JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2023, VOL. 41, NO. 1, 228–240
https://doi.org/10.1080/07350015.2021.2008406

Taylor & Francis
Taylor & Francis Group

Testing for Structural Change of Predictive Regression Model to Threshold Predictive Regression Model

Fukang Zhu[a], Mengya Liu[b], Shiqing Ling[c], and Zongwu Cai[d]

[a]School of Mathematics, Jilin University, Changchun, Jilin, China; [b]School of Mathematics, Central China Normal University, Wuhan, Hubei, China; [c]Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, China; [d]Department of Economics, University of Kansas, Lawrence, KS

ABSTRACT

This article investigates two test statistics for testing structural changes and thresholds in predictive regression models. The generalized likelihood ratio (GLR) test is proposed for the stationary predictor and the generalized *F* test is suggested for the persistent predictor. Under the null hypothesis of no structural change and threshold, it is shown that the GLR test statistic converges to a function of a centered Gaussian process, and the generalized *F* test statistic converges to a function of Brownian motions. A Bootstrap method is proposed to obtain the critical values of test statistics. Simulation studies and a real example are given to assess the performances of the proposed tests.

ARTICLE HISTORY
Received December 2020
Accepted November 2021

KEYWORDS
Bootstrap method; Generalized *F*-test; Generalized likelihood ratio test; Predictive regression; Structural change; Threshold model

1. Introduction

The question of whether asset returns are predictable or not is one of the most studied and contentious issues in financial economics. Predictors commonly considered for returns include various lagged financial variables, such as log dividend-price ratio, log earnings-price ratio, log book-to-market ratio, dividend yield, term spread, term structure of interest rates, and default premia. Predictive regression (PR) is a conventional method to check whether some financial variables have the explanatory power on the stock return predictability. It is extensively used in studies of mutual fund performance, conditional capital asset pricing and optimal asset allocation; see, for example, the survey article by Liao, Cai, and Chen (2018). The classical PR model takes the following simple form

$$y_t = \psi_0 + \psi_1 x_{t-1} + u_t, \tag{1}$$

where u_t is commonly assumed to be an independent and identically distributed (iid) innovation process with mean 0 and variance $\sigma^2 > 0$ and the predictor x_t is modeled as

$$x_t = \rho x_{t-1} + v_t, \tag{2}$$

where v_t is a zero-mean innovation process. For convenience, we set $x_0 = 0$. Inferences in PR models are complicated due to the joint interaction of the highly persistent nature ($|\rho|$ in (2) can be either one or very close to one) of the commonly used predictors with endogeneity problems arising from the correlation of u_t and v_t. There is growing literature aiming to developing valid and reliable inferences for such settings, see Campbell and Yogo (2006), Jansson and Moreira (2006), Cai and Wang (2014), Zhu, Cai, and Peng (2014a), and Breitung and Demetrescu (2015), Yang et al. (2020, 2021), among others. For more recent developments in this area, the reader is referred to the survey article by Liao, Cai, and Chen (2018).

It is well documented that predictability may be time-varying and the impact of predictors may be evolving over time. In a comprehensive study on the predictability of the equity premium, Welch and Goyal (2008) found significant instabilities in predictability as highlighted by others. The sensitivity analysis conducted in Kostakis, Magdalinos, and Stamatogiannis (2015) also highlighted significant variations in test conclusions which depend on whether one considers pre- or post-50s data. There are two common modeling tools to deal with parameter instability in PR models: structural change and threshold model.

Testing for structural changes has always been an important issue in econometrics because a myriad of political and economic factors can cause the relationships among economic variables to change over time. The great depression, oil price shocks, technical progress, and abrupt policy and regulations changes all are such examples. The earliest references go back to Chow (1960) and Quandt (1960). The Chow's test assumes that the time of structural change is known a priori, while Quandt's test takes the largest Chow test statistic over all possible times of the structural change. This type of test statistics needs a normalization and has a Darling-Erdös-type limit, see Horváth (1993) and Ling (2007). Another method is to restrict the change-point interval $(0, 1)$ to a closed subinterval, see Andrews (1993) and Bai and Perron (1998). Recently, Pitarakis (2017) studied two cumulative squared residuals-based tests and Georgiev et al. (2018) considered the *SupF* and Cramér-von-Mises type statistics for a change point in PR models.

As pointed by Gonzalo and Pitarakis (2012), the predictive impact of a variable may alternate in strength across different episodes (e.g., periods of rapid versus slow growth, periods of

CONTACT Zongwu Cai *caiz@ku.edu* Department of Economics, University of Kansas, 1460 Jayhawks Blvd, Lawrence, KS 66045.

high versus low stock market valuation, periods of high versus low consumer confidence). Ignoring such phenomena by proceeding within a linear framework may mask the forecasting ability of a particular variable and the presence of interesting and economically meaningful dynamics. To this end, Gonzalo and Pitarakis (2012) investigated the threshold PR (TPR) model as follows:

$$y_t = \psi_0 + \psi_1 x_{t-1} + (\phi_0 + \phi_1 x_{t-1}) I(q_{t-1} \leq r) + u_t, \quad (3)$$

where q_t is a threshold (stationary) variable and r is an unknown threshold parameter, and they found that a threshold variable provides a trigger to predictability. Besides the previous methods, some researchers believe that parameters should smoothly change, such as Cai, Wang, and Wang (2015) which developed a test against smooth parameter variation in the parameters of the PR model.

The evidence for both structural changes and threshold effects for the predictability of stock returns has been well documented in the literature. For example, Paye and Timmermann (2006) examined evidence of instability in models of ex post predictable components in stock returns related to structural changes in the coefficients of state variables such as the lagged dividend yield, short interest rate, term spread, and default premium. Also, Rapach and Wohar (2006) investigated the structural stability of predictive regression models of U.S. quarterly aggregate real stock returns over the postwar era and found strong evidence of structural changes in S&P 500 returns. Furthermore, Pettenuzzo and Timmermann (2011) found a strong evidence of multiple breaks in return prediction models based on the dividend yield or a short interest rate. Currently, Smith and Timmermann (2021) developed a new approach to modeling and predicting stock returns in the presence of breaks that simultaneously affect a large cross-section of stocks and found that out-of-sample return forecasts are significantly more accurate than those from existing approaches. Per threshold effect, McMillan (2001) found threshold effects between stock market returns and interest rates. Recently, Gonzalo and Pitarakis (2012, 2017), applied model (3) to the prediction of stock returns with dividend yields and found the presence of regimes in which predictability kicks in solely during bad economic times, as a result, their analysis illustrated the fact that the presence of regimes may make predictability appear as nonexistent when assessed within a linear model. Finally, Kiliç (2018) revealed presence of asymmetric regime-dependence (threshold effects) and variability in the strength and size of predictability across asset-related (e.g., dividend/price ratio) versus other (e.g., default yield spread) predictors.

Existing test statistics are just for the change of parameters or the form of the model (i.e., threshold). The co-existence of threshold effects and structural changes has been documented in the literature. The spread between long- and short-term interest rates is often used to predict recessions. Based on the evidence in the literature that the spread predicts negative output growth but is not useful when there is a boom jointly with the evidence that the spread could have lost its predictive power and the fact that the volatility of output growth has decreased, Galvão (2006) found that the timing of the 2001 U.S. recession can be anticipated correctly using the spread as the leading indicator when dealing with nonlinearity and a structural change simultaneously. Okun's law[1] refers to the empirical observation that there is an inverse relationship between output and unemployment gaps, or between cyclical unemployment and cyclical output. Based on the Hodrick-Prescott[2] and band-pass filtered data for Canada, Huang and Chang (2005) found strong support of structural change as well as threshold nonlinearity when reevaluating the empirical validity of Okun's law.

In practice, the structure of a model may be changed in terms of both time horizon and states. For example, the threshold AR-ARCH models were used to fit 11 nonoverlapping 2-year period Hong Kong Hang Seng index from 1970 to 1991, but Wong and Li (1997) found that 2 out of 11 periods should follow the AR-ARCH models. When the form of models is changed, the conventional likelihood ratio test for the standard change-point problem is not locally most powerful any more. Therefore, how to efficiently detect the change of the structural forms of the time series models is of great interest. Indeed, Berkes et al. (2011) and Zhu and Ling (2012) considered this kind of problem in autoregressive models and provided several supporting real examples.

In this article, we propose two test statistics for testing structural changes and thresholds in PR models. The generalized likelihood ratio (GLR) test is proposed for the stationary predictor and the generalized F (GF) test is suggested for the persistent predictor. Under the null hypothesis of no structural change and threshold, it is shown that the GLR test statistic converges to a function of a centered Gaussian process, and the GF test statistic converges to a function of Brownian motions. Finally, to make the proposed tests useful in practice, a Bootstrap method is proposed to obtain the critical values of test statistics.

This article is organized as follows. Section 2 presents the test statistics and their limiting distributions. Section 3 discusses a Bootstrap method to obtain the critical values of test statistics. Simulation studies and a real example are given to assess the finite sample performances of the proposed tests in Sections 4 and 5, respectively. Section 6 concludes the article. All detailed proofs are given in Appendix.

2. Test Statistics and Main Results

To establish our setting, let $\{(y_t, x_t, q_t), t = 1, \ldots, n\}$ be a sequence of observations from the following model

$$y_t = \begin{cases} \psi_0 + \psi_1 x_{t-1} + u_t, & t = 1, \ldots, k, \\ \psi_0 + \psi_1 x_{t-1} + (\phi_0 + \phi_1 x_{t-1}) & \\ \qquad I(q_{t-1} \leq r) + u_t, & t = k+1, \ldots, n, \end{cases} \quad (4)$$

where x_t is defined as in (2). We consider the null and alternative hypotheses as follows:

$$H_0 : k = n \quad \text{versus} \quad H_1 : k = k^* \text{ and } \phi \neq 0, \quad (5)$$

where $0 < k^* < n$ and $\phi = (\phi_0, \phi_1)^\top$. The model in (4) is a PR model under H_0 and it changes to a TPR model after the time k^*

[1]In economics, Okun's law is an empirically observed relationship between unemployment and losses in a country's production.

[2]The Hodrick-Prescott filter (also known as Hodrick-Prescott decomposition) is a mathematical tool used in macroeconomics, especially in real business cycle theory, to remove the cyclical component of a time series from raw data.

under H_1. Under H_0, no change has occurred and no threshold effect exists, while under H_1, a change has occurred at time k^* (if $k^* > 0$) and a threshold effect exists. When $r = \infty$, (5) is for testing the parameter change in the PR model, which was studied recently by Pitarakis (2017) and Georgiev et al. (2018). When $k = 0$, (5) is for testing the threshold in the TPR model investigated by Gonzalo and Pitarakis (2012, 2017). Therefore, under H_0, both k^* and r are absent, which is the main difficulty in our setting.

Let $Y = (y_1, \dots, y_n)^\top, X_t = (1, x_{t-1})^\top, X = (X_1, \dots, X_n)^\top, u = (u_1, \dots, u_n)^\top, \psi = (\psi_0, \psi_1)^\top$, and $\theta = (\psi^\top, \phi^\top)^\top$. Denote $X_{kr} = \operatorname{diag}\{I(q_{i-1} \le r, i > k), i = 1, \dots, n\}X$ and $\widetilde{X}_{kr} = (X, X_{kr})$. Here, X and $\widetilde{X}_{kr}$ are $n \times 2$ and $n \times 4$, respectively. Under H_0, clearly, the model in (4) becomes to $Y = X\psi + u$, while under H_1, it reduces to $Y = \widetilde{X}_{kr}\theta + u$. Therefore, the least-squares based estimates of σ^2 under H_0 and H_1 are given by

$$\hat{\sigma}_n^2 = \frac{1}{n}[Y^\top Y - (Y^\top X)(X^\top X)^{-1}(X^\top Y)] \equiv \frac{1}{n}\mathrm{RSS}_0 \quad (6)$$

and

$$\hat{\sigma}_n^2(k, r) = \frac{1}{n}[Y^\top Y - (Y^\top \widetilde{X}_{kr})(\widetilde{X}_{kr}^\top \widetilde{X}_{kr})^{-1}(\widetilde{X}_{kr}^\top Y)] \equiv \frac{1}{n}\mathrm{RSS}_1, \quad (7)$$

respectively, where RSS_0 and RSS_1 are the residual sum of squares (RSS) under H_0 and H_1, respectively.

2.1. Stationary Predictor

This section is devoted to the case when the predictor x_t is stationary. That is, $|\rho| < 1$, where ρ is given in (2). Let the filtration $\mathcal{F}_t$ be the σ-field generated by $\{(u_i, v_i, q_i) : 1 \le i \le t\}$ and $F(\cdot)$ be the distribution function of q_t. We impose the following assumptions.

Assumption 1. The process $\{(X_t u_t, q_{t-1})\}$ is a stationary α-mixing process with geometric rate, and $|\rho| < 1$, where $X_t = (1, x_{t-1})^\top$ and ρ is given in (2).

