Y_t=\alpha_0\frac{1-\gamma_0^t}{1-\gamma_0}+\gamma_0^tY_0+\sum_{j=0}^{t-1}\gamma_0^{t-1-j}\beta_0X_j+\sum_{j=1}^{t}\gamma_0^{t-j}U_j, \quad (A.4)$$

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

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_{t1}(0,0)\\&=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}U_tY_{t-1}+\frac{d_1}{n}\sum_{t=1}^{n}Y_{t-1}Y_{t-1}+\frac{d_2}{n}\sum_{t=1}^{n}X_{t-1}Y_{t-1}\\&=\frac{1}{\sqrt{n}}\sum_{t=3}^{n}U_t\{\alpha_0\frac{1-(d_1/\sqrt{n})^{t-1}}{1-d_1/\sqrt{n}}+(\frac{d_1}{\sqrt{n}})^{t-1}Y_0\\&\quad+\sum_{j=0}^{t-2}(\frac{d_1}{\sqrt{n}})^{t-2-j}\frac{d_2}{\sqrt{n}}X_j\\&\quad+\sum_{j=1}^{t-2}(\frac{d_1}{\sqrt{n}})^{t-1-j}U_j+U_{t-1}\}+d_1\{E(U_1^2)+\alpha_0^2\}\\&\quad+d_2E\{X_1(\alpha_0+U_1)\}+o_p(1)\\&=\frac{1}{\sqrt{n}}\sum_{t=3}^{n}U_t(\alpha_0+U_{t-1})+d_1\{E(U_1^2)+\alpha_0^2\}\\&\quad+d_2E\{X_1(\alpha_0+U_1)\}+o_p(1)\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}Z_{t2}(0,0)\\&=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}U_t\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}+\frac{d_1}{n}\sum_{t=1}^{n}\frac{Y_{t-1}X_{t-1}}{\sqrt{1+X_{t-1}^2}}\\&\quad+\frac{d_2}{n}\sum_{t=1}^{n}\frac{X_{t-1}^2}{\sqrt{1+X_{t-1}^2}}\\&=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{U_tX_{t-1}}{\sqrt{1+X_{t-1}^2}}+d_1E(\frac{(\alpha_0+U_1)X_1}{\sqrt{1+X_1^2}})\\&\quad+d_2E(\frac{X_1^2}{\sqrt{1+X_1^2}})+o_p(1),\end{aligned}$$

which imply (A.1). Similarly, we can prove (A.2) and (A.3).

(ii) By noting that $Y_t-\alpha_0-\beta_0X_{t-1}-U_t=o_p(1)$ as $t\to\infty$, 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 $t\to\infty$, results follow from similar arguments in proving (i). □

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

(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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{Z}_t^*(0,0)=\tilde{\boldsymbol{W}}_1+\begin{pmatrix}d_1\{E(U_1^2)+\alpha_0^2\}+d_2\alpha_0\int_0^1J_{V,\rho}(s)\,ds\\ d_1\alpha_0+d_2\int_0^1J_{V,\rho}(s)\,ds\end{pmatrix}+o_p(1), \quad (A.5)$$

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Z}_t^*(0,0)\boldsymbol{Z}_t^{*T}(0,0) = E(U_1^2)\begin{pmatrix} E(U_1^2)+\alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix} + o_p(1)$$

$$:= \tilde{\Sigma}_1 + o_p(1), \tag{A.6}$$

$$\text{and} \quad \max_{1\le t\le n} ||\boldsymbol{Z}_t^*(0,0)|| = o_p(n^{1/2}), \tag{A.7}$$

