A New Test on Asset Return Predictability with Structural Breaks*

Zongwu Cai†
University of Kansas

Seong Yeon Chang‡
Soongsil University

February 4, 2022

Abstract

This paper considers predictive regressions in which a structural break is allowed on an unknown date. We establish novel testing procedures for asset return predictability using empirical likelihood methods based on weighted-score equations. The theoretical results are useful in practice because our unified framework does not require distinguishing whether the predictor variables are stationary or nonstationary. Simulations show that the empirical likelihood-based tests perform well in terms of size and power in finite samples. As an empirical analysis, we test asset returns predictability using various predictor variables.

JEL Classification: C12, C14, C32, G12

Keywords: Autoregressive process; Empirical likelihood; Structural break; Unit root; Weighted estimation

---

*We are grateful to Dukpa Kim, Yuichi Kitamura, Whitney Newey, Youngki Shin, Liangjun Su, Yanping Yi, and Fukang Zhu. Chang's research was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A8067828).

†Department of Economics, University of Kansas, Lawrence, KS 66045 USA (Email: *caiz@ku.edu*)

‡*Corresponding author*: Department of Economics, Soongsil University, Seoul 06978, Korea (Email: *sychang@ssu.ac.kr*)

1 Introduction

Asset return predictability has been studied for decades as a cornerstone research topic in economics and finance. It has been widely examined in many financial applications, such as mutual fund performance, conditional capital asset pricing, and optimal asset allocations. There are two major facets to dealing with asset returns predictability: first, checking whether the return series is autocorrelated, a random walk, or a martingale difference sequence (MDS) and second, using financial (state) variables as predictors to determine whether the financial (state) variables can predict asset returns. There is a vast amount of literature devoted to testing whether asset returns are autocorrelated, random walk, MDS, or other types of dependent structures; see Campbell et al. (1997) and the references therein. Recently, numerous empirical studies have documented the predictability of asset returns using various lagged financial or state variables, such as the log dividend-price ratio, log earnings-price ratio, log book-to-market ratio, dividend yield, term spread, default premium, interest rates, and other economic variables. Although much research has been conducted, empirical evidence remains inconclusive.

Predictive regression is a conventional method used to check whether some financial variables have explanatory power for stock return predictability. Classical predictive regression is a structural predictive linear model:

$$\begin{cases} y_t = \alpha + \beta\, x_{t-1} + u_t, \\ x_t = \theta + \phi\, x_{t-1} + v_t \end{cases} \tag{1}$$

for $1 \le t \le T$, where $|\phi| < 1$, $u_t = \rho\, v_t + \epsilon_t$, $(\epsilon_t, v_t) \sim N(\mathbf{0}, \mathrm{diag}(\sigma_\epsilon^2, \sigma_v^2))$ are independent and identically distributed (i.i.d.) series and $x_0 \sim N(\theta(1-\phi)^{-1}, \sigma_v^2(1-\phi^2)^{-1})$. In a predictive regression, the future values of some scalar time series y_t can be predicted from the lagged values of a financial variable x_{t-1}. Therefore, the null hypothesis is no predictability; that is, $\mathbb{H}_0 : \beta = 0$. Note that this normality assumption may not be required in if the sample size T is large enough.

Notably, predictive regressions contain econometric issues that have crucial effects on testing predictability. Campbell (2008), Phillips and Lee (2013), and Phillips (2015) gave an overview of econometric issues and remedies in predictive regressions. Here, we briefly review their arguments for completeness. First, the correlation between u_t and v_t plays an important role in many applications; see Table 4 in Campbell and Yogo (2006), which creates the so-called "embedded endogeneity". Stambaugh (1999) showed that the ordinary least squares (OLS) estimator for β in (1) is biased in finite samples due to the correlation between u_t and v_t under normality and with stationary regression ($|\phi| < 1$), denoted by I(0). More precisely, the bias of the OLS estimator $\hat{\beta}$ can be represented as