Assumption 2. x_t is stationary and ergodic, and $E(u_t|\mathcal{F}_{t-1}) = 0$ and $0 < E[\|X_t u_t\|^{2+\delta}] < \infty$ for some $\delta > 0$, where $\|\cdot\|$ denotes the Euclidean norm.

Assumption 3. $F(\cdot)$ is continuous and strictly increasing and the corresponding density function is bounded away from zero and ∞ over each bounded set.

Assumption 1 implies that $\{y_t\}$ is a stationary α-mixing process with geometric rate and Assumption 2 allows that u_t is a martingale difference sequence (MDS), and it permits correlation between innovations u_t and v_t. The moment restriction in Assumption 2 is more relaxed than the usual $4 + \delta$ moment restriction in the literature.

For fixed k and r, the GLR-type statistic for testing (5) is given by

$$\mathrm{GLR}_n(k, r) = n\left(\ln \hat{\sigma}_n^2 - \ln \hat{\sigma}_n^2(k, r)\right), \quad (8)$$

which is the exact likelihood ratio if the distribution of u_t is $N(0, \sigma^2)$; see, for example, Berkes et al. (2011) and Zhu and Ling (2012) for details. Note that $\mathrm{GLR}_n(k, r)$ in (8) can be expressed as

$$\mathrm{GLR}_n(k, r) = n \ln\left(\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1} + 1\right) \approx n\left(\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1}\right),$$

and by Taylor expansion, it can be approximated by the F-type test statistic, termed as generalized F-test (GF) in Cai and Tiwari (2000) and Cai et al. (2019); see (10). Therefore, the GLR test is asymptotically equivalent to the GF test. Since the exact values of k and r are unknown under H_0, it is natural to construct our test by using the maxima of $\mathrm{GLR}_n(k, r)$ on the ranges of k and r. That is, our test statistic is defined as

$$\mathrm{GLR}_n = \sup_{0 \le k \le k_1} \sup_{r_1 \le r \le r_2} \mathrm{GLR}_n(k, r),$$

where $-\infty < r_1 < r_2 < \infty$, $k_1 = \lfloor n\pi_1 \rfloor$ with $0 < \pi_1 < 1$, and $\lfloor n\pi_1 \rfloor$ denotes the integer part of $n\pi_1$.

Let $I_{t-1}^* = I(q_{t-1} \le r)$. Then, $I_{t-1}^* = I(F(q_{t-1}) \le F(r)) \equiv I(U_{t-1} \le \lambda)$, where $\lambda \equiv F(r)$ and U_t is the uniformly distributed random variable on $[0, 1]$. Define $\lambda_1 \equiv F(r_1)$ and $\lambda_2 \equiv F(r_2)$. Then, $\lambda \in [\lambda_1, \lambda_2]$. Note that the range of π is $[0, \pi_1]$. The following theorem gives the limiting distribution of GLR_n.

Theorem 1. If Assumptions 1–3 hold, then under H_0, it follows that

$$\mathrm{GLR}_n \xrightarrow{d} \sup_{0 \le \pi \le \pi_1} \sup_{\lambda_1 \le \lambda \le \lambda_2} \xi_{\pi\lambda}^\top M_{\pi\lambda}^{-1} \xi_{\pi\lambda} \ \textit{as } n \to \infty,$$

where $\xrightarrow{d}$ denotes convergence in distribution, $M_{\pi\lambda} = \Sigma_{\pi\lambda} - \Sigma_{\pi\lambda}\Sigma^{-1}\Sigma_{\pi\lambda}$, $\Sigma = E(X_t X_t^\top)$, $\Sigma_{\pi\lambda} = (1 - \pi)E(I_{t-1}^* X_t X_t^\top)$, $\xi_{\pi\lambda}$ is a Gaussian process with mean zero and covariance kernel

$$\mathrm{cov}(\xi_{\pi\lambda}, \xi_{\pi^*\lambda^*}) = \Sigma_{\max(\pi,\pi^*),\min(\lambda,\lambda^*)} - \Sigma_{\pi\lambda}\Sigma^{-1}\Sigma_{\pi^*\lambda^*}.$$

If x_t and q_t are independent, then $\Sigma_{\pi\lambda} = a_{\pi\lambda}\Sigma$ with $a_{\pi\lambda} = (1 - \pi)\lambda$. Thus, $M_{\pi\lambda} = a_{\pi\lambda}(1 - a_{\pi\lambda})\Sigma$ and the covariance kernel of $\xi_{\pi\lambda}$ is $[a_{\max(\pi,\pi^*),\min(\lambda,\lambda^*)} - a_{\pi\lambda}a_{\pi^*\lambda^*}]\Sigma$.

Corollary 1. Under the assumption of Theorem 1, if x_t and q_t are independent, then

$$\mathrm{GLR}_n \xrightarrow{d} \sup_{0 \le \pi \le \pi_1} \sup_{\lambda_1 \le \lambda \le \lambda_2} \frac{\xi_{\pi\lambda}^{*\top}\xi_{\pi\lambda}^*}{a_{\pi\lambda}(1 - a_{\pi\lambda})}$$

as $n \to \infty$, where $\xi_{\pi\lambda}^*$ is a Gaussian process with mean zero and covariance kernel $\mathrm{cov}(\xi_{\pi\lambda}^*, \xi_{\pi^*\lambda^*}^*) = [a_{\max(\pi,\pi^*),\min(\lambda,\lambda^*)} - a_{\pi\lambda}a_{\pi^*\lambda^*}]I_2$.

Remark 1. Since $\lambda_2 < 1$, $a_{\pi\lambda}$ is always less than 1 such that the range of π can be $[0, \pi_1]$. If we allows $\lambda = \lambda_2 = 1$(i.e., the pure structural change case), we need to cut the range of π as $[\pi_0, \pi_1]$ with $0 < \pi_0 < \pi_1$. In this case, the limiting distribution is identical to the one in Theorem 3 (c) in Andrews (1993), that is, a quadratic form in normalized Brownian bridges: $\sup_{\pi_0 \le \pi \le \pi_1} \xi_{\pi 1}^{*\top}\xi_{\pi 1}^*/[\pi(1 - \pi)]$, where $\xi_{\pi 1}^*$ is a standard bivariate Brownian bridge. If $\pi = 0$ (i.e., the pure threshold case), then the limiting distribution is $\sup_{\lambda_1 \le \lambda \le \lambda_2} \xi_{0\lambda}^{*\top}\xi_{0\lambda}^*/[\lambda(1 - \lambda)]$, where $\xi_{0\lambda}^*$ is a standard bivariate Brownian bridge, and this distribution is identical to the one in Equation (4) in Gonzalo and Pitarakis (2012), but the difference is that they considered a Wald type test statistic when the persistent predictor $\rho = \rho_n = 1 - c/n$ for some $c > 0$, which will be discussed in the next section.

2.2. Persistent Predictor

This section considers the test statistic for (5) when x_t is persistent, that is, unit root (or integrated or $\rho = 1$) and near-integrated (or local-to-unity or $\rho = \rho_n = 1 - c/n$ for some $c > 0$) predictor. Let $w_t = (u_t, v_t)^\top$ and the filtration $\mathcal{F}_t = \sigma(w_s, q_s | s \le t)$. We impose the following maintained conditions on the model.

Assumption 4. $E(w_t|\mathcal{F}_{t-1}) = 0, E(w_t w_t^\top|\mathcal{F}_{t-1}) > 0, \sup_t Eu_t^4 < \infty$ and $\sup_t Ev_t^4 < \infty$.

Assumption 5. The threshold variable q_t is strictly stationary, ergodic, and strong mixing with mixing coefficients α_m satisfying $\sum_{m=1}^{\infty} \alpha_m^{1/2-1/\kappa} < \infty$ for $\kappa > 2$.

Again, w_t in Assumption 4 is a MDS, so that it can be used to establish the weak convergence result of the empirical process $\sum_{t=1}^{\lfloor ns \rfloor} I_{t-1}^* u_t$ as in Caner and Hansen (2001). It can be replaced by the following two assumptions: (i) $\{u_t, \mathcal{F}_t\}$ is a stationary and ergodic MDS with $E(u_t^2|\mathcal{F}_{t-1}) = \sigma^2$, $E(u_t) = 0$, and $E(u_t^4) < \infty$. (ii) $\sum_{t=1}^{\lfloor ns \rfloor} w_t/\sqrt{n} \Longrightarrow (B_u(s), B_v(s))^\top$ as $n \to \infty$, where $(B_u(s), B_v(s))$ is a vector of Brownian motions with a positive definite long-run covariance matrix and "$\Longrightarrow$" represents weak convergence. Similar assumptions were imposed by Chen (2015) by studying the robust estimation and inference of threshold models with integrated regressors. Assumption 5 is very conventional in the literature of threshold models. As pointed out by Gonzalo and Pitarakis (2012), Assumption 4 implies that a functional central limit theorem holds for the joint process w_t, that is, $\sum_{t=1}^{\lfloor ns \rfloor} w_t/\sqrt{n} \Longrightarrow (B_u(s), B_v(s))^\top$, with the long-run variance of the bivariate Brownian motion being given by $\sum_{k=-\infty}^{\infty} E(w_0 w_k^\top)$. Similar to Theorem 1 of Caner and Hansen (2001), we can show that

$$\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} v_t, \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} I_{t-1}^* u_t \right) \Longrightarrow [B_v(s), \sigma W(s, \lambda)] \text{ in } D^2[0, 1], \tag{9}$$

where $W(s, \lambda)$ is a two-parameter standard Brownian motion on $(s, \lambda) \in [0, 1]^2$ and $D[0, 1]$ denotes the Skorohod space. The two-parameter Brownian motion is a special tool to derive the limiting distribution in threshold models.

We define the generalized F-type statistic as follows:

$$\mathrm{GF}_n = \sup_{0 \le k \le k_1} \sup_{r_1 \le r \le r_2} \mathrm{GF}_n(k, r), \quad \text{where}$$

$$\mathrm{GF}_n(k, r) = n \left(\frac{\mathrm{RSS}_0 - \mathrm{RSS}_1}{\mathrm{RSS}_1} \right), \tag{10}$$

from which, we can see that this statistic is equivalent to $\mathrm{GLR}_n(k, r)$ in (8) in Section 2.1 if x_t is stationary. Assume that x_t is persistent, that is,

$$x_t = \rho_n x_{t-1} + v_t, \ \rho_n = 1 - c/n, \ c \ge 0. \tag{11}$$

We also make use of the diffusion process $K_c(s) = \int_0^s e^{c(s-u)} dB_v(u)$ with $K_c(s)$ such that $dK_c(s) = cK_c(s) + dB_v(s)$ and $K_c(0) = 0$. Note that we can also write

$$K_c(s) = B_v(s) + c \int_0^s e^{c(s-u)} B_v(u) du.$$

It is obvious that when $c = 0$, $K_c(s)$ becomes to $B_v(s)$ given in (9). More properties about $K_c(s)$ can be found in Section 3 of Phillips (1988). Under our assumptions, it follows directly from Lemma 3.1 in Phillips (1988) that $x_{\lfloor ns \rfloor}/\sqrt{n} \Longrightarrow K_c(s)$. The following theorem establishes the asymptotic behavior of GF_n when x_t is a persistent process.

Theorem 2. If Assumptions 3–5 hold and x_t is given by (11), then under H_0, we have

$$\mathrm{GF}_n \xrightarrow{d} \sup_{0 \le \pi \le \pi_1} \sup_{\lambda_1 \le \lambda \le \lambda_2} \zeta_{\pi\lambda}^\top L_{\pi\lambda}^{-1} \zeta_{\pi\lambda}$$

as $n \to \infty$, where $L_{\pi\lambda} = \lambda \Xi_\pi - \lambda^2 \Xi_\pi \Xi_0^{-1} \Xi_\pi$, $\zeta_{\pi\lambda} = \int_\pi^1 \overline{K}_c(s) dW(s, \lambda) - \lambda \Xi_\pi \Xi_0^{-1} \int_0^1 \overline{K}_c(s) dW(s, 1)$, and $\Xi_\pi = \int_\pi^1 \overline{K}_c(s) \overline{K}_c^\top(s) ds$, $\overline{K}_c(s) = (1, K_c(s))^\top$.

The following corollary gives a special case of Theorem 2, that is, $c = 0$ and x_t is a unit root process.

Corollary 2. When $c = 0$, that is, x_t is a unit root process, then the limiting distribution in Theorem 2 reduces to $\sup_{0 \le \pi \le \pi_1} \sup_{\lambda_1 \le \lambda \le \lambda_2} \eta_{\pi\lambda}^\top N_{\pi\lambda}^{-1} \eta_{\pi\lambda}$, where $N_{\pi\lambda} = \lambda \Omega_\pi - \lambda^2 \Omega_\pi \Omega_0^{-1} \Omega_\pi$, $\eta_{\pi\lambda} = \int_\pi^1 \overline{B}_v(s) dW(s, \lambda) - \lambda \Omega_\pi \Omega_0^{-1} \int_0^1 \overline{B}_v(s) dW(s, 1)$, and $\Omega_\pi = \int_\pi^1 \overline{B}_v(s) \overline{B}_v^\top(s) ds$, $\overline{B}_v(s) = (1, B_v(s))^\top$.