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

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{Z}_t^*(0,\beta_0) = \tilde{\boldsymbol{W}}_2 + \begin{pmatrix} d_1\alpha_0\beta_0\int_0^1 J_{V,\rho}(s)\,ds \\ d_1\beta_0\int_0^1 J_{V,\rho}(s)\,ds \end{pmatrix} + o_p(1), \tag{A.8}$$

$$\begin{aligned} &\tfrac{1}{n}\textstyle\sum_{t=1}^{n} \boldsymbol{Z}_t^*(0,\beta_0)\boldsymbol{Z}_t^{*T}(0,\beta_0) \\ &= E(U_1^2)\begin{pmatrix} E(U_1-\beta_0\sum_{j=0}^{\infty}\psi_j V_{1-j})^2+\alpha_0^2 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix} + o_p(1) \\ &:= \tilde{\Sigma}_2 + o_p(1), \end{aligned} \tag{A.9}$$

$$\text{and} \quad \max_{1\le t\le n} ||\boldsymbol{Z}_t^*(0,\beta_0)|| = o_p(n^{1/2}), \tag{A.10}$$

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

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{Z}_t^*(\gamma_0,0) = \tilde{\boldsymbol{W}}_3 + \begin{pmatrix} d_2\frac{\alpha_0}{1-\gamma_0}\int_0^1 J_{V,\rho}(s)\,ds \\ d_2\int_0^1 J_{V,\rho}(s)\,ds \end{pmatrix} + o_p(1), \tag{A.11}$$

$$\begin{aligned} &\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Z}_t^*(\gamma_0,0)\boldsymbol{Z}_t^{*T}(\gamma_0,0) \\ &= E(U_1^2)\lim_{t\to\infty}\begin{pmatrix} E(\sum_{j=1}^{t}\gamma_0^{t-j}U_j)^2+(\frac{\alpha_0}{1-\gamma_0})^2 & \frac{\alpha_0}{1-\gamma_0} \\ \frac{\alpha_0}{1-\gamma_0} & 1 \end{pmatrix} \\ &\quad + o_p(1) \\ &:= \tilde{\Sigma}_3 + o_p(1), \end{aligned} \tag{A.12}$$

$$\text{and} \quad \max_{1\le t\le n} ||\boldsymbol{Z}_t^*(\gamma_0,0)|| = o_p(n^{1/2}), \tag{A.13}$$

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

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

$$\frac{1}{\sqrt{n}}X_{[nr]} \xrightarrow{D} J_{V,\rho}(r) \quad \text{in the space} \quad D[0,1], \tag{A.14}$$

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 (A.4) and (A.14), we have $Y_t - \alpha_0 - U_t = o_p(1)$ as $t \to \infty$. Hence,

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{t1}(0,0) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_tY_{t-1} + \frac{d_1}{n}\sum_{t=1}^{n} Y_{t-1}Y_{t-1} \\ &\quad + \frac{d_2}{n^{3/2}}\sum_{t=1}^{n} X_{t-1}Y_{t-1} \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t(\alpha_0+U_{t-1}) + d_1\{E(U_1^2)+\alpha_0^2\} \\ &\quad + d_2E(Y_1)\int_0^1 J_{V,\rho}(s)\,ds + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t(\alpha_0+U_{t-1}) + d_1\{E(U_1^2)+\alpha_0^2\} \\ &\quad + d_2\alpha_0\int_0^1 J_{V,\rho}(s)\,ds + o_p(1) \end{aligned} \tag{A.15}$$

and

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{t2}(0,0) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \frac{d_1}{n}\sum_{t=1}^{n}\frac{Y_{t-1}X_{t-1}}{\sqrt{1+X_{t-1}^2}} \\ &\quad + \frac{d_2}{n^{3/2}}\sum_{t=1}^{n}\frac{X_{t-1}^2}{\sqrt{1+X_{t-1}^2}} \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \frac{d_1\alpha_0}{n}\sum_{t=1}^{n}\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} \\ &\quad + \frac{d_1}{n}\sum_{t=1}^{n} U_t\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} \\ &\quad + \frac{d_2}{n^{3/2}}\sum_{t=1}^{n}\frac{X_{t-1}^2}{\sqrt{1+X_{t-1}^2}} + o_p(1). \end{aligned} \tag{A.16}$$