$$\mathbb{E}[\hat{\beta} - \beta] = \rho \, \mathbb{E}[\hat{\phi} - \phi],$$

where $\rho = \text{cov}(u_t, v_t)/\text{var}(v_t)$. The autoregressive bias function $\mathbb{E}[\hat{\phi} - \phi]$ depends only on ϕ and the sample size T. Thus, the sample autocorrelation is biased downward about $-(1 + 3\phi)/T$, and the predictive slope β is biased upward with $\rho < 0$. Stambaugh (1999) suggested the first-order bias-correction estimator while Amihud and Hurvich (2004) considered a linear projection of u_t on v_t as $u_t = \rho \, v_t + \epsilon_t$ and then, regressed y_t on $\hat{v}_t$ and x_{t-1} with intercept; that is,

$$y_t = \alpha + \beta \, x_{t-1} + \rho \, \hat{v}_t + \epsilon_t, \tag{2}$$

where $\hat{v}_t$ is obtained from the second equation in (1). The OLS estimator of β, denoted by $\hat{\beta}$, is a two-stage approach. However, Amihud and Hurvich (2004) assumed that x_t is stationary.

Second, the autoregressive parameter ϕ in x_t is assumed to be persistent, which is crucial for the statistical inference on β. Important contributions about nearly integrated or integrated regressors include Phillips (1987), Elliott and Stock (1994), Cavanagh et al. (1995), Lewellen (2004), Torous et al. (2005), Campbell and Yogo (2006), Jansson and Moreira (2006), Amihud et al. (2009), Chen and Deo (2009), Cai and Wang (2014), and the references therein. In the literature, a persistent regressor x_t is represented in a local-to-unity framework; that is, $\phi = 1 - c/T$, $c \geq 0$, denoted by NI(1) if $c > 0$ or $I(1)$ if $c = 0$. Indeed, Cai and Wang (2014) showed that $\hat{\beta}$ in (2) has the following asymptotic distribution $T(\hat{\beta} - \beta) \xrightarrow{d} \xi_c$ where ξ_c is a random variable involving the integration of a geometric Brownian motion and "$\xrightarrow{d}$" denotes convergence in distribution; see Cai and Wang (2014) for details. Therefore, the asymptotic results, in particular the limiting distribution, depend on c, which is not estimable consistently, although its estimate has a limiting distribution.

Recently, a series of studies have considered some uniform inferences on predictive regressions in the sense that the testing procedure for predictability is robust to general time-series characteristics on the regressor and errors. These include, but not limited to, the papers by Campbell and Yogo (2006), Phillips and Magdalinos (2007, 2009), Chen and Deo (2009), Elliott (2011), Phillips and Lee (2013), Zhu et al. (2014), Kostakis et al. (2015), Breitung and Demetrescu (2015), Lee (2016), Fang and Lee (2019), Liu et al. (2019), Yang et al. (2020), Hosseinkouchack and Demetrescu (2021), Demetrescu and Rodrigues (2020), Yang et al. (2021), Zhu et al. (2021). The reader is referred to the recent survey paper by Liao et al.

(2018) for more discussions. Actually, Campbell and Yogo (2006) proposed a new method called the Q-test based on the Bonferroni idea to construct a confidence interval for β for each ϕ. Chen and Deo (2009) found that the intercept parameter in predictive regression with persistent covariates makes inference difficult, and they proposed the restricted likelihood method, which is free of such nuisance intercept parameters. More importantly, the bias of the restricted maximum likelihood estimates is much less than that of the OLS estimates near the unit root without loss of efficiency. Phillips and Magdalinos (2009) and Kostakis et al. (2015) introduced a data-filtering procedure called IVX estimation, which restricts the degree of persistence of data-filtered IVX instruments within the class of near-stationary process defined in Phillips and Magdalinos (2007). A standard instrumental variable estimation with the constructed instruments is robust to the general time-series characteristics of regressors in the sense that the derived estimator converges in distribution to a mixed normal limit. Hence, the corresponding Wald statistic asymptotically follows the chi-square distribution under the null. Phillips and Lee (2013) considered the IVX estimation to long-horizon predictive regressions with persistent covariates while Lee (2016) and Fang and Lee (2019) extended the IVX filtering method to predictive quantile regression. Zhu et al. (2014) proposed an empirical likelihood (EL) approach together with a weighted least squares idea to construct a confidence interval for β, recently, extended by Liu et al. (2019) and Yang et al. (2021).