Remark 2. When $\pi = 0$, Theorem 2 includes Proposition 1 in Gonzalo and Pitarakis (2012) as a special case because we have the same null hypothesis (no threshold and structural change) although the alternative hypothesis is different. Note that the formulation of our model (4) is different from model (1) in Gonzalo and Pitarakis (2012). Unlike the case with stationary predictor, $E(I_{t-1}^* X_t X_t^\top)$ cannot be simplified to $\lambda \Sigma$ when x_t and q_t are independent. A similar discussion about PR model without structural change was given by Gonzalo and Pitarakis (2012, p. 232).

3. Bootstrapped Critical Values

We have seen in the previous section that structural change or nonstationarity affects the asymptotic distributions of the test statistics in complicated ways. Under the null, the test statistic has a limiting distribution given by a functional of a zero-mean Gaussian process whose covariance function depends on the data generating process (DGP). The asymptotic critical values thus depend on the DGP and cannot be tabulated. In this section, we use a Bootstrap method to obtain the critical value for our tests, see Hansen (1996) and Zhu, Yu, and Li (2014b) in different settings.

We first give some notations. Let $z_t(k, r) = (X_t^\top, I(q_t \le r, t > k) X_t^\top)^\top$ and $u_t(\theta, k, r) = y_t - z_t^\top(k, r)\theta$. Then,

$$\theta_n(k, r) \equiv \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} u_t^2(\theta, k, r) = (\widetilde{X}_{kr}^\top \widetilde{X}_{kr})^{-1} \widetilde{X}_{kr}^\top Y,$$

where $\widetilde{X}_{kr} = (X, X_{kr}) = (z_1(k, r), \ldots, z_n(k, r))^\top$ and Θ is a compact parameter space of θ. Assume that $\{\varepsilon_t\}_{t=1}^n$ is a sequence of iid $N(0, 1)$ random variables. Let $\hat{u} = (\hat{u}_1 \varepsilon_1, \ldots, \hat{u}_n \varepsilon_n)^\top$, where $\hat{u}_t = y_t - z_t^\top(k, r)\theta(k, r)$. By using $\{\hat{u}_t\}$, we will construct our Bootstrap statistics in this section.

3.1. Stationary Predictor

Denote $T^{\star}_{kr} = n^{-1/2}[X^{\top}_{kr} - X^{\top}_{kr}X(X^{\top}X)^{-1}X^{\top}]\hat{u}$. For each (k, r), we define the Bootstrap GLR test statistic as follows:

$$\widehat{\mathrm{GLR}}_n(k, r) = T^{\star\top}_{kr}\left\{\frac{X^{\top}_{kr}X_{kr}}{n} - \frac{X^{\top}_{kr}X}{n}\left(\frac{X^{\top}X}{n}\right)^{-1}\frac{X^{\top}X_{kr}}{n}\right\}^{-1}T^{\star}_{kr}/\hat{\sigma}^2_n.$$

Note that $T^{\star}_{kr}$ can also be written as

$$\left(-\frac{X^{\top}_{kr}X}{n}\left(\frac{X^{\top}X}{n}\right)^{-1}, I_2\right)\hat{Z}_n(k, r) \text{ with } \hat{Z}_n(k, r) = \frac{1}{\sqrt{n}}\widetilde{X}^{\top}_{kr}\hat{u} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t(k, r)\hat{u}_t\varepsilon_t.$$

Our Bootstrap GLR test statistic is defined as follows:

$$\widehat{\mathrm{GLR}}_n \equiv \sup_{0\le\pi\le\pi_1}\sup_{\lambda_1\le\lambda\le\lambda_2}\widehat{\mathrm{GLR}}^{*}_n(\pi, \lambda) \equiv \sup_{0\le k\le k_1}\sup_{r_1\le r\le r_2}\widehat{\mathrm{GLR}}_n(k, r). \tag{12}$$

Before stating our result, we mention the concept "weak convergence in probability," which generalizes convergence in distribution to allow for conditional (i.e., random) distribution functions. The asymptotic theory of $\widehat{\mathrm{GLR}}_n$ is stated in the following theorem, which says that the conditional limiting distribution is the same as the null distribution as in Theorem 1.

Theorem 3. If Assumptions 1–3 hold, then under H_0, it follows that

$$\widehat{\mathrm{GLR}}_n|w^*_1, \ldots, w^*_n \xrightarrow{d} \sup_{0\le\pi\le\pi_1}\sup_{\lambda_1\le\lambda\le\lambda_2}\xi^{\top}_{\pi\lambda}M^{-1}_{\pi\lambda}\xi_{\pi\lambda}$$

in probability as $n \to \infty$, where $w^*_t = (y_t, x_t, q_t), t = 1, \ldots, n$.

From Theorem 3, conditional on the data sample $\{w^*_1, \ldots, w^*_n\}$, our Bootstrap procedure to obtain the critical value at significance level α is as follows:

i. generate iid $N(0, 1)$ samples $\{\varepsilon_t\}^n_{t=1}$ and calculate $\widehat{\mathrm{GLR}}_n$ via (12);
ii. repeat step (i) J times to get $\{\widehat{\mathrm{GLR}}^{(1)}_n, \ldots, \widehat{\mathrm{GLR}}^{(J)}_n\}$;
iii. choose $c^J_{n,\alpha}$ as the α-th upper percentile of $\{\widehat{\mathrm{GLR}}^{(1)}_n, \ldots, \widehat{\mathrm{GLR}}^{(J)}_n\}$.

The critical value for the GLR test is $c^J_{n,\alpha}$. The following corollary guarantees that our Bootstrapped critical value $c^J_{n,\alpha}$ is asymptotically valid, which is called unconditional Bootstrap validity and its proof is similar to that for Corollary 2 in Zhu, Yu, and Li (2014b) and hence is omitted.

Corollary 3. If Assumptions 1–3 hold, under H_0, we have $\lim_{n\to\infty}\lim_{J\to\infty} P(\mathrm{GLR}_n \ge c^J_{n,\alpha}) = \alpha$.

3.2. Persistent Predictor

It is well known that standard Bootstrap techniques fail in the context of nonstationary AR models. One has to employ the subsample Bootstrap method which faces the difficult issue of choosing the subsample size. When testing for structural change in conditional models, the asymptotic distributions of some statistics are not invariant to structural change in the regressors. To solve the size problem, Hansen (2000) proposed the *fixed regressor Bootstrap* method to achieve the first-order asymptotic distribution and it possess reasonable size properties in small samples. 2018; 2019 established the validity of this Bootstrap method for test statistics for the parameter instability in predictive regression models.

Using a similar idea as that in Section 3.1, we construct the Bootstrap GF test as follows:

$$\widehat{\mathrm{GF}}_n = \sup_{0\le k\le k_1}\sup_{r_1\le r\le r_2} T^{\star\top}_{kr}R^{-1}_{kr}T^{\star}_{kr}/\hat{\sigma}^2_n(k, r),$$

where $T^{\star}_{kr} = D_nX^{\top}_{kr}\hat{u} - D_nX^{\top}_{kr}XD_n(D_nX^{\top}XD_n)^{-1}D_nX^{\top}\hat{u}$ and R_{kr} is given in the proof of Theorem 2, that is, $R_{kr} = D_nX^{\top}_{kr}X_{kr}D_n - D_nX^{\top}_{kr}XD_n(D_nX^{\top}XD_n)^{-1}D_nX^{\top}X_{kr}D_n$ and D_n $=\mathrm{diag}\{1/\sqrt{n}, 1/n\}$. To formulate a useful asymptotic result, a weaker convergence mode is required, that is, the terminology of weak convergence of random measures, which is weaker than weak convergence in probability. It reduces to the latter when the limit distribution is nonrandom. For the definition of this terminology and its application to Bootstrap, we refer to the article by Cavaliere and Georgiev (2020). Following Georgiev et al. (2019), we need to strengthen Assumption 4 as follows:

Assumption 6. Assumption 4 holds together with the following conditions: (a) w_t is drawn from a doubly infinite strictly stationary and ergodic sequence $\{w_t\}^{\infty}_{t=-\infty}$ which is a MDS with respect to its own past. (b) $\{u_t\}^{\infty}_{t=-\infty}$ is a MDS also with respect to $\mathcal{G}^1_t \vee \mathcal{G}^2_t$, where $\mathcal{G}^1_t$ and $\mathcal{G}^2_t$ are the σ-algebras generated by $\{v_t\}^{\infty}_{t=-\infty}$ and $\{u_t\}^{\infty}_{t=-\infty}$, respectively, and $\mathcal{G}^1_t \vee \mathcal{G}^2_t$ denotes the smallest σ-algebra containing both $\mathcal{G}^1_t$ and $\mathcal{G}^2_t$.

The asymptotic theory of our Bootstrapped statistic $\widehat{\mathrm{GF}}_n$ is given as follows, which is different from Theorem 3.

Theorem 4. If Assumptions 3–6 hold and x_t is given by (11), then under H_0, it follows that

$$\widehat{\mathrm{GF}}_n|w^*_1, \ldots, w^*_n \longrightarrow \sup_{0\le\pi\le\pi_1}\sup_{\lambda_1\le\lambda\le\lambda_2}\zeta^{\top}_{\pi\lambda}L^{-1}_{\pi\lambda}\zeta_{\pi\lambda}\Big|B_v$$

as $n \to \infty$, where $w^*_t = (y_t, x_t, q_t)$, $t = 1, \ldots, n$ and B_v is given in (9).

Similar to the Bootstrap procedure in Section 3.1, we first obtain the Bootstrap sample $\{\hat{W}^{(1)}_n, \ldots, \hat{W}^{(J)}_n\}$ and then use its α-th upper percentile $d^J_{n,\alpha}$ as the approximating critical value of the GF test statistic GF_n. The following corollary guarantees that our Bootstrapped critical value $d^J_{n,\alpha}$ is asymptotically valid, which is called conditional Bootstrap validity and its proof is similar to that for Corollary 1 in Georgiev et al. (2019) and hence is omitted.

Corollary 4. If Assumptions 3–6 hold, under H_0, then $\lim_{n\to\infty}\lim_{J\to\infty} P(\mathrm{GF}_n \ge d^J_{n,\alpha}|w^*_1, \ldots, w^*_n) = \alpha$.

4. Simulation Studies

In this section, we report the performances of GLR_n and GF_n for both stationary and unit root cases in the finite sample, respectively.

We first study the size of two tests at the nominal $\alpha = 1\%$ and 5% level, respectively. The data are generated from the null model (4) with $(\psi_0, \psi_1) = (0.1, 0.2)$ and $k = n$, where the predictor x_t is modeled as $x_t = \rho x_{t-1} + \nu_t$ with $\rho = 0.2, 0.9$, and 1 and $\nu_t \sim N(0, 1)$, and the threshold variable $q_t \sim N(0, 1)$. Based on the Bootstrapped critical values, we consider rejection frequencies from 5000 replications with the sample sizes 200, 500, and 1000. The results are presented in Table 1. We find that the sizes of the tests are acceptable no matter in stationary or unit root cases. Additionally, both GLR_n and GF_n perform better when the sample size is not less than 500.

Next, in the stationary cases with $\rho = 0.2$ and 0.9, we explore the power of the GLR_n test against different alternative models, where ϕ_0 and ϕ_1 take values in $\{-2, -1, -0.6, 0.6, 1, 2\}$. In the nonstationary cases with $\rho = 1$, the power of GF_n test against different alternative models is considered, where ϕ_0 and ϕ_1 take values in $\{-1, -0.6, -0.2, 0.2, 0.6, 1\}$. At the same time, we consider whether our tests will be affected by the time of change k^* and the true threshold r in the stationary and nonstationary cases, that is, $(k^*, r) = (100, 0), (150, 0)$, and $(150, 0.2)$. The results of the stationary and unit root cases are summarized in Tables 2 and 3, respectively. The first point we should pay attention is that the farther (ϕ_0, ϕ_1) away from (0, 0), the stronger the power. But the power is seriously affected by ϕ_1. This is reasonable since the predictor x_t plays an important in the PR model. The second is that different k^* and r do not cause huge differences in power, which implies that our test is applicable and credible when the time of change and threshold are unknown.

It is interesting to study the local power properties of the proposed tests, for example, how is power affected if there exists a break and no thresholds or vice-versa. We consider two case: a structural change with no thresholds, a threshold with no structural changes, and other settings are given in Tables 4 and 5, respectively. If the data are generated by a structural change model, then the proposed test has some powers when $\phi_0 = 1$ or 2; while if the data are generated by a threshold model, then the powers of the proposed test are small, so the proposed test can detect breaks not thresholds well. A similar conclusion was made be Carrasco (2002), who showed that tests designed for a threshold alternative have power against parameter instability originating from structural change models, but tests for structural change have no power if the data are generated by a threshold model. In other words, testing only for a structural change might be very misleading and might result in adopting a linear model while the data are generated by another nonlinear model, but the stability test based on a misspecified threshold autoregressive model can detect parameter instability originating from structural change models, which suggest that the structural change model is easy to distinguish from the threshold model.