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

$$\begin{aligned} &\tfrac{1}{\sqrt{n}}\textstyle\sum_{t=1}^{n} U_t\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} \\ &= \tfrac{1}{\sqrt{n}}\textstyle\sum_{t=1}^{n}(S_t-S_{t-1})\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} \\ &= \tfrac{1}{\sqrt{n}}S_n\frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} + \tfrac{1}{\sqrt{n}}\textstyle\sum_{t=1}^{n} S_t\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} - \frac{X_t}{\sqrt{1+X_t^2}}\}. \end{aligned} \tag{A.17}$$

It follows from Taylor expansion that

$$\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} - \frac{X_t}{\sqrt{1+X_t^2}} = (1+\xi_t^2)^{-3/2}(X_{t-1}-X_t), \tag{A.18}$$

where ξ_t lies between X_{t-1} and X_t. By (A.14), 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 $t \to \infty$, which imply that

$$|\xi_t|/t^a \xrightarrow{p} \infty \quad \text{for any} \quad a \in (0,1/2) \quad \text{as} \quad t \to \infty. \tag{A.19}$$

It follows from (A.18) and (A.19) that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} S_t\left(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} - \frac{X_{t-2}}{\sqrt{1+X_{t-2}^2}}\right) = o_p(1). \tag{A.20}$$

By (A.17) and (A.20), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} = \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} U_t + o_p(1). \tag{A.21}$$

Similarly, we have

$$\begin{cases} \frac{1}{n}\sum_{t=1}^{n}\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} = \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}} + o_p(1) \\ \frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{X_{t-1}^2}{\sqrt{1+X_{t-1}^2}} = \frac{X_{n-1}}{\sqrt{1+X_{n-1}^2}}\frac{1}{n}\sum_{t=1}^{n}\frac{X_{t-1}}{n^{1/2}} + o_p(1). \end{cases} \tag{A.22}$$

Hence, (A.5) follows from (A.14), (A.15), (A.16), (A.21), and (A.22).

(ii) It follows from (A.4) and (A.14) that

$$\frac{1}{\sqrt{n}}Y_{[nr]} = \frac{\beta_0}{\sqrt{n}}X_{[nr]} + o_p(1) \xrightarrow{D} \beta_0 J_{V,\rho}(r) \quad \text{in the space} \quad D[0,1]. \tag{A.23}$$

Hence,

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

and

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

which imply (A.8). Similarly, we can prove (A.9) and (A.10).

(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 $t\to\infty$, we have

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

and

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

which imply (A.11). Similarly, we can prove (A.12) and (A.13). □

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{Z}_t(0,0)=\bar{\boldsymbol{W}}_1+\begin{pmatrix} d_1\{E(U_1^2)+\alpha_0^2\}+d_2\alpha_0\theta_0\int_0^1\frac{1-e^{-\rho s}}{s}\,ds\\ d_1\alpha_0 sgn(\theta_0)+d_2|\theta_0|\int_0^1\frac{1-e^{-\rho s}}{\rho}\,ds\end{pmatrix}+o_p(1), \tag{A.24}$$

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{Z}_t(0,0)\boldsymbol{Z}_t^T(0,0)=E(U_1^2)\begin{pmatrix}E(U_1^2)+\alpha_0^2 & \alpha_0\\ \alpha_0 & 1\end{pmatrix}+o_p(1):=\bar{\Sigma}_1+o_p(1), \tag{A.25}$$

$$\text{and}\quad \max_{1\le t\le n}||\boldsymbol{Z}_t(0,0)||=o_p(n^{1/2}), \tag{A.26}$$

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

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{Z}_t(0,\beta_0)=\bar{\boldsymbol{W}}_2+\begin{pmatrix} d_1\beta_0\theta_0(\alpha_0-\beta_0\theta_0)\int_0^1\frac{1-e^{-\rho s}}{\rho}\,ds\\ d_1\beta_0|\theta_0|\int_0^1\frac{1-e^{-\rho s}}{\rho}\,ds\end{pmatrix}+o_p(1), \tag{A.27}$$