In this line of work, predictive regression models investigated in the afoermentioned literature are assumed to be stable; that is, no structural breaks are allowed. However, as Stock and Watson (1996) and Lettau and Van Nieuwerburgh (2008) found, economic and financial variables are subject to smooth or structural changes, which makes it reasonable to allow for the possibility of structural changes in predictive regression models. Subsequent research formally considered structural breaks in the predictive regressions. For example, Viceira (1997), Paye and Timmermann (2006), Rapach and Wohar (2006), and Zhu et al. (2021) tested for structural breaks and found strong evidence of instability in predictive regression models. Lettau and Van Nieuwerburgh (2008) focused on level shifts in the predictor variables and explained that the forecasting relationship may be unstable unless such shifts are included in the analysis. Recently, Cai et al. (2015) considered a model with coefficients changing smoothly over time, and then proposed a nonparametric testing procedure to determine whether the time-varying coefficients change over time. They found that the coefficients were indeed unstable. Recently, Gonzalo and Pitarakis (2012, 2017) and Zhu et al. (2021) developed tests for the null hypothesis of no predictability against threshold predictability in a predictive regression model with threshold effects. A practical question is how to specify the form of time-varying coefficients; that is, how the coefficients change over time. In this study, we assume that the coefficients are piecewise constant with structural changes. Since the work by Perron (1989), it is well known that structural changes in the data-generating process (DGP) should be considered appropriately to make statistical inferences reliable. Despite the large body of literature on estimating predictive regression models, studies pertaining to testing and estimating predictability allowing for structural changes are scarce.

The main contributions of this study are twofold. First, we consider predictive regressions in which the model parameters exhibit a structural break on an unknown date. When the true break date is unknown, we estimate it and propose testing procedures for predictability based on a consistent estimate of the break fraction, which contrasts sharply with conventional predictability tests. To test for a structural break or parameter instability, important contributions include Andrews (1993) and Andrews and Ploberger (1994). Bai (1994, 1997) showed that the break fraction can be estimated consistently by minimizing the sum of squared residuals (SSR) from the unrestricted model. They derived the limiting distribution of the estimate of the break date, which can be applied to constructing confidence intervals for the true break date. Bai and Perron (1998, 2003) considered statistical inference related to multiple structural changes under general conditions. Elliott and Müller (2006) considered the problem of testing for general types of parameter variations, including infrequent breaks, and established a partial-sums-type test based on the residuals obtained from the restricted model. The proposed tests were optimal as they nearly obtained the local Gaussian power envelop. The estimator of the break date is referenced in the literature, for instance, in Bai (1994, 1997), Bai and Perron (1998), Bai et al. (1998), and Kurozumi and Arai (2006).

Second, we propose the EL method based on weighted score equations, first introduced by Zhu et al. (2014), without allowing for a structural break. Prior studies have considered weighted estimating procedures, and a normal limiting distribution has been obtained (see, for example, Ling, 2005; Chan and Peng, 2005). Remarkably, Chan et al. (2012) extended a weighted estimation method to first-order autoregression, denoted as AR(1), to estimate the autoregressive parameter and found that the estimate maintains a normal limit regardless of whether the autoregressive process is I(0), I(1), NI(1), or even explosive ($c < 0$). They suggested using the EL method for the weighted score equation of the weighted least squares estimate to construct confidence intervals for all values of the AR parameter and showed that confidence intervals obtained by the EL method perform better in finite samples than those constructed using the weighted least squares method proposed by So and Shin (1999).

Recently, Li et al. (2014) established EL tests for causality of bivariate first-order autoregressive processes. Zhu et al. (2014), Liu et al. (2019) and Yang et al. (2021), most relevant to this article, considered predictive regressions and applied EL methods to test the null hypothesis of $\beta = 0$ and constructed confidence intervals for β. Chan et al. (2012) and Zhu et al. (2014) shed new light on predictive regressions as we can avoid estimating the autoregressive parameter ϕ and test for predictability based on the EL method, which performs well in finite samples.