Table 1. Sizes (in percentage) of GLR_n and GF_n under the null model (4).

		$\alpha = 0.01$			$\alpha = 0.05$		
	n	200	500	1000	200	500	1000
Stationary case (GLR_n)	$\rho = 0.2$	1.17	0.90	1.05	5.23	5.12	5.03
	$\rho = 0.9$	0.82	1.08	1.04	4.81	4.92	5.01
Unit root case (GF_n)	$\rho = 1$	1.22	1.07	0.99	5.20	4.96	5.01

Table 2. Powers (in percentage) against model (4) with $(\psi_0, \psi_1) = (0.1, 0.2)$ and $x_t = 0.9x_{t-1} + \nu_t$.

	$\alpha = 0.01$						$\alpha = 0.05$					
	$k^* = 100, r = 0$						$k^* = 100, r = 0$					
$\phi_1 \backslash \phi_0$	−2.0	−1.0	−0.6	0.6	1.0	2.0	−2.0	−1.0	−0.6	0.6	1.0	2.0
−2.0	57.4	56.3	54.5	54.6	56.7	56.9	80.5	79.0	78.1	77.6	79.4	79.9
−1.0	23.9	20.4	17.8	16.9	19.7	24.0	46.0	38.6	36.9	37.9	38.8	45.7
−0.6	8.4	5.0	4.7	4.3	4.8	8.7	17.1	14.3	13.7	13.5	14.1	16.9
0.6	8.6	4.9	4.5	4.2	4.6	8.3	17.0	14.1	13.8	13.7	14.2	17.2
1.0	23.6	20.1	16.9	17.6	20.2	24.0	46.2	38.7	37.3	37.2	38.9	46.5
2.0	57.2	56.2	54.0	54.3	56.7	57.0	80.2	79.1	78.0	77.8	79.3	80.0
	$k^* = 150, r = 0$						$k^* = 150, r = 0$					
$\phi_1 \backslash \phi_0$	−2.0	−1.0	−0.6	0.6	1.0	2.0	−2.0	−1.0	−0.6	0.6	1.0	2.0
−2.0	57.2	56.2	54.1	54.5	56.8	57.3	80.5	78.7	78.4	77.9	79.2	80.4
−1.0	24.2	19.7	17.6	17.4	20.6	24.4	45.9	38.8	37.2	38.0	39.2	46.3
−0.6	8.5	4.8	4.7	4.4	4.7	8.5	17.2	14.2	13.6	13.8	13.9	17.3
0.6	8.3	4.8	4.5	4.3	4.5	8.7	17.1	14.0	13.5	13.6	14.3	17.2
1.0	24.0	20.3	16.7	17.9	20.4	23.8	46.2	38.9	37.4	37.4	39.3	46.0
2.0	57.0	56.4	54.3	54.4	56.7	57.1	80.3	79.0	78.3	77.4	79.1	79.8
	$k^* = 150, r = 0.2$						$k^* = 150, r = 0.2$					
$\phi_1 \backslash \phi_0$	−2.0	−1.0	−0.6	0.6	1.0	2.0	−2.0	−1.0	−0.6	0.6	1.0	2.0
−2.0	56.5	56.0	53.9	53.7	55.8	56.3	79.2	77.8	77.2	76.9	78.0	78.9
−1.0	23.2	19.0	16.7	15.7	18.6	23.5	44.8	37.4	34.8	35.7	37.8	44.9
−0.6	7.9	4.8	4.4	4.3	4.6	7.8	16.0	13.3	12.4	12.5	13.8	16.2
0.6	8.0	4.5	4.1	4.0	4.4	7.9	16.2	13.4	12.7	12.8	13.7	16.1
1.0	23.0	19.1	15.8	15.6	19.4	23.1	45.3	37.2	35.4	35.6	37.4	45.2
2.0	56.3	56.1	54.0	53.7	56.0	56.4	79.3	77.6	77.0	76.8	78.2	79.0

Table 3. Powers (in percentage) against model (4) with $(\psi_0, \psi_1) = (0.1, 0.2)$ and $x_t = x_{t-1} + \nu_t$.

	$\alpha = 0.01$							$\alpha = 0.05$					
	$k^* = 100, r = 0$							$k^* = 100, r = 0$					
$\phi_1 \backslash \phi_0$	-1.0	-0.6	-0.2	0.2	0.6	1.0	$\phi_1 \backslash \phi_0$	-1.0	-0.6	-0.2	0.2	0.6	1.0
-1.0	85.8	84.9	84.7	84.1	84.6	86.2	-1.0	99.1	97.8	97.3	97.2	98.0	99.4
-0.6	53.7	51.9	50.4	50.1	51.6	54.0	-0.6	73.2	72.3	72.6	71.7	72.5	73.1
-0.2	4.9	3.9	2.5	2.7	3.8	4.7	-0.2	10.4	9.6	8.5	8.9	9.5	10.7
0.2	5.0	4.4	2.4	2.8	3.6	4.9	0.2	10.8	9.7	8.4	8.7	9.6	10.5
0.6	53.8	52.6	51.4	50.6	52.8	53.4	0.6	73.6	72.4	71.9	71.6	72.4	73.0
1.0	85.6	84.2	84.4	84.6	84.2	86.8	1.0	99.3	98.5	97.2	97.4	98.9	99.6
	$k^* = 150, r = 0$							$k^* = 150, r = 0$					
$\phi_1 \backslash \phi_0$	-1.0	-0.6	-0.2	0.2	0.6	1.0	$\phi_1 \backslash \phi_0$	-1.0	-0.6	-0.2	0.2	0.6	1.0
-1.0	85.7	85.0	84.5	84.3	84.8	86.9	-1.0	99.2	98.7	97.4	97.3	98.8	99.6
-0.6	53.0	52.6	51.9	51.2	52.0	54.6	-0.6	73.4	72.5	71.8	71.4	72.6	73.0
-0.2	5.0	3.8	2.8	2.6	4.2	4.9	-0.2	10.7	9.4	8.7	8.8	9.7	10.6
0.2	5.1	4.0	2.9	2.8	3.9	4.9	0.2	10.5	9.7	8.6	8.7	9.6	10.8
0.6	53.9	52.5	50.2	50.3	52.4	54.2	0.6	73.6	72.1	72.4	71.6	72.8	73.9
1.0	85.7	85.6	84.3	84.5	84.8	86.4	1.0	99.1	98.9	97.3	97.2	98.8	99.7
	$k^* = 150, r = 0.2$							$k^* = 150, r = 0.2$					
$\phi_1 \backslash \phi_0$	-1.0	-0.6	-0.2	0.2	0.6	1.0	$\phi_1 \backslash \phi_0$	-1.0	-0.6	-0.2	0.2	0.6	1.0
-1.0	83.7	82.6	81.1	80.9	82.4	83.6	-1.0	97.1	96.7	96.4	96.0	96.5	97.1
-0.6	51.3	50.5	49.4	48.7	50.6	51.4	-0.6	72.0	70.2	69.7	69.4	69.8	72.1
-0.2	4.1	3.0	2.0	1.9	2.9	3.8	-0.2	9.3	8.7	7.4	7.8	8.5	9.2
0.2	4.1	3.5	2.1	2.2	3.0	3.8	0.2	9.5	8.6	7.2	7.4	8.8	9.3
0.6	51.8	50.3	48.2	48.4	50.2	51.6	0.6	72.0	69.8	69.8	69.6	70.2	71.9
1.0	83.9	82.6	80.9	80.8	82.3	83.8	1.0	97.1	96.5	95.6	96.2	96.4	97.0

Table 4. Powers (in percentage) against model (2.1) with $(\psi_0, \psi_1) = (0.1, 0.2)$ and $x_t = 0.9x_{t-1} + \nu_t$.

		$\alpha = 0.01$				$\alpha = 0.05$		
		$k^* = 150, r = \infty$				$k^* = 150, r = \infty$		
A structural change and no thresholds	$\phi_1 \backslash \phi_0$	0.6	1.0	2.0	$\phi_1 \backslash \phi_0$	0.6	1.0	2.0
	0.6	3.8	11.7	37.3	0.6	10.6	29.0	61.9
	1	2.4	13.4	38.2	1	9.4	30.5	62.4
		$k^* = 0, r = 0$				$k^* = 0, r = 0$		
A threshold and no structural changes	$\phi_1 \backslash \phi_0$	0.6	1.0	2.0	$\phi_1 \backslash \phi_0$	0.6	1.0	2.0
	0.6	1.2	4.6	12.9	0.6	5.6	12.8	30.1
	1.0	3.2	4.7	11.6	1.0	8.0	14.0	28.1

Table 5. Powers (in percentage) against model (2.1) with $(\psi_0, \psi_1) = (0.1, 0.2)$ and $x_t = x_{t-1} + \nu_t$.

		$\alpha = 0.01$				$\alpha = 0.05$		
		$k^* = 150, r = \infty$				$k^* = 150, r = \infty$		
A structural change and no thresholds	$\phi_1 \backslash \phi_0$	0.6	1.0	2.0	$\phi_1 \backslash \phi_0$	0.6	1.0	2.0
	0.6	5.0	22.1	35.7	0.6	25.6	47.2	79.9
	1.0	7.0	16.6	40.7	1.0	24.6	42.2	73.4
		$k^* = 0, r = 0$				$k^* = 0, r = 0$		
A threshold and no structural changes	$\phi_1 \backslash \phi_0$	0.6	1.0	2.0	$\phi_1 \backslash \phi_0$	0.6	1.0	2.0
	0.6	0.0	0.5	2.0	0.6	0.5	4.0	4.1
	1.0	0.0	0.6	1.5	1.0	2.5	3.5	5.0

5. A Real Example

In this section, GLR_n and GF_n are applied to test whether commonly used financial variables have threshold effects and change points with respect to excess stock market returns. We consider the datasets analyzed in Welch and Goyal (2008) and Kostakis, Magdalinos, and Stamatogiannis (2015), but the period of data is 1927–2019, which is updated by Amit Goyal's Web. Following the setting in Welch and Goyal (2008), S&P 500 value-weighted log excess returns are used as excess market returns. We consider the following six variables as predictors: book-to-market value ratio (**b/m**), T-bill rate (**tbl**), net equity expansion (**ntis**), inflation rate (**inf**), long-term return (**ltr**), stock variance (**svar**), which are quarterly. Each dataset $\{x_t\}$ contains 372 observations

Table 6. GLR_n and GF_n tests for six predictive regressors.

	p-value for unit root test*	GLR_n or GF_n		r^*	k^*	Regime		
		$\alpha = 0.01$	$\alpha = 0.05$			$t \le k^*$	$t > k^*$	
							$q_{t-1} > r^*$	$q_{t-1} \le r^*$
b/m	0.0281	1	1	0.7934	119	119	33	220
tbl	0.5293	1	1	0.0552	43	43	78	251
ntis	< 0.01	0	1	0.0283	74	74	57	241
infl	< 0.01	1	1	0.0147	74	74	67	231
ltr	< 0.01	0	1	0.0429	74	74	75	223
svar	< 0.01	0	1	0.0084	74	74	44	254

NOTE: Alternative hypothesis: stationary. "1" means reject the null hypothesis in (5).

and the candidates for the threshold are $\{x_{(t)}\}_{t=38}^{335}$, where $\{x_{(t)}\}$ is an ascending rearrangement of $\{x_t\}$.

To determine if we should use GLR_n or GF_n, we first perform unit root tests for these six predictors and the corresponding p-values are presented in Table 6. From Table 6, one can find that only the T-bill rate is a unit root process at the significance level 0.05. Thus, we use GF_n to test the null in (5) when x_t is the T-bill rate, and use GLR_n to test for the null in (5) when x_t is other five predictors. The results are reported in Table 6, from which, one can observe the following findings:

i. GF_n rejects the null in (5) at the levels 0.01 and 0.05 when x_t is T-bill rate data. When the prediction variable x_t is book-to-market value ratio or stock variance, GLR_n test rejects the null in (5) at both levels 0.01 and 0.05. When x_t is inflation rate or long-term return or net equity expansion, GLR_n rejects the null at the level 0.05 but cannot reject the null at level 0.01. These findings illustrate that there is indeed a new regime jointly determined by a threshold and a change point in each dataset, and our test can be used for testing the change-point and threshold effect in PR model simultaneously.
ii. There exist the same change-point estimates ($k^* = 74$) when the prediction variable x_t is inflation rate, long-term return, net equity expansion or stock variance. Looking back at history, $t = 75$ corresponds to the second quarter of 1945, the end of World War II, which is indeed an important point for financial markets.

6. Conclusion and Discussion

In this article, two tests, termed as GLR and GF, are proposed to test structural changes and thresholds in linear predictive regression models. The former is for the case that regressor is stationary and the latter is for the situation that regressor is nonstationary such as unit root process or nearly integrated process. The limiting distribution of the proposed test statistics are derived under the null hypotheses. It turns out that the limiting distributions are functionals of some Gaussian processes with complicated covariance structures and depend heavily on the characteristics of regress, in the sense that regressor is stationary or nonstationary. Due to the complexity of the limiting distributions of the proposed test statistics, a Bootstrap approach is suggested for computing the critical values of the proposed tests.