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{Z}_t(0,\beta_0)\boldsymbol{Z}_t^T(0,\beta_0)$$
$$=E(U_1^2)\begin{pmatrix}E(\alpha_0+U_1-\beta_0\theta_0-\beta_0\sum_{j=0}^{\infty}\psi_jV_{1-j})^2 & (\alpha_0-\beta_0\theta_0)\mathrm{sgn}(\theta_0)\\ (\alpha_0-\beta_0\theta_0)\mathrm{sgn}(\theta_0) & 1\end{pmatrix}+o_p(1)$$
$$:=\bar{\Sigma}_2+o_p(1), \tag{A.28}$$

$$\text{and}\quad \max_{1\le t\le n}||\boldsymbol{Z}_t(0,\beta_0)||=o_p(n^{1/2}), \tag{A.29}$$

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

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{Z}_t(\gamma_0,0)=\bar{\boldsymbol{W}}_3+\begin{pmatrix} d_2\theta_0\frac{\alpha_0}{1-\gamma_0}\int_0^1\frac{1-e^{-\rho s}}{s}\,ds\\ d_2|\theta_0|\int_0^1\frac{1-e^{-\rho s}}{\rho}\,ds\end{pmatrix}+o_p(1), \tag{A.30}$$

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{Z}_t(\gamma_0,0)\boldsymbol{Z}_t^T(\gamma_0,0)$$
$$=E(U_1^2)\lim_{t\to\infty}\begin{pmatrix}E(\sum_{j=1}^{t-1}\gamma_0^{t-1-j}U_j)^2+(\frac{\alpha_0}{1-\gamma_0})^2 & \frac{\alpha_0}{1-\gamma_0}\mathrm{sgn}(\theta_0)\\ \frac{\alpha_0}{1-\gamma_0}\mathrm{sgn}(\theta_0) & 1\end{pmatrix}+o_p(1)$$
$$:=\bar{\Sigma}_3+o_p(1), \tag{A.31}$$

$$\text{and}\quad \max_{1\le t\le n}||\boldsymbol{Z}_t(\gamma_0,0)||=o_p(n^{1/2}), \tag{A.32}$$

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

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 $t \to \infty$, 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 $t \to \infty$, results follow from the same arguments in proving Lemma 2(iii). □

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 $|\beta - \beta_0| \le n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2+\delta)$ as $n \to \infty$, and β^* and $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^*(\beta^*)$ satisfy $Q_{1n}(\beta^*, \boldsymbol{\lambda}^*) = 0$ and $Q_{2n}(\beta^*, \boldsymbol{\lambda}^*) = 0$, where

$$Q_{1n}(\beta, \boldsymbol{\lambda}) := \frac{1}{n}\sum_{t=1}^n \frac{\boldsymbol{Z}_t(0,\beta)}{1+\boldsymbol{\lambda}^T \boldsymbol{Z}_t(0,\beta)} \quad \text{and}$$

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

(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 $|\gamma - \gamma_0| \le n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2+\delta)$ as $n \to \infty$, and γ^* and $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^*(\gamma^*)$ satisfy $Q_{3n}(\gamma^*, \boldsymbol{\lambda}^*) = 0$ and $Q_{4n}(\gamma^*, \boldsymbol{\lambda}^*) = 0$, where

$$Q_{3n}(\gamma, \boldsymbol{\lambda}) := \frac{1}{n}\sum_{t=1}^n \frac{\boldsymbol{Z}_t(\gamma,0)}{1+\boldsymbol{\lambda}^T \boldsymbol{Z}_t(\gamma,0)}, \quad \text{and}$$

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

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

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

$$\tilde{Q}_{1n}(\tilde{\beta}, \tilde{\boldsymbol{\lambda}}) := \frac{1}{n}\sum_{t=1}^n \frac{\tilde{\boldsymbol{Z}}_t^*(0,\tilde{\beta})}{1+\tilde{\boldsymbol{\lambda}}^T \tilde{\boldsymbol{Z}}_t^*(0,\tilde{\beta})} \quad \text{and}$$

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

(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 $|\gamma - \gamma_0| \le n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2+\delta)$ as $n \to \infty$, 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