In this study, we extend the analysis of Zhu et al. (2014) in a practical direction; that is, it is to test the instability of model parameters in predictive regressions and to show that the EL method based on some weighted score equations works well under such general circumstances. The simulation results indicate that the proposed EL tests have good finite-sample properties in terms of both size and power. To the best of our knowledge, this study is the first to incorporate an estimate of the break date and adopt a unified framework in predictive regressions.

The remainder of this paper is organized as follows. Section 2 introduces predictive regression models that exhibit a structural break on an unknown date. The EL-based methodologies are considered, and useful asymptotic results are presented. Section 3 provides simulation results to support the usefulness of the proposed EL method. In Section 4, techniques are applied to test the predictability of stock returns using a variety of predictive regressors. Finally, Section 5 provides brief concluding remarks. All technical derivations are presented in the appendix.

2 Econometric Approaches and Related Theories

We consider a linear predictive regression model that experiences a structural change on an unknown date. The standard model (1) is modified to allow for a structural change at date T_1^0 as follows: for $t = 1, \ldots, T$,

$$\begin{cases} y_t = (\alpha_1 + \beta_1 x_{t-1})\, \mathbf{1}_{t \le T_1^0} + (\alpha_2 + \beta_2 x_{t-1})\, \mathbf{1}_{t > T_1^0} + u_t, \\ x_t = \theta + \phi\, x_{t-1} + \sum_{j=0}^{\infty} \psi_j v_{t-j}, \end{cases} \tag{3}$$

where, in what follows, the linear process $\sum_{j=0}^{\infty} \psi_j v_{t-j}$ is assumed to be strictly stationary[1], and $\{u_t,\, v_t\}$ is a sequence of i.i.d. random vectors with means zero and finite variances. To

[1]If $\{\psi_j\}$ in model (3) satisfies some condition, say $\sum_{j=0}^{\infty} |\psi_j| < \infty$, it is straightforward to show that $\sum_{j=0}^{\infty} \psi_j v_{t-j}$ is strictly stationary (see Brockwell and Davis, 1991, p.89)

test the predictability of model (3), we first consider the joint null hypothesis $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$ and propose a unified EL test regardless of whether x_t is I(0) or I(1) or NI(1). Moreover, we consider testing predictability in the pre- and post-break subsamples; that is, $\mathbb{H}_0 : \beta_1 = 0$ and $\mathbb{H}_0 : \beta_2 = 0$, respectively.

2.1 Testing the joint null hypothesis $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$

Note that the null hypothesis of interest is no predictability; that is, $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$. Under the null hypothesis, the predictive regression model in (3) reduces to a change in mean model as follows:

$$y_t = \alpha_1 \mathbf{1}_{t \le T_1^0} + \alpha_2 \mathbf{1}_{t > T_1^0} + u_t. \tag{4}$$

Now, we state some assumptions as follows.

Assumption 1. *The magnitude of the level shift can be expressed as* $|\alpha_2 - \alpha_1| = \delta_T T^{-1/2}$ *where* $\delta_T = O(T^{\epsilon})$ *for some* $\epsilon \in (0, 1/2]$.

Assumption 2. $T_1^0 = [T\lambda_0]$ *where* $\lambda_0 \in (\pi, 1-\pi)$ *for some* $\pi \in (0, 1/2)$.

Under Assumption 1, the magnitude of the level shift either is independent of the sample size or shrinks to zero at a rate slower than $T^{-1/2}$; thus, the break fraction can be estimated consistently regardless of whether x_t is either stationary or (nearly) integrated under the null hypothesis (see, for example, Bai, 1994; Bai and Perron, 1998; Kurozumi and Arai, 2006). Assumption 2 is the standard for ensuring that the pre- and post-break subsamples are asymptotically large enough. We propose a new testing procedure for the null hypothesis $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$ as follows:

- Step 1: Estimate the break dates in the change in mean model (4) using the procedure recommended in Bai and Perron (1998, 2006).
- Step 2: Split the whole sample into disjoint subsamples in accordance with the estimated break dates.
- Step 3: Compute the EL-based test statistic in each subsample, and add them up to construct the final statistic.