Finally, we note several possible extensions of the present study. First, based on the unified tests in Chen, Deo, and Yi (2013), Li, Li, and Peng (2017), Liu et al. (2019), and Yang et al. (2021), it may be of interest to have some unified testing procedures free of the characteristics of regressor (regardless of the persistence of regressor). Second, it might allow regressor to have structural changes or thresholds too. Last, there exists the model imbalance or model inconsistency issue in the linear predictive regression when stock returns are predicted with a pure unit root predictor. To this end, Cai and Gao (2017) proposed a nonlinear predictive model to compress the strong signal of the unit root process involved, and Ren, Tu, and Yi (2019) proposed a balanced predictive regression by augmenting it with an additional lag of the predictors. Therefore, the results can be generalized to these settings. We leave such extensions as possible future research topics.

Appendix: Mathematical Proofs

Before giving the proof of Theorem 1, we first state the following Lemma 1, which is Theorem 8 in Li, Ling, and Zhang (2016) and is about the weak convergence of a general marked empirical process.

Let $\mathcal{F}_t$ be the σ-field. Assume $\mathbf{Z}_t$ and $\xi_t, t = 0, \pm 1, \ldots$, are $\mathcal{F}_t$-measurable $p \times 1$ random vectors and univariate random variables, respectively. We consider the general marked empirical process

$$W_n(x, \tau) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor n\tau \rfloor} \mathbf{Z}_t I(\xi_{t-d} \le x), \quad (x, \tau) \in [-\infty, \infty] \times [0, 1],$$

where d is a positive integer.

Lemma 1. Let $\mathbf{K}_x = E(\mathbf{Z}_t \mathbf{Z}_t^\top I(\xi_{t-d} \le x))$. Assume (i) $\{(\mathbf{Z}_t, \xi_{t-d})\}$ is an α-mixing process with geometric rate; (ii) $E(\mathbf{Z}_t | \mathcal{F}_{t-1}) = 0$ and $0 < E(\|\mathbf{Z}_t\|^2 |\ln \|\mathbf{Z}_t\||^5) < \infty$; (iii) $\mathbf{K}_x$ and $\mathbf{K}_x - \mathbf{K}_y$ are positive definite for any $x, y \in \mathbb{R}$ with $x > y$. Then, $W_n(x, \tau) \Rightarrow G(x, \tau)$ in $\mathbb{D}([-\infty, \infty] \times [0, 1])$, where $\{G(x, \tau) : (x, \tau) \in [-\infty, \infty] \times [0, 1]\}$ is a Gaussian process with mean zero and covariance kernel $\mathrm{cov}(G(x, \tau_1), G(x, \tau_2)) = (\tau_1 \wedge \tau_2)\mathbf{K}_{x \wedge y}$; almost all paths of $G(x, \tau)$ are continuous in x and τ.

The following lemma gives the relation between the moments assumptions in Assumption 2 and Lemma 1.

Lemma 2. If $E\|X_t u_t\|^{2+\delta} < \infty$ for some $\delta > 0$, then $E[\|X_t u_t\|^2 |\ln \|X_t u_t\||^5] < \infty$

Proof of Lemma 2: In fact, if a random variable $X > 0$ with $EX^\delta < \infty$ for some δ, then for any $m \ge 1$, we have

$$\begin{aligned} &E[\ln(1 + X)]^m \\ &\quad = \epsilon^{-m} E[\ln(1 + X)^\epsilon]^m \le O(1) E[\ln(1 + X^\epsilon)]^m \le EX^{m\epsilon} < \infty, \end{aligned}$$

by taking ϵ smaller so that $\epsilon m < \min\{\delta, 1\}$. Thus,

$$E(\|X_t u_t\|^2 (\ln \|X_t u_t\|)^5 I\{\|X_t u_t\| \geq 1\}) = E(\|X_t u_t\|^2 [\ln(1+b_t)]^5)$$
$$\leq \{E\|X_t u_t\|^{2+\delta}\}^{\frac{2}{2+\delta}} \{E[\ln(1+b_t)]^{\frac{5(2+\delta)}{\delta}}\}^{\frac{\delta}{2+\delta}} < \infty, \quad \text{(A.1)}$$

where $b_t = \{\|X_t u_t\|^5 - 1\} I\{\|X_t u_t\| \geq 1\}$. Note that, when $x \in (0,1)$, $x^2 |\ln x|^5 \leq x^2 (\ln x)^6 \to 0$ as $x \to 0$. Thus, for any constant $c > 0$, there exists a constant $x_0 > 0$ such that $x^2 |\ln x|^5 \leq c$ as $x \in (0, x_0)$. Thus,

$$E(\|X_t u_t\|^2 |\ln \|X_t u_t\||^5 I\{\|X_t u_t\| < 1\})$$
$$= E(\|X_t u_t\|^2 |\ln \|X_t u_t\||^5 I\{\|X_t u_t\| \leq x_0\})$$
$$+ E(\|X_t u_t\|^2 |\ln \|X_t u_t\||^5 I\{x_0 < \|X_t u_t\| < 1\})$$
$$\leq c + E(|\ln \|X_t u_t\||^5 I\{x_0 < \|X_t u_t\| < 1\}) \leq c + |\ln x_0^{-1}|^5 < \infty. \quad \text{(A.2)}$$

By (A.1) and (A.2), the conclusion holds.

Proof of Theorem 1: Using a two-term Taylor expansion, we have

$$\mathrm{GLR}_n(k,r) = \frac{S_n(k,r)}{\hat{\sigma}_n^2} + \frac{S_n^2(k,r)}{2n\xi_n^2(k,r)}, \quad \text{(A.3)}$$

where $S_n(k,r) = n(\hat{\sigma}_n^2 - \hat{\sigma}_n^2(k,r))$ and $\xi_n^2(k,r)$ is between $\hat{\sigma}_n^2$ and $\hat{\sigma}_n^2(k,r)$. By (6) and (7), we can show that (see, e.g., Example 7.1 in Schott (2017))

$$S_n(k,r) = T_{kr}^\top \left\{ \frac{X_{kr}^\top X_{kr}}{n} - \frac{X_{kr}^\top X}{n} \left(\frac{X^\top X}{n} \right)^{-1} \frac{X^\top X_{kr}}{n} \right\}^{-1} T_{kr}, \quad \text{(A.4)}$$

where $T_{kr} = n^{-1/2}[X_{kr}^\top - X_{kr}^\top X(X^\top X)^{-1} X^\top]Y$. Note that $T_{kr} = n^{-1/2}[X_{kr}^\top - X_{kr}^\top X(X^\top X)^{-1} X^\top]u$ if H_0 holds. Rescale the time axis by setting $k = \lfloor n\pi \rfloor$ with $\pi \in [0,1]$, and denote $X_{\pi\lambda}^* = X_{\lfloor n\pi \rfloor, r}$ and $T_{\pi\lambda}^* = T_{\lfloor n\pi \rfloor, r}$.

Using arguments similar to (iv) of Lemma 2.1 in Chan (1990), we show that $M_{\pi\lambda}$ is invertible for every $(\pi, \lambda) \in [0, \pi_1] \times [\lambda_1, \lambda_2]$ as follows. It is easy to know that $\Sigma, \Sigma_{\pi\lambda}$ and $\Sigma - \Sigma_{\pi\lambda}$ are positive definite, then exit orthogonal matrix Q and diagonal matrix $D = \mathrm{diag}(d_1, d_2)$ such that $Q\Sigma Q^\top = \mathrm{diag}(d_1^*, d_2^*)$ and $Q\Sigma_{\pi\lambda} Q^\top = D$, and $0 < d_i < d_i^*, i = 1, 2$. Hence, $M_{\pi\lambda} = \Sigma_{\pi\lambda} - \Sigma_{\pi\lambda}\Sigma^{-1}\Sigma_{\pi\lambda}$ is also positive definite.

By the ergodic theorem, we know that

$$\frac{1}{n} X^\top X \to \Sigma, \quad \frac{1}{n} X_{\pi\lambda}^{*\top} X \to \Sigma_{\pi\lambda}, \quad \frac{1}{n} X_{\pi\lambda}^{*\top} X_{\pi\lambda}^* \to \Sigma_{\pi\lambda} \text{ a.s.}$$

as $n \to \infty$. Then using results in Kaczor and Nowak (2001, p.85), we know that

$$\sup_{0 \leq \pi \leq \pi_1} \sup_{\lambda_1 \leq \lambda \leq \lambda_2} \left| \frac{X_{\pi\lambda}^{*\top} X}{n} - \Sigma_{\pi\lambda} \right| = o_p(1), \quad \text{and}$$
$$\sup_{0 \leq \pi \leq \pi_1} \sup_{\lambda_1 \leq \lambda \leq \lambda_2} \left| \frac{X_{\pi\lambda}^{*\top} X_{\pi\lambda}^*}{n} - \Sigma_{\pi\lambda} \right| = o_p(1).$$

Thus, we have

$$\sup_{0 \leq \pi \leq \pi_1} \sup_{\lambda_1 \leq \lambda \leq \lambda_2} \left| \left\{ \frac{X_{\pi\lambda}^{*\top} X_{\pi\lambda}^*}{n} - \frac{X_{\pi\lambda}^{*\top} X}{n} \left(\frac{X^\top X}{n} \right)^{-1} \frac{X^\top X_{\pi\lambda}^*}{n} \right\}^{-1} - M_{\pi\lambda}^{-1} \right| = o_p(1), \quad \text{(A.5)}$$

and

$$\sup_{0 \leq \pi \leq \pi_1} \sup_{\lambda_1 \leq \lambda \leq \lambda_2} \left| T_{\pi\lambda}^* - (-\Sigma_{\pi\lambda}\Sigma^{-1}, I_2) \frac{1}{\sqrt{n}} (X, X_{\pi\lambda}^*)^\top u \right| =$$
$$\sup_{0 \leq \pi \leq \pi_1} \sup_{\lambda_1 \leq \lambda \leq \lambda_2} \left| T_{\pi\lambda}^* + \frac{1}{\sqrt{n}} \Sigma_{\pi\lambda}\Sigma^{-1} X^\top u - \frac{1}{\sqrt{n}} X_{\pi\lambda}^{*\top} u \right| = o_p(1),$$

where I_2 is an identity matrix with order 2.

Let $\mathbf{K}_x = E(u_t^2 X_t X_t^\top I(q_{t-1} \leq x))$, then it is obvious that $\mathbf{K}_x$ and $\mathbf{K}_x - \mathbf{K}_y$ are positive definite for any x, y with $x > y$. Now, under H_0, based on Assumptions 1–3 and Lemma 2, applying Lemma 1 with $\tau = 1, \mathbf{Z}_t = X_t u_t, \xi_{t-d} = q_{t-1}$, we know that $\{T_{\pi\lambda}^*\}$ converges weakly to $\{\sigma\xi_{\pi\lambda}\}$ in $\mathbb{D}([0,1] \times [-\infty, \infty])$, where $\mathbb{D}(A)$ is the space of real-valued functions on the set A, which are right continuous and have left-hand limits and it is equipped with the Skorohod topology.

We use the method in Zhu and Ling (2012) to finish the proof. From the convergence of $T_{\pi\lambda}^*$, we know that $\sup_{0 \leq \pi \leq \pi_1} \sup_{\lambda_1 \leq \lambda \leq \lambda_2} T_{\pi\lambda}^* = O_p(1)$. Combing it with (A.5), it follows that

$$\sup_{0 \leq k \leq k_1} \sup_{r_1 \leq r \leq r_2} S_n(k,r) = \sup_{0 \leq \pi \leq \pi_1} \sup_{\lambda_1 \leq \lambda \leq \lambda_2} T_{\pi\lambda}^{*\top} M_{\pi\lambda}^{-1} T_{\pi\lambda}^* + o_p(1). \quad \text{(A.6)}$$

Similar to Theorem 2.3 in Chan (1990), define functional $L : x(\cdot) \to \sup_{0 \leq \pi \leq \pi_1} \sup_{\lambda_1 \leq \lambda \leq \lambda_2} x(\pi, \lambda)^\top M_{\pi\lambda}^{-1} x(\pi, \lambda)$. Note that $M_{\pi\lambda}$ is a continuous matrix function over $[0, \pi_1] \times [\lambda_1, \lambda_2]$, and $\sup_{0 \leq \pi \leq \pi_1} \sup_{\lambda_1 \leq \lambda \leq \lambda_2} |x(\pi, \lambda)| < \infty$, thus L is a continuous functional, so $L(T_{\pi\lambda}^*)$ converge weakly to $L(\sigma\xi_{\pi\lambda})$. Then, by (A.6),

$$\sup_{0 \leq k \leq k_1} \sup_{r_1 \leq r \leq r_2} S_n(k,r) \xrightarrow{d} \sigma^2 \sup_{0 \leq \pi \leq \pi_1} \sup_{\lambda_1 \leq \lambda \leq \lambda_2} \xi_{\pi\lambda}^\top M_{\pi\lambda}^{-1} \xi_{\pi\lambda}, \quad \text{(A.7)}$$

which shows that $\sup_{0 \leq k \leq k_1} \sup_{r_1 \leq r \leq r_2} S_n(k,r) = O_p(1)$. Then we have $\sup_{0 \leq k \leq k_1} \sup_{r_1 \leq r \leq r_2} |\hat{\sigma}_n^2(k,r) - \hat{\sigma}_n^2| = o_p(1)$. Note that $\xi_n^2(k,r)$ is between $\hat{\sigma}_n^2$ and $\hat{\sigma}_n^2(k,r)$ and $\hat{\sigma}_n^2$ converges to σ^2 a.s., so we have $\sup_{0 \leq k \leq k_1} \sup_{r_1 \leq r \leq r_2} \frac{1}{\xi_n^2(k,r)} = O_p(1)$, then if follows that

$$\frac{1}{2n} \sup_{0 \leq k \leq k_1} \sup_{r_1 \leq r \leq r_2} \frac{S_n^2(k,r)}{\xi_n^2(k,r)} = o_p(1). \quad \text{(A.8)}$$

This theorem follows from (A.3), (A.7), and (A.8). This completes the proof. □

The following lemma is Lemma 1 in Gonzalo and Pitarakis (2012), which is used for proving Theorem 2 and stated here for convenience.