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

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

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

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 $\bar{\beta} = \beta n$, $\bar{\beta}_0 = \beta_0 n$, $\bar{\boldsymbol{Z}}_t(\gamma, \bar{\beta}) = \boldsymbol{Z}_t(\gamma, \beta)$ 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 $|\bar{\beta} - \bar{\beta}_0| \le n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2+\delta)$ as $n \to \infty$, 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

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

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

(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 $|\gamma - \gamma_0| \le n^{-1/\delta_0}$ for some $\delta_0 \in (2, 2+\delta)$ as $n \to \infty$, 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

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

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

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

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

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

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 (A.14) that

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

and

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

which imply that

$$\frac{\partial \tilde{Q}_{1n}(\tilde{\beta}_0, 0)}{\partial \tilde{\beta}} = \begin{pmatrix} -\alpha_0 \int_0^1 J_{V,\rho}(s)\,ds \\ -\int_0^1 J_{V,\rho}(s)\,ds \end{pmatrix} + o_p(1) =: \tilde{S}_2^* + o_p(1). \quad \text{(A.33)}$$

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

$$\begin{cases} \frac{\partial \tilde{Q}_{1n}(\bar{\beta}_0,0)}{\partial \bar{\boldsymbol{\lambda}}} = -\tilde{\Sigma}_2 + o_p(1), \\ \tilde{Q}_{1n}(\bar{\beta}_0, 0) = O_p(n^{-1/2}), \quad \frac{\tilde{Q}_{1n}(\bar{\beta}_0,0)}{\partial \bar{\beta}} = O_p(1), \\ \frac{\partial \tilde{Q}_{2n}(\bar{\beta}_0,0)}{\partial \bar{\beta}} = 0, \quad \frac{\partial \tilde{Q}_{2n}(\bar{\beta}_0,0)}{\partial \bar{\boldsymbol{\lambda}}} = \tilde{S}_2^* + o_p(1) = O_p(1), \end{cases} \tag{A.34}$$

where $\tilde{\Sigma}_2$ is defined in Lemma 2(ii). By (A.33) and (A.34), 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

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

and

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

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

$$\begin{aligned} \sqrt{n}\bar{\boldsymbol{\lambda}}^* &= \tilde{\Sigma}_2^{-1}\sqrt{n}\tilde{Q}_{1n}(\bar{\beta}_0, 0) + \tilde{\Sigma}_2^{-1}\tilde{S}_2^*\sqrt{n}(\bar{\beta}^* - \bar{\beta}_0) + o_p(1) \\ &= \{\tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1}\tilde{S}_2^*\tilde{S}_2^{*T}\tilde{\Sigma}_2^{-1}}{\tilde{S}_2^{*T}\tilde{\Sigma}_2^{-1}\tilde{S}_2^*}\}\sqrt{n}\tilde{Q}_{1n}(\bar{\beta}_0, 0) + o_p(1) \\ &= \{\tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1}\tilde{S}_2\tilde{S}_2^T\tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T\tilde{\Sigma}_2^{-1}\tilde{S}_2}\}\sqrt{n}\tilde{Q}_{1n}(\bar{\beta}_0, 0) + o_p(1), \end{aligned} \tag{A.35}$$

where $\tilde{S}_2 = -(\alpha_0, 1)^T$. It follows from (A.35) and Taylor expansion that