The break date can be estimated using a global least squares criterion:

$$\hat{T}_1 = \underset{T_1 \in \Lambda}{\arg\min}\ S_T(T_1) \tag{5}$$

where $\Lambda = T\Lambda_\nu$, $\Lambda_\nu = (\nu, 1-\nu)$ for some small trimming ν, and $S_T(T_1)$ is the SSR.

$$S_T(T_1) = \sum_{t=1}^{T_1}\left(y_t - T_1^{-1}\sum_{t=1}^{T_1} y_t\right)^2 + \sum_{t=T_1+1}^{T}\left(y_t - (T-T_1)^{-1}\sum_{t=T_1}^{T} y_t\right)^2$$

for an admissible break date T_1. Let $\hat{\lambda}_1 = \hat{T}_1/T$ denote the break fraction estimate. The consistency of $\hat{\lambda}$ is well established in the literature.

Remark 1. *If no level shift is allowed in (3); that is, $\alpha_1 = \alpha_2 = \alpha$. Then, $y_t = \alpha + u_t$ $(t = 1, \ldots, T)$ under $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$. The EL method established by Zhu et al. (2014) can be applied to test the null hypothesis without modification. Hence, in this study, we focus on the change in mean model (4) under the null hypothesis.*

However, in both subsamples, the intercepts α_1 and α_2 are unknown, and the EL method may fail (see, for example, Chan et al. (2012) and Zhu et al. (2014) for a detailed explanation of this issue.) To apply the EL method, we considered the following estimation equations:

$$\sum_{t=1}^{T_1^0}(y_t - \alpha_1 - \beta_1 x_{t-1}) = 0, \sum_{t=1}^{T_1^0}(y_t - \alpha_1 - \beta_1 x_{t-1})\, w(x_{t-1}) = 0, \tag{6}$$

and

$$\sum_{t=T_1^0+1}^{T}(y_t - \alpha_2 - \beta_2 x_{t-1}) = 0, \sum_{t=T_1^0+1}^{T}(y_t - \alpha_2 - \beta_2 x_{t-1})\, w(x_{t-1}) = 0, \tag{7}$$

where the weight $w(x_{t-1}) \equiv x_{t-1}/(1 + x_{t-1}^2)^{1/2}$. Solving Equations (6) and (7) yields the weighted OLS estimates of α_j and β_j for $j = 1, 2$. When x_t is (nearly) integrated, $(T_1^0)^{-1}\sum_{t=1}^{T_1^0} u_t\, w(x_{t-1})$ does not converge in probability to a constant but converges in distribution to a random variable as $T \to \infty$ because of the intercept term (see, e.g., Chan and Wei, 1987; Chan et al., 2012). This suggests that the joint limit of $(T_1^0)^{-1/2}\sum_{t=1}^{T_1^0}(y_t - \alpha_1 - \beta_1 x_{t-1})$ and $(T_1^0)^{-1/2}\sum_{t=1}^{T_1^0}(y_t - \alpha_1 - \beta_1 x_{t-1})\, w(x_{t-1})$ cannot follow a bivariate normal distribution. Similarly, the joint limits of $(T - T_1^0)^{-1/2}\sum_{t=T_1^0+1}^{T}(y_t - \alpha_2 - \beta_2 x_{t-1})$ and $(T - T_1^0)^{-1/2}\sum_{t=T_1^0+1}^{T}(y_t - \alpha_2 - \beta_2 x_{t-1})\, w(x_{t-1})$ cannot be a bivariate normal distribution. Hence, when the predictor variable x_t is nonstationary, the EL method based on weighted score equations fails as Wilks's theorem does not hold.

Zhu et al. (2014) suggested an avenue to avoid this problem. Here, we briefly review this argument. We can eliminate α_j, $j = 1, 2$, using the first difference. However, this comes at a cost. When $\phi = 1$, that is, x_t is an I(1) process, the sequence $\{x_t - x_{t-1}\}$ is a stationary process. The inference of β_j, $j = 1, 2$, becomes less efficient with the first difference as the rate of convergence is $T^{1/2}$ rather than T with an I(1) process x_t. Furthermore, the noise components, $\{u_t - u_{t-1}\}$, are not independent. To accommodate these difficulties, Zhu et al. (2014) used the first difference with a large lag. Let $m = [n/2]$ where n is the number of observations in the sample. The observables $\{y_t, x_t\}$ take the first difference with m-horizon differences. Then, we have $\tilde{y}_t = y_t - y_{t+m}$, $\tilde{x}_t = x_t - x_{t+m}$, and $\tilde{u}_t = u_t - u_{t+m}$ for $t = 1, 2, \ldots, m$. The EL function for β_j, $j = 1, 2$ can be constructed using $\tilde{y}_t$ and $\tilde{x}_t$.