Lemma 3. Recall $I_t^* = I(q_t \leq r) = I(U_t \leq \lambda)$. Under Assumptions 4 and 5, when x_t is persistent given by (11), we have

$$\frac{1}{n} \sum I_t^* \xrightarrow{p} \lambda, \quad \frac{1}{n^{3/2}} \sum x_t \Rightarrow \int_0^1 K_c(s) ds, \quad \frac{1}{n^2} \sum x_t^2 \Rightarrow \int_0^1 K_c^2(s) ds,$$
$$\frac{1}{n^{3/2}} \sum x_t I_t^* \Rightarrow \lambda \int_0^1 K_c(s) ds, \quad \frac{1}{n^2} \sum x_t^2 I_t^* \Rightarrow \lambda \int_0^1 K_c^2(s) ds,$$
$$\frac{1}{n} \sum x_{t-1} u_t \Rightarrow \sigma \int_0^1 K_c(s) dW(s, 1),$$
$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor n\pi \rfloor} u_t I_t^* \Rightarrow \sigma W(\pi, \lambda), \quad \frac{1}{n} \sum x_{t-1} u_t I_{t-1}^* \Rightarrow \sigma \int_0^1 K_c(s) dW(s, \lambda),$$

where $K_c(s)$ and $W(s, \lambda)$ are defined in Section 2.2.

Proof of Theorem 2: As shown in (A.4), we have $S_n(k,r) = n(\hat{\sigma}_n^2 - \hat{\sigma}_n^2(k,r)) = T_{kr}^\top R_{kr}^{-1} T_{kr}$ with

$$T_{kr} = D_n X_{kr}^\top u - D_n X_{kr}^\top X D_n (D_n X^\top X D_n)^{-1} D_n X^\top u,$$
$$R_{kr} = D_n X_{kr}^\top X_{kr} D_n - D_n X_{kr}^\top X D_n (D_n X^\top X D_n)^{-1} D_n X^\top X_{kr} D_n,$$

where $D_n = \mathrm{diag}\{1/\sqrt{n}, 1/n\}$. Rescale the time axis by setting $k = \lfloor n\pi \rfloor$ with $\pi \in [0,1]$, and note that $\lambda = F(r)$. Using Lemma 3, we have

$$D_n X_{kr}^\top X_{kr} D_n = D_n X_{kr}^\top X D_n \Longrightarrow \lambda \begin{pmatrix} (1-\pi) & \int_\pi^1 K_c(s)ds \\ \int_\pi^1 K_c(s)ds & \int_\pi^1 K_c^2(s)ds \end{pmatrix}$$
$$\equiv \lambda \int_\pi^1 \overline{K}_c(s)\overline{K}_c^\top(s)ds \equiv \lambda \Xi_\pi,$$

where $\overline{K}_c(s) = (1, K_c(s))^\top$. Similarly, we have

$$D_n X^\top X D_n \Longrightarrow \int_0^1 \overline{K}_c(s)\overline{K}_c^\top(s)ds = \Xi_0.$$

It now follows from the continuous mapping theorem that $R_{kr} \Longrightarrow L_{\pi\lambda}$. Using Lemma 3, we have

$$D_n X_{kr}^\top u \Longrightarrow \sigma \begin{pmatrix} W(1-\pi,\lambda) \\ \int_\pi^1 K_c(s)dW(s,\lambda) \end{pmatrix} = \sigma \int_\pi^1 \overline{K}_c(s)dW(s,\lambda),$$

and

$$D_n X^\top u \Longrightarrow \sigma \begin{pmatrix} W(1,1) \\ \int_0^1 K_c(s)dW(s,1) \end{pmatrix} = \sigma \int_0^1 \overline{K}_c(s)dW(s,1).$$

Thus, we have $T_{kr} \Longrightarrow \sigma\zeta_{\pi\lambda}$. This establishes that $S_n(k,r) \Longrightarrow \sigma^2\zeta_{\pi\lambda}^\top L_{\pi\lambda}^{-1}\zeta_{\pi\lambda}$. It follows from that

$$\begin{aligned}\hat{\sigma}_n^2 &= \frac{1}{n}u^\top u - \frac{1}{n}u^\top X(X^\top X)^{-1}X^\top u \\ &= \frac{1}{n}u^\top u - \frac{1}{n}\{u^\top X D_n (D_n X^\top X D_n)^{-1} D_n X^\top u\}. \qquad (A.9)\end{aligned}$$

Using a similar argument as for the weak convergence of T_{kr}, we know that the second term in (A.9) converges to 0. Furthermore, since $u^\top u/n$ converges to σ^2 a.s., we know that $\hat{\sigma}_n^2$ converges to σ^2 a.s. Note that $\hat{\sigma}_n^2(k,r) = \hat{\sigma}_n^2 - S_n(k,r)/n$ and $S_n(k,r)$ weakly converges to a limit. We know that $\hat{\sigma}_n^2(k,r)$ converges to σ^2 a.s. The theorem follows by standard manipulations. □

Although the proof for Corollary 2 can be derived from Theorem 2, here we give a detailed one for completeness.
Proof of Corollary 2: First, we focus on $S_n(k,r) = n(\hat{\sigma}_n^2 - \hat{\sigma}_n^2(k,r))$. As shown in (A.4), we have $S_n(k,r) = T_{kr}^\top R_{kr}^{-1} T_{kr}$ with

$$T_{kr} = D_n X_{kr}^\top Y - D_n X_{kr}^\top X D_n (D_n X^\top X D_n)^{-1} D_n X^\top Y,$$
$$R_{kr} = D_n X_{kr}^\top X_{kr} D_n - D_n X_{kr}^\top X D_n (D_n X^\top X D_n)^{-1} D_n X^\top X_{kr} D_n,$$

where $D_n = \mathrm{diag}\{1/\sqrt{n}, 1/n\}$. If H_0 holds, we have

$$T_{kr} = D_n X_{kr}^\top u - D_n X_{kr}^\top X D_n (D_n X^\top X D_n)^{-1} D_n X^\top u.$$

Rescale the time axis by setting $k = \lfloor n\pi \rfloor$ with $\pi \in [0,1]$, and note that $\lambda = F(r)$. Following Assumption 4 and Theorem 3 in Caner and Hansen (2001), we have

$$\begin{aligned} D_n X_{kr}^\top X_{kr} D_n &= D_n X_{kr}^\top X D_n \\ &= \begin{pmatrix} \frac{1}{n}\sum_{t=k+1}^n I_{t-1}^* & \frac{1}{n^{3/2}}\sum_{t=k+1}^n I_{t-1}^* x_{t-1} \\ \frac{1}{n^{3/2}}\sum_{t=k+1}^n I_{t-1}^* x_{t-1} & \frac{1}{n^2}\sum_{t=k+1}^n I_{t-1}^* x_{t-1}^2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{n}\sum_{t=1}^n I_{t-1}^* & \frac{1}{n^{3/2}}\sum_{t=1}^n I_{t-1}^* x_{t-1} \\ \frac{1}{n^{3/2}}\sum_{t=1}^n I_{t-1}^* x_{t-1} & \frac{1}{n^2}\sum_{t=1}^n I_{t-1}^* x_{t-1}^2 \end{pmatrix} \\ &\quad - \begin{pmatrix} \frac{1}{n}\sum_{t=1}^k I_{t-1}^* & \frac{1}{n^{3/2}}\sum_{t=1}^k I_{t-1}^* x_{t-1} \\ \frac{1}{n^{3/2}}\sum_{t=1}^k I_{t-1}^* x_{t-1} & \frac{1}{n^2}\sum_{t=1}^k I_{t-1}^* x_{t-1}^2 \end{pmatrix} \\ &\Longrightarrow \lambda \begin{pmatrix} 1 & \int_0^1 B_v(s)ds \\ \int_0^1 B_v(s)ds & \int_0^1 B_v^2(s)ds \end{pmatrix} \\ &\quad - \lambda \begin{pmatrix} \pi & \int_0^\pi B_v(s)ds \\ \int_0^\pi B_v(s)ds & \int_0^\pi B_v^2(s)ds \end{pmatrix} \\ &= \lambda \begin{pmatrix} (1-\pi) & \int_\pi^1 B_v(s)ds \\ \int_\pi^1 B_v(s)ds & \int_\pi^1 B_v^2(s)ds \end{pmatrix} \equiv \lambda \int_\pi^1 \overline{B}_v(s)\overline{B}_v^\top(s)ds \equiv \lambda\Omega_\pi, \end{aligned}$$

where $\overline{B}_v(s) = (1, B_v(s))^\top$. Theorem 3 in Caner and Hansen (2001) directly implies that

$$\begin{aligned} D_n X^\top X D_n &= \begin{pmatrix} 1 & \frac{1}{n^{3/2}}\sum_{t=1}^n x_{t-1} \\ \frac{1}{n^{3/2}}\sum_{t=1}^n x_{t-1} & \frac{1}{n^2}\sum_{t=1}^n x_{t-1}^2 \end{pmatrix} \\ &\Longrightarrow \begin{pmatrix} 1 & \int_0^1 B_v(s)ds \\ \int_0^1 B_v(s)ds & \int_0^1 B_v^2(s)ds \end{pmatrix} \\ &= \int_0^1 \overline{B}_v(s)\overline{B}_v^\top(s)ds = \Omega_0. \end{aligned}$$

It now follows from the continuous mapping theorem that $R_{kr} \Longrightarrow N_{\pi\lambda}$. Using Theorems 1 and 2 in Caner and Hansen (2001), we have

$$\begin{aligned} D_n X_{kr}^\top u &= \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{t=k+1}^n I_{t-1}^* u_t \\ \frac{1}{n}\sum_{t=k+1}^n I_{t-1}^* x_{t-1} u_t \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^n I_{t-1}^* u_t \\ \frac{1}{n}\sum_{t=1}^n I_{t-1}^* x_{t-1} u_t \end{pmatrix} \\ &\quad - \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^k I_{t-1}^* u_t \\ \frac{1}{n}\sum_{t=1}^k I_{t-1}^* x_{t-1} u_t \end{pmatrix} \\ &\Longrightarrow \sigma \begin{pmatrix} W(1,\lambda) \\ \int_0^1 B_v(s)dW(s,\lambda) \end{pmatrix} 0 - \sigma \begin{pmatrix} W(\pi,\lambda) \\ \int_0^\pi B_v(s)dW(s,\lambda) \end{pmatrix} \\ &= \sigma \begin{pmatrix} W(1-\pi,\lambda) \\ \int_\pi^1 B_v(s)dW(s,\lambda) \end{pmatrix} = \sigma \int_\pi^1 \overline{B}_v(s)dW(s,\lambda), \end{aligned}$$

$$D_n X^\top u = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t \\ \frac{1}{n} \sum_{t=1}^n x_{t-1} u_t \end{pmatrix} \Longrightarrow \sigma \begin{pmatrix} W(1,1) \\ \int_0^1 B_v(s) dW(s,1) \end{pmatrix}$$
$$= \sigma \int_0^1 \overline{B}_v(s) dW(s,1).$$

Thus, we have $T_{kr} \Longrightarrow \sigma \eta_{\pi\lambda}$. This establishes that $S_n(k,r) \Longrightarrow \sigma^2 \eta_{\pi\lambda}^\top N_{\pi\lambda}^{-1} \eta_{\pi\lambda}$. It follows from that

$$\hat{\sigma}_n^2 = \frac{1}{n} u^\top u - \frac{1}{n} u^\top X (X^\top X)^{-1} X^\top u$$
$$= \frac{1}{n} u^\top u - \frac{1}{n} \{u^\top X D_n (D_n X^\top X D_n)^{-1} D_n X^\top u\}. \quad \text{(A.10)}$$