$$\begin{aligned} &-2\log L^{P1}(0) \\ =&-2\log \bar{L}(0, \bar{\beta}^*) \\ =&2\sum_{t=1}^n \bar{\boldsymbol{\lambda}}^{*T}\boldsymbol{Z}_t(0, \bar{\beta}^*) - \sum_{t=1}^n \bar{\boldsymbol{\lambda}}^{*T}\boldsymbol{Z}_t(0, \bar{\beta}^*)\boldsymbol{Z}_t^T(0, \bar{\beta}^*)\bar{\boldsymbol{\lambda}}^* \\ &+ o_p(1) \\ =&2n\bar{\boldsymbol{\lambda}}^{*T}\tilde{Q}_{1n}(\bar{\beta}_0, 0) + 2n\bar{\boldsymbol{\lambda}}^{*T}\frac{\partial \tilde{Q}_{1n}(\bar{\beta}_0, 0)}{\partial \bar{\beta}}(\bar{\beta}^* - \bar{\beta}_0) \\ &- n\bar{\boldsymbol{\lambda}}^{*T}\tilde{\Sigma}_2\bar{\boldsymbol{\lambda}}^* + o_p(1) \\ =&2\sqrt{n}\bar{\boldsymbol{\lambda}}^{*T}\{\sqrt{n}\tilde{Q}_{1n}(\bar{\beta}_0, 0) + \tilde{S}_2\sqrt{n}(\bar{\beta}^* - \bar{\beta}_0)\} - n\bar{\boldsymbol{\lambda}}^{*T}\tilde{\Sigma}_2\bar{\boldsymbol{\lambda}}^* \\ &+ o_p(1) \\ =&2\sqrt{n}\bar{\boldsymbol{\lambda}}^{*T}\tilde{\Sigma}_2\sqrt{n}\bar{\boldsymbol{\lambda}}^* - n\bar{\boldsymbol{\lambda}}^{*T}\tilde{\Sigma}_2\bar{\boldsymbol{\lambda}}^* + o_p(1) \\ =&\{\sqrt{n}\tilde{Q}_{1n}(\bar{\beta}_0, 0)\}^T\{\tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1}\tilde{S}_2\tilde{S}_2^T\tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T\tilde{\Sigma}_2^{-1}\tilde{S}_2}\}\tilde{\Sigma}_2\{\tilde{\Sigma}_2^{-1} \\ &- \frac{\tilde{\Sigma}_2^{-1}\tilde{S}_2\tilde{S}_2^T\tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T\tilde{\Sigma}_2^{-1}\tilde{S}_2}\}\{\sqrt{n}\tilde{Q}_{1n}(\bar{\beta}_0, 0)\} + o_p(1) \\ =&(\tilde{\Sigma}_2^{-1/2}\tilde{\boldsymbol{W}}_2)^T\{I_{2\times 2} - \frac{\tilde{\Sigma}_2^{-1/2}\tilde{S}_2\tilde{S}_2^T\tilde{\Sigma}_2^{-1/2}}{\tilde{S}_2^T\tilde{\Sigma}_2^{-1}\tilde{S}_2}\}(\tilde{\Sigma}_2^{-1/2}\tilde{\boldsymbol{W}}_2) + o_p(1), \end{aligned} \tag{A.36}$$

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

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

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

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

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

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

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

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 $n \to \infty$.

Therefore, it follows from 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. □

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 noncentral parameters in Theorems 3(ii) and 4(ii), we use the facts that

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

and

$$\begin{aligned} &\tilde{\boldsymbol{D}}_2^T\left\{\tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1}\tilde{S}_2\tilde{S}_2^T\tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T\tilde{\Sigma}_2^{-1}\tilde{S}_2}\right\}\tilde{\boldsymbol{D}}_2 \\ &= d_1^2\beta_0^2\tilde{S}_2^T\left\{\tilde{\Sigma}_2^{-1} - \frac{\tilde{\Sigma}_2^{-1}\tilde{S}_2\tilde{S}_2^T\tilde{\Sigma}_2^{-1}}{\tilde{S}_2^T\tilde{\Sigma}_2^{-1}\tilde{S}_2}\right\}\tilde{S}_2 = 0. \end{aligned}$$

□

Acknowledgments

We thank the editor, Professor Jianqing Fan, an associate editor, and one reviewer for their insightful and constructive comments that greatly improve the overall presentation.

Funding

Yang's research was partly supported by Humanity and Social Science Youth Foundation of Ministry of Education of China (19C10421028), NSF of China (71401066), and the Major Program of the National Social Science Foundation of China (17ZDA073). Liu's research was partly supported by NSF of China (11971208, 11601197), China Postdoctoral Science Foundation funded project (2016M600511, 2017T100475), and NSF of Jiangxi Province (2018ACB21002, 20171ACB21030). Peng's research was partly supported by the Simons Foundation. Cai's research was partly supported by NSF of China (71631004, key project).