The true break date T_1^0 is generally unknown. We use consistent estimates to apply the EL method. Without losing generality, we assume that $\hat{T}_1 < T_1^0$. Let $\hat{m}_1 = [\hat{T}_1/2]$, and $\hat{m}_2 = [(T - \hat{T}_1)/2]$. The difference series $\{\tilde{y}_t\}$ is obtained as follows for each subsample: (i) In the pre-$\hat{T}_1$ subsample, for $t = 1, 2, \ldots, \hat{m}_1$, we define $\tilde{y}_t = y_t - y_{t+\hat{m}_1}$, $\tilde{x}_t = x_t - x_{t+\hat{m}_1}$, and $\tilde{u}_t = u_t - u_{t+\hat{m}_1}$.

$$\tilde{y}_t = \beta_1 \tilde{x}_{t-1} + \tilde{u}_t, \quad \tilde{x}_t = \phi \tilde{x}_{t-1} + \sum_{j=0}^{\infty} \psi_j \tilde{v}_{t-j}.$$

Taking the difference with a large lag helps ensure that $|\tilde{x}_t| \xrightarrow{p} \infty$ when $|x_t| \xrightarrow{p} \infty$ as $t \to \infty$ where "$\xrightarrow{p}$" denotes convergence in probability. (ii) In the post-$\hat{T}_1$ subsample, for $t = \hat{T}_1 + 1, \ldots, \hat{T}_1 + \hat{m}_2$, we define $\tilde{y}_t = y_t - y_{t+\hat{m}_2}$, $\tilde{x}_t = x_t - x_{t+\hat{m}_2}$, and $\tilde{u}_t = u_t - u_{t+\hat{m}_2}$. It is noteworthy that the post-$\hat{T}_1$ subsample should be divided into two parts if the estimate of the break date differs from the true break date, for instance, $\hat{T}_1 < T_1^0$. More precisely,

$$y_t = \begin{cases} \alpha_1 + \beta_1 x_{t-1} + u_t, & \text{for } t = \hat{T}_1 + 1, \ldots, T_1^0, \\ \alpha_2 + \beta_2 x_{t-1} + u_t, & \text{for } t = T_1^0 + 1, \ldots, T, \end{cases}$$

and

$$y_{t+\hat{m}_2} = \alpha_2 + \beta_2 x_{t-1} + u_t, \quad t = \hat{T}_1 + 1, \ldots, \hat{T}_1 + \hat{m}_2.$$

Hence, we have

$$\tilde{y}_t = \beta_2 \tilde{x}_{t-1} + \tilde{u}_t, \quad \tilde{x}_t = \phi \tilde{x}_{t-1} + \sum_{j=0}^{\infty} \psi_j \tilde{v}_{t-j},$$

where

$$\tilde{u}_t = \begin{cases} (u_t - u_{t+\hat{m}_2}) + (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2) x_{t-1}, & \text{for } t = \hat{T}_1 + 1, \ldots, T_1^0, \\ (u_t - u_{t+\hat{m}_2}) & \text{for } t = T_1^0 + 1, \ldots, \hat{T}_1 + \hat{m}_2. \end{cases}$$

Correspondingly, we let $\tilde{l}_1(\beta_1)$ denote the EL-based statistics calculated using data for $t = 1, \ldots, \hat{T}_1$ and $\tilde{l}_2(\beta_2)$ using data for $t = \hat{T}_1 + 1, \ldots, T$. More precisely, based on the preceding equations, the EL function is defined as

$$\tilde{L}_1(\beta_1) = \sup \left\{ \prod_{t=1}^{\hat{m}_1} (\hat{m}_1 p_t) : p_1 \geq 0, \ldots, p_{\hat{m}_1} \geq 0, \sum_{t=1}^{\hat{m}_1} p_t = 1, \sum_{t=1}^{\hat{m}_1} p_t \tilde{H}_t(\beta_1) = 0 \right\},$$