Using a similar argument as for the weak convergence of T_{kr}, we know that the second term in (A.10) converges to 0. Furthermore, since $u^\top u/n$ to σ^2 a.s., we know that $\hat{\sigma}_n^2$ converges to σ^2 a.s. Note that $\hat{\sigma}_n^2(k,r) = \hat{\sigma}_n^2 - S_n(k,r)/n$ and $S_n(k,r)$ weakly converges to a limit. We know that $\hat{\sigma}_n^2(k,r)$ converges to σ^2 a.s. The theorem follows by standard manipulations. □

Proof of Theorem 3: We use the method as the proof of Theorem 2 in Hansen (1996), also see Theorem 3 in Zhu, Yu, and Li (2014b). Rescale the time axis and denote $z_t^*(\pi,\lambda) = z_t(k,r), \theta_n^*(\pi,\lambda) = \theta_n(k,r)$ and $\hat{Z}_n^*(\pi,\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^n z_t^*(\pi,\lambda) \hat{u}_t \varepsilon_t$. First, let W denote the set of samples ω for which

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \sup_{\pi,\lambda} \|z_t^*(\pi,\lambda)\| u_t^2 < \infty \quad \text{(A.11)}$$

and

$$\lim_{n\to\infty} \sup_{\pi,\lambda,\pi^*,\lambda^*} \left\| \frac{1}{n} \sum_{t=1}^n z_t^*(\pi,\lambda) z_t^{*\top}(\pi^*,\lambda^*) u_t^2 - \sigma^2 E(z_t^*(\pi,\lambda) z_t^{*\top}(\pi^*,\lambda^*)) \right\| \to 0 \; a.s. \quad \text{(A.12)}$$

Since $\sup_{\pi,\lambda} \|z_t^*(\pi,\lambda)\| \le \sqrt{2}\|X_t\|$ and $E(\|X_t\| u_t^2) < \infty$ due to Assumption 2, by the ergodic theorem we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \sup_{\pi,\lambda} \|z_t^*(\pi,\lambda)\| u_t^2 \le \lim_{n\to\infty} \frac{\sqrt{2}}{n} \sum_{t=1}^n \|X_t\| u_t^2 < \infty \; a.s,$$

that is, (A.11) holds. By Assumptions 1–3 and similar arguments for Theorem 1, it is not hard to see that (A.12) holds. Thus, $P(W) = 1$. Take any $\omega \in W$. For the remainder of the proof, all operations are conditionally on ω, and hence all of the randomness appears in the iid $N(0,1)$ variables $\{\varepsilon_t\}$.

Second, define

$$Z_n^*(\pi,\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^n z_t^*(\pi,\lambda) u_t \varepsilon_t.$$

Then, using the same argument for Lemma 1, we have $Z_n^*(\pi,\lambda) \Longrightarrow \sigma G_{\pi\lambda}$ a.s. as $n \to \infty$, where $G_{\pi\lambda}$ is a Gaussian process with zero-mean function and covariance kernel $E(z_t^*(\pi,\lambda) z_t^{*\top}(\pi^*,\lambda^*))$. Note that

$$\sup_{\pi,\lambda} \|\hat{Z}_n^*(\pi,\lambda) - Z_n^*(\pi,\lambda)\|$$
$$\le \sup_{\pi,\lambda} \left\| \frac{1}{n} \sum_{t=1}^n z_t^*(\pi,\lambda) z_t^{*\top}(\pi,\lambda) \varepsilon_t \right\| \sup_{\pi,\lambda} \|\sqrt{n}(\theta_n^*(\pi,\lambda) - \theta^0)\|,$$

where θ^0 is the true value of θ. Using the same argument for $Z_n^*(\pi,\lambda)$, we have

$$\frac{1}{n} \sum_{t=1}^n z_t^*(\pi,\lambda) z_t^{*\top}(\pi,\lambda) \varepsilon_t \Longrightarrow 0 \text{ a.s. as } n \to \infty.$$

Next, we show that $\sup_{\pi,\lambda} \|\sqrt{n}(\theta_n^*(\pi,\lambda) - \theta^0)\| = O_p(1)$ using arguments similar to Lemma A.6 in Zhu, Yu, and Li (2014b). Rescale the time axis and denote $u_t^*(\theta,\pi,\lambda) = u_t(\theta,k,r)$, then $u_t^*(\theta,\pi,\lambda) = y_t - z_t^{*\top}(\pi,\lambda)\theta$. For any (π,λ), by Taylor's expansion we have

$$\sum_{t=1}^n [u_t^{*2}(\theta_n^*(\pi,\lambda),\pi,\lambda) - u_t^{*2}(\theta^0,\pi,\lambda)]$$
$$= -(\theta_n^*(\pi,\lambda) - \theta^0)^\top \left(\sum_{t=1}^n 2u_t^*(\theta^0,\pi,\lambda) z_t^*(\pi,\lambda) \right)$$
$$+ (\theta_n^*(\pi,\lambda) - \theta^0)^\top \left(\sum_{t=1}^n z_t^*(\pi,\lambda) z_t^{*\top}(\pi,\lambda) \right) (\theta_n^*(\pi,\lambda) - \theta^0)$$
$$+ o_p(|\theta_n^*(\pi,\lambda) - \theta^0|^2).$$

Define $G_n^*(\pi,\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^n u_t^*(\theta^0,\pi,\lambda) z_t^*(\pi,\lambda)$. Similar to $Z_n^*(\pi,\lambda)$, we have that $G_n^*(\pi,\lambda) = O_p(1)$. Let $\lambda_{\min} > 0$ be the minimum eigenvalue of $E(z_t^*(\pi,\lambda) z_t^{*\top}(\pi,\lambda))$. Then, for any $\eta > 0$, there exists a $M(\eta) > 0$ such that

$$P\left(\sup_{\pi,\lambda} \sqrt{n}|\theta_n^*(\pi,\lambda) - \theta^0| > M(\eta) \right)$$
$$\le P\Bigg(\sqrt{n}|\theta_n^*(\pi,\lambda) - \theta^0| > M(\eta), \sum_{t=1}^n [u_t^{*2}(\theta_n^*(\pi,\lambda),\pi,\lambda)$$
$$- u_t^{*2}(\theta^0,\pi,\lambda)] \le 0 \text{ for some } (\pi,\lambda) \Bigg)$$
$$\le P(\sqrt{n}|\theta_n^*(\pi,\lambda) - \theta^0| > M(\eta), -2\sqrt{n}|\theta_n^*(\pi,\lambda) - \theta^0||G_n^*(\pi,\lambda)|$$
$$+ n|\theta_n^*(\pi,\lambda) - \theta^0|^2[\lambda_{\min} + o_p(1)] + o_p(|\theta_n^*(\pi,\lambda) - \theta^0|^2)$$
$$\le 0 \text{ for some } (\pi,\lambda))$$
$$\le P(M(\eta) < \sqrt{n}|\theta_n^*(\pi,\lambda) - \theta^0| \le [\lambda_{\min} + o_p(1)]^{-1}[2|G_n^*(\pi,\lambda)|$$
$$+ o_p(|\theta_n^*(\pi,\lambda) - \theta^0|/\sqrt{n})] \text{ for some } (\pi,\lambda))$$
$$\le P(|G_n^*(\pi,\lambda)| > M(\eta)[\lambda_{\min} + o_p(1)]/2 + o_p(1) \text{ for some } (\pi,\lambda))$$
$$\le \eta,$$

so $\sup_{\pi,\lambda} \|\sqrt{n}(\theta_n^*(\pi,\lambda) - \theta^0)\| = O_p(1)$ holds. Then, it follows that $\hat{Z}_n^*(\pi,\lambda) - Z_n^*(\pi,\lambda) \Longrightarrow 0$ in probability as $n \to \infty$. Thus, we have $\hat{Z}_n^*(\pi,\lambda) \Longrightarrow \sigma G_{\pi\lambda}$ in probability as $n \to \infty$.

Now, we consider the functional

$$L : z(\cdot,\cdot) \in \mathbb{D}([0,\pi_1] \times [\lambda_1,\lambda_2])$$
$$\to \frac{1}{\sigma^2} \sup_{0\le\pi\le\pi_1} \sup_{\lambda_1\le\lambda\le\lambda_2} z(\pi,\lambda)^\top K_{\pi\lambda} z(\pi,\lambda),$$

where $K_{\pi\lambda} = (-\Sigma_{\pi\lambda}\Sigma^{-1}, I_2)^\top M_{\pi\lambda}^{-1}(-\Sigma_{\pi\lambda}\Sigma^{-1}, I_2)$. Clearly, $L(\cdot)$ is a continuous functional. By the continuous mapping theory, it follows that $L(\hat{Z}_n^*(\pi,\lambda)) \Longrightarrow L(\sigma G_{\pi\lambda})$ in probability as $n \to \infty$. Using the facts that $\hat{\sigma}_n^2$ converges to σ^2 a.s. and

$$\left(-\frac{X_{kr}^\top X}{n} \left(\frac{X^\top X}{n} \right)^{-1}, I_2 \right)^\top \left\{ \frac{X_{kr}^\top X_{kr}}{n} - \frac{X_{kr}^\top X}{n} \left(\frac{X^\top X}{n} \right)^{-1} \frac{X^\top X_{kr}}{n} \right\}^{-1}$$
$$\left(-\frac{X_{kr}^\top X}{n} \left(\frac{X^\top X}{n} \right)^{-1}, I_2 \right)$$

converges to $K_{\pi\lambda}$ uniformly in (π,λ), we have that

$$\sup_{0\le\pi\le\pi_1}\sup_{\lambda_1\le\lambda\le\lambda_2} G\hat{L}R_n^*(\pi,\lambda) = L(\hat{Z}_n^*(\pi,\lambda)) + o_p(1).$$

Based on above discussions, we have

$$G\hat{L}R_n|w_1^*,\dots,w_n^* \xrightarrow{d} \sup_{0\le\pi\le\pi_1}\sup_{\lambda_1\le\lambda\le\lambda_2} G_{\pi\lambda}^\top K_{\pi\lambda}^{-1} G_{\pi\lambda}$$

in probability as $n \to \infty$. Note that the covariance kernel of $(-\Sigma_{\pi\lambda}\Sigma^{-1}, I_2)G_{\pi\lambda}$ is

$$\begin{aligned}&(-\Sigma_{\pi\lambda}\Sigma^{-1}, I_2)E(z_t^*(\pi,\lambda)z_t^{*\top}(\pi^*,\lambda^*))(-\Sigma_{\pi^*\lambda^*}\Sigma^{-1}, I_2)^\top\\
&= (-\Sigma_{\pi\lambda}\Sigma^{-1}, I_2)\begin{pmatrix}\Sigma & \Sigma_{\pi^*\lambda^*}\\ \Sigma_{\pi\lambda} & \Sigma_{\max(\pi,\pi^*),\min(\lambda,\lambda^*)}\end{pmatrix}\begin{pmatrix}-\Sigma^{-1}\Sigma_{\pi^*\lambda^*}\\ I_2\end{pmatrix}\\
&= \Sigma_{\max(\pi,\pi^*),\min(\lambda,\lambda^*)} - \Sigma_{\pi\lambda}\Sigma^{-1}\Sigma_{\pi^*\lambda^*}.\end{aligned}$$

Thus, the conclusion holds. □

The following lemma is a special case of Theorem 3 in Georgiev et al. (2019), which is used for proving Theorem 4 and stated here for convenience.

Lemma 4. Let $\tilde{e}_{nt}(t = 1,\dots,n)$ be scalar measurable functions of $y_i, x_i, q_i (i = 1,\dots,n)$ and such that $\frac{1}{n}\sum_{t=1}^{[nr]}\tilde{e}_{nt}^2 \xrightarrow{P} \int_0^r m^2(s)ds$ for $r \in [0,1]$, where $m(\cdot)$ is a square-integrable real function on $[0,1]$. Introduce $\varepsilon_{tb} = \tilde{e}_{nt}\varepsilon_t (t = 1,\dots,n)$, and $B^*(r) = \int_0^r m(s)dB_1^*(s)$, where B_1^* is a standard Brownian motion independent of (B_u, B_v). Under Assumption 6, the following converge jointly as

$$\begin{aligned}&\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}u_t, \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}v_t, \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}v_s u_t\right)\\
&\xrightarrow{w} \left(B_u(r), B_v(r), \int_0^1 B_u(s)dB_v(s)\right)\Big|B_v,\end{aligned}$$

$r \in [0,1]$, in the sense of weak convergence of random measures on $\mathcal{D}^2 \times \mathbb{R}$, and

$$\begin{aligned}&\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}v_t, \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_{tb}, \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}v_s\varepsilon_{tb}\right)\\
&\xrightarrow{w} \left(B_v(r), B^*(r), \int_0^1 B_v(s)dB^*(s)\right)\Big|B_v\end{aligned}$$

for $r \in [0,1]$, in the sense of weak convergence of random measures on $\mathcal{D}^2 \times \mathbb{R}$, where $\mathcal{D}^k := D_k[0,1]$ is the space of right continuous with left limit functions from $[0,1]$ to $\mathbb{R}^k$, equipped with the Skorokhod topology, and $\mathcal{D} := \mathcal{D}^1$.