References

Amihud, Y., and Hurvich, C. M. (2004), "Predictive Regressions: A Reduced-Bias Estimation Method," *Journal of Financial and Quantitative Analysis*, 39, 813–841. [685]

Bickel, P., and Doksum, K. A. (2001), *Mathematical Statistics I* (2nd ed.), Upper Saddle River, NJ: Prentice Hall. [686]

Billingsley, P. (1999), *Convergence of Probability Measures* (2nd ed.), New York: Wiley. [695]

Breitung, J., and Demetrescu, M. (2015), "Instrumental Variable and Variable Addition Based Inference in Predictive Regressions," *Journal of Econometrics*, 187, 358–375. [684,685]

Cai, Z., and Wang, Y. (2014), "Testing Predictive Regression Models With Nonstationary Regressors," *Journal of Econometrics*, 178, 4–14. [684,685,691,693]

Campbell, J. Y. (2008), "Viewpoint: Estimating the Equity Premium," *Canadian Journal of Economics/Revue canadienne d'économique*, 41, 1–21. [684]

Campbell, J. Y., and Yogo, M. (2006), "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27–60. [684,685,693]

Cavanagh, C. L., Elliott, G., and Stock, J. H. (1995), "Inference in Models With Nearly Integrated Regressors," *Econometric Theory*, 11, 1131–1147. [684,685]

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. [685]

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. [685]

Choi, Y., Jacewitz, S., and Park, J. Y. (2016), "A Reexamination of Stock Return Predictability," *Journal of Econometrics*, 192, 168–189. [685]

Demetrescu, M. (2014), "Enhancing the Local Power of IVX Based Tests in Predictive Regressions," *Economics Letters*, 124, 269–273. [685,686,689,690,691,692,693]

Demetrescu, M., and Rodrigues, P. M. M. (2016), "Residual-Augmented IVX Predictive Regression," Working Paper, Economics and Research Department, Banco de Portugal. [684]

Fama, E. F., and French, K. R. (2010), "Luck Versus Skill in the Cross-Section of Mutual Fund Returns," *Journal of Finance*, 65, 1915–1947. [686]

Jansson, M., and Moreira, M. J. (2006), "Optimal Inference in Regression Models With Nearly Integrated Regressors," *Econometrica*, 74, 681–714. [684]

Jensen, M. C. (1968), "The Performance of Mutual Funds in the Period 1945–1964," *Journal of Finance*, 23, 389–416. [686]

Jing, B. Y., Yuan, J., and Zhou, W. (2009), "Jackknife Empirical Likelihood," *Journal of American Statistical Association*, 104, 1224–1232. [687]

Kosowski, R., Timmermann, A., Wermers, R. and White, H. (2006), "Can Mutual Fund "Stars" Really Pick Stocks? New Evidence From a Bootstrap Analysis," *Journal of Finance*, 61, 2551–2595. [686]

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015), "Robust Econometric Inference for Stock Return Predictability," *Review of Financial Studies*, 28, 1506–1553. [684,685,686,689,690,691,692,693]

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. [687]

Liu, J. M., Chen, R., and Yao, Q. (2010), "Nonparametric Transfer Function Models," *Journal of Econometrics*, 157, 151–164. [687]

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. [685]

Owen, A. (2001), *Empirical Likelihood*, New York: Chapman and Hall. [686,697]

Phillips, P. C. B. (1987), "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547. [695]

——— (2015), "Pitfalls and Possibilities in Predictive Regression," *Journal of Financial Econometrics*, 13, 521–555. [684]

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. [684,685]

Phillips, P. C. B., and Magdalinos, T. (2007), "Limit Theory for Moderate Deviations From a Unit Root," *Journal of Econometrics*, 136, 115–130. [685]

Qin, J., and Lawless, J. (1994), "Empirical Likelihood and General Estimating Functions," *Annals of Statistics*, 22, 300–325. [686,697]

Stambaugh, R. F. (1999), "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421. [685]

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. [687]

Zhu, F., Cai, Z., and Peng, L. (2014), "Predictive Regressions for Macroeconomic Data," *Annals of Applied Statistics*, 8, 577–594. [685,686,687,689,693]