Proof of Theorem 4: First we focus on $\hat{S}_n(k,r) = T_{kr}^{\star\top}R_{kr}^{-1}T_{kr}^{\star}$. Rescale the time axis by setting $k = \lfloor n\pi\rfloor$ with $\pi \in [0,1]$, and note that $\lambda = F(r)$. Since $n^{-3/2}\sum_{t=1}^n x_{t-1} \Longrightarrow \int_0^1 B_v^\top(s)ds$ by Assumption 4 and the continuous mapping theorem, we have

$$D_nX_{kr}^\top X_{kr}D_n = D_nX_{kr}^\top XD_n \Longrightarrow \lambda\int_\pi^1 \overline{B}_v(s)\overline{B}_v^\top(s)ds \equiv \lambda\Omega_\pi,$$

where $\overline{B}_v(s) = (1, B_v(s))^\top$. And similarly,

$$D_nX^\top XD_n \Longrightarrow \int_0^1 \overline{B}_v(s)\overline{B}_v^\top(s)ds = \Omega_0.$$

It now follows from the continuous mapping theorem again that $R_{kr} \Longrightarrow N_{\pi\lambda}$.

Next, it follows from Lemma 1 with $\tilde{e}_{nt} = \hat{u}_t$ that

$$\begin{aligned}&\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}v_t, \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}I_{t-1}^*\hat{u}_t, \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1}I_{t-1}^*x_{t-1}\hat{u}_t\right)\\
&\xrightarrow{w} \left(B_v(r), \sigma W(r,1), \sigma\int_0^1 B_v(s)dW(s,1)\right)\Big|B_v,\end{aligned}$$

where $n^{-1}\sum_{t=1}^{[nr]}I_{t-1}^*\hat{u}_t^2 \Longrightarrow \sigma^2W^2(r,1)$ and $\mathrm{var}(\sigma W(r,1)) = \sigma^2\mathrm{var}(W(r,1)) = \sigma^2E(W^2(r,1))$ satisfying the framework of Lemma 4. Then,

$$\begin{aligned}D_nX_{kr}^\top\hat{u} &= \begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}I_{t-1}^*\hat{u}_t\\ \frac{1}{n}\sum_{t=k+1}^{n}I_{t-1}^*x_{t-1}\hat{u}_t\end{pmatrix} = \begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}I_{t-1}^*\hat{u}_t\\ \frac{1}{n}\sum_{t=1}^{n}I_{t-1}^*x_{t-1}\hat{u}_t\end{pmatrix}\\
&\quad - \begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{k}I_{t-1}^*\hat{u}_t\\ \frac{1}{n}\sum_{t=1}^{k}I_{t-1}^*x_{t-1}\hat{u}_t\end{pmatrix}\\
&\Longrightarrow \sigma\begin{pmatrix}W(1,\lambda)\\ \int_0^1 B_v(s)dW(s,\lambda)\end{pmatrix}\Big|B_v - \sigma\begin{pmatrix}W(\pi,\lambda)\\ \int_0^\pi B_v(s)dW(s,\lambda)\end{pmatrix}\Big|B_v\\
&= \sigma\begin{pmatrix}W(1-\pi,\lambda)\\ \int_\pi^1 B_v(s)dW(s,\lambda)\end{pmatrix}\Big|B_v = \sigma\int_\pi^1 \overline{B}_v(s)dW(s,\lambda)\Big|B_v,\end{aligned}$$

and

$$\begin{aligned}D_nX^\top\hat{u} &= \begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{u}_t\\ \frac{1}{n}\sum_{t=1}^{n}x_{t-1}\hat{u}_t\end{pmatrix}\\
&\Longrightarrow \sigma\begin{pmatrix}W(1,1)\\ \int_0^1 B_v(s)dW(s,1)\end{pmatrix}\Big|B_v = \sigma\int_0^1 \overline{B}_v(s)dW(s,1)\Big|B_v.\end{aligned}$$

Hence, $T_{kr}^\star|w_1^*,\dots,w_n^* \Longrightarrow \sigma\eta_{\pi\lambda}|B_v$. Thus, $\hat{S}_n(k,r)|w_1^*,\dots,w_n^* \Longrightarrow \sigma^2\eta_{\pi\lambda}^\top N_{\pi\lambda}^{-1}\eta_{\pi\lambda}|B_v$. Thus, (a) holds by standard manipulations. Finally, using Theorem 2 and a similar argument, we can show that (b) holds. □

Acknowledgments

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

Funding

Zhu's research is supported by National Natural Science Foundation of China (NSFC) Grants (11871027 and 11731015) and Natural Science Foundation of Jilin Province Grants (20210101143JC). Ling's research is supported by Hong Kong Research Grants Commission Grants (16303118, 16301620 and 16300621), RGC-Theme-based Research Scheme T31-604/18-N), and the Australian Research Council. Cai's research is partially supported by NSFC Grants (71631004 and 72033008).

References

Andrews, D. W. K. (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856. [228,230]

Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models With Multiple Structural Changes," *Econometrica*, 66, 47–78. [228]

Berkes, I., Horváth, L., Ling, S., and Schauer, J. (2011), "Testing for Structural Change of AR Model to threshold AR Model," *Journal of Time Series Analysis*, 32, 547–565. [229,230]

Breitung, J., and Demetrescu, M. (2015), "Instrumental Variable and Variable Addition Based Inference in Predictive Regressions," *Journal of Econometrics*, 187, 358–375. [228]

Cai, B., and Gao, J. (2017), "A Simple Nonlinear Predictive Model for Stock Returns," Working Paper, *https://ideas.repec.org/p/msh/ebswps/2017-18.html* [235]

Cai, Z., Fang, Y., Lin, M., and Su, J. (2019), "Inferences for a Partially Varying Coefficient Model With Endogenous Regressors," *Journal of Business & Economic Statistics*, 37, 158–170. [230]

Cai, Z., and Tiwari, R.C. (2000), "Application of a Local Linear Autoregressive Model to BOD Time Series," *Environmetrics*, 11, 341–350. [230]

Cai, Z., and Wang, Y. (2014), "Testing Predictive Regression Models With Nonstationary Regressors," *Journal of Econometrics*, 178, 4–14. [228]

Cai, Z., Wang, Y., and Wang, Y. (2015), "Testing Instability in a Predictive Regression Model With Nonstationary Regressors," *Econometric Theory*, 31, 953–980. [229]

Campbell, J. Y., and Yogo, M. (2006), "Efficient Tests of Stock Return Predictability," *Journal of Financial Economics*, 81, 27–60. [228]

Caner, M., and Hansen, B. E. (2001), "Threshold Autoregression With a Unit Root," *Econometrica*, 69, 1555–1596. [231,237]

Carrasco, M. (2002), "Misspecified Structural Change, Threshold, and Markov-Switching Models," *Journal of Econometrics*, 109, 239–273. [233]

Cavaliere, G., and Georgiev, I. (2020), "Inference Under Random Limit Bootstrap Measures," *Econometrica*, 88, 2547–2574. [232]

Chen, H. (2015), "Robust Estimation and Inference for Threshold Models With Integrated Regressors," *Econometric Theory*, 31, 778–810. [231]

Chan, K. S. (1990), "Testing for Threshold Autoregression," *Annals of Statistics*, 18, 1886–1894. [236]

Chen, W. W., Deo, R. S. and Yi, Y. (2013), "Uniform Inference in Predictive Regression Models," *Journal of Business & Economic Statistics*, 31, 525–533. [235]

Chow, G. C. (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28, 591–605. [228]

Galvão, A. B. C. (2006), "Structural Break Threshold VARs for Predicting US Recessions Using the Spread," *Journal of Applied Econometrics*, 21, 463–487. [229]

Georgiev, I., Harvey, D. I., Leybourne, S. J., and Taylor, A. M. R. (2018), "Testing for Parameter Instability in Predictive Regression Models," *Journal of Econometrics*, 204, 101–118. [228,230,232]

Georgiev, I., Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2019), "A Bootstrap Stationarity Test for Predictive Regression Invalidity," *Journal of Business & Economic Statistics*, 37, 528–541. [232,239]

Gonzalo, J., and Pitarakis, J. (2012), "Regime-Specific Predictability in Predictive Regressions," *Journal of Business & Economic Statistics*, 30, 229–241. [228,229,230,231,236]

Gonzalo, J., and Pitarakis, J. (2017), "Inferring the Predictability Induced by a Persistent Regressor in a Predictive Threshold Model," *Journal of Business & Economic Statistics*, 35, 202–217. [229,230]

Hansen, B. E. (1996), "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430. [231,238]

Hansen, B. E. (2000), "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, 97, 93–115. [232]

Horváth, L. (1993), "The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations," *Annals of Statistics*, 21, 671–680. [228]

Huang, H., and Chang, Y. (2005), "Investigating Okun's Law by the Structural Break With Threshold Approach: Evidence From Canada," *The Manchester School*, 73, 599–611. [229]

Jansson, M., and Moreira, M. J. (2006), "Optimal Inference in Regression Models With Nearly Integrated Regressors," *Econometrica*, 74, 681–714. [228]

Kaczor, W. J., and Nowak, M. T. (2001), *Problems in Mathematical Analysis II*, Providence: AMS. [236]

Kiliç, R. (2018), "Robust Inference for Predictability in Smooth Transition Predictive Regressions," *Econometric Reviews*, 37, 1067–1094. [229]

Kostakis, A., Magdalinos, T., and Stamatogiannis, M.P. (2015), "Robust Econometric Inference for Stock Return Predictability," *Review of Financial Studies*, 28, 1506–1553. [228,234]

Li, C., Li, D., and Peng, L. (2017), "Uniform Test for Predictive Regression With AR Errors," *Journal of Business & Economic Statistics*, 35, 29–39. [235]

Li, D., Ling, S., and Zhang, R. (2016), "On a Threshold Double Autoregressive Model," *Journal of Business & Economic Statistics*, 34, 68–80. [235]

Liao, X., Cai, Z., and Chen, H. (2018), "A Perspective on Recent Methods on Testing Predictability of Asset Returns," *Applied Mathematics-A Journal of Chinese Universities*, 33, 127–144. [228]

Ling, S. (2007), "Testing for Change Points in Time Series Models and Limiting Theorems for NED Sequences," *Annals of Statistics*, 35, 1213–1237. [228]

Liu, X., Yang, B., Cai, Z., and Peng, L. (2019), "A Unified Test for Predictability of Asset Returns Regardless of Properties of Predicting Variables," *Journal of Econometrics*, 208, 141–159. [235]

McMillan, D. G. (2001), "Nonlinear Predictability of Stock Market Returns: Evidence From Nonparametric and Threshold Models," *International Review of Economics and Finance*, 10, 353–368. [229]

Paye, B. S., and Timmermann, A. (2006), "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315. [229]

Pettenuzzo, D., and Timmermann, A. (2011), "Predictability of Stock Returns and Asset Allocation Under Structural Breaks," *Journal of Econometrics*, 164, 60–78. [229]

Phillips, P. C. B. (1988), "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56, 1021–1043. [231]

Pitarakis, J. (2017), "A Simple Approach for Diagnosing Instabilities in Predictive Regressions," *Oxford Bulletin of Economics and Statistics*, 79, 851–874. [228,230]

Quandt, R. E. (1960), "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324–330. [228]

Rapach, D. E., and Wohar, M. E. (2006), "Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns," *Journal of Financial Econometrics*, 4, 238–274. [229]

Ren, Y., Tu, Y., and Yi, Y. (2019), "Balanced Predictive Regressions," *Journal of Empirical Finance*, 54, 118–142. [235]

Schott, J. R. (2017), *Matrix Analysis for Statistics*, 3rd ed., Hoboken, NJ: Wiley. [236]

Smith, S. C., and Timmermann, A. (2021), "Break Risk," *Review of Financial Studies*, 34, 2045–2100. [229]

Welch, I., and Goyal, A. (2008), "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508. [228,234]

Wong, C. S., and Li, W. K. (1997), "Testing for Threshold Autoregression With Conditional Heteroscedasticity," *Biometrika*, 84, 407–418. [229]

Yang, B., Liu, X., Peng, L., and Cai, Z. (2021), "Unified Tests for a Dynamic Predictive Regression," *Journal of Business & Economic Statistics*, 39, 684–699. [228,235]

Yang, B., Long, W., Peng, L., and Cai, Z. (2020), "Testing the Predictability of U.S. Housing Price Index Returns Based on an IVX-AR Model," *Journal of the American Statistical Association*, 115, 1598–1619. [228]

Zhu, F., Cai, Z., and Peng, L. (2014a), "Predictive Regressions for Macroeconomic Data," *Annals of Applied Statistics*, 8, 577–594. [228]

Zhu, K., and Ling, S. (2012), "Likelihood Ratio Tests for the Structural Change of an AR(p) Model to a Threshold AR(p) Model," *Journal of Time Series Analysis*, 33, 223–232. [229,230,236]

Zhu, K., Yu, P. L. H., and Li, W. K. (2014b), "Testing for the Buffered Autoregressive Processes," *Statistica Sinica*, 24, 971–984. [231,232,238]