where $\tilde{H}_t(\beta_1) = (\tilde{y}_t - \beta_1 \tilde{x}_{t-1}) \tilde{x}_{t-1}/(1 + \tilde{x}_{t-1}^2)^{1/2}$. Note that the supremum is taken with respect to p_t. After applying the Lagrange multiplier technique, we obtain

$$\tilde{l}_1(\beta_1) = -2\log \tilde{L}_1(\beta_1) = 2\sum_{t=1}^{\hat{m}_1} \log\{1 + \tilde{\lambda}_1 \tilde{H}_t(\beta_1)\},$$

where $\tilde{\lambda}_1 = \tilde{\lambda}_1(\beta_1)$ satisfies

$$\sum_{t=1}^{\hat{m}_1} \frac{\tilde{H}_t(\beta_1)}{1 + \tilde{\lambda}_1 \tilde{H}_t(\beta_1)} = 0.$$

Similarly,

$$\tilde{L}_2(\beta_2) = \sup\left\{ \prod_{t=\hat{T}_1+1}^{\hat{T}_1+\hat{m}_2} (\hat{m}_2 p_t) : p_{\hat{T}_1+1} \geq 0, \ldots, p_{\hat{T}_1+\hat{m}_2} \geq 0, \sum_{t=\hat{T}_1+1}^{\hat{T}_1+\hat{m}_2} p_t = 1, \sum_{t=\hat{T}_1+1}^{\hat{T}_1+\hat{m}_2} p_t \tilde{H}_t(\beta_2) = 0 \right\},$$

where $\tilde{H}_t(\beta_2) = (\tilde{y}_t - \beta_2 \tilde{x}_{t-1}) \tilde{x}_{t-1}/(1 + \tilde{x}_{t-1}^2)^{1/2}$ and

$$\tilde{l}_2(\beta_2) = -2\log \tilde{L}_2(\beta_2) = 2\sum_{t=\hat{T}_1+1}^{\hat{T}_1+\hat{m}_2} \log\{1 + \tilde{\lambda}_2 \tilde{H}_t(\beta_2)\},$$

where $\tilde{\lambda}_2 = \tilde{\lambda}_2(\beta_2)$ satisfies

$$\sum_{t=\hat{T}_1+1}^{\hat{T}_1+\hat{m}_2} \frac{\tilde{H}_t(\beta_2)}{1 + \tilde{\lambda}_2 \tilde{H}_t(\beta_2)} = 0.$$

To validate Wilks's theorem for the aforementioned EL method, we assume the following regularity condition:

- **Condition A**: $\mathbf{E}[u_1] = 0$, $\mathbf{E}[v_1] = 0$, $\mathbf{E}[|u_1|^{2+\kappa} + |v_1|^{2+\kappa}] < \infty$ for some $\kappa > 0$ and $\{u_t, v_t\}$ is a sequence of i.i.d. random vectors.

Theorem 1. *Under Assumptions 1 and 2, suppose that model (4) holds with coefficients ψ_j satisfying that the linear process $\sum_{j=0}^{\infty} \psi_j v_{t-j}$ is a strictly stationary process, and either (i) $|\phi| < 1$ independent of T (stationary case), (ii) $\phi = 1 - c/T$ for some $c > 0$ (NI(1) case), or (iii) $\phi = 1$ (I(1) case). Then, under Condition A, we have $\tilde{l}_\beta(\beta_{1,0}, \beta_{2,0}) \equiv \tilde{l}_1(\beta_{1,0}) + \tilde{l}_2(\beta_{2,0}) \xrightarrow{d} \chi^2(2)$ as $T \to \infty$ where $(\beta_{1,0}, \beta_{2,0})$ denotes the true value of (β_1, β_2).*

Theorem 1 states that a unified EL test rejects the null hypothesis of no predictability in both regimes $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$ at significance level τ if $\tilde{l}_\beta(0,0) > \chi^2_{2.1-\tau}$ where $\chi^2_{2.1-\tau}$ denotes the $(1-\tau)$th quantile of a chi-squared distribution with two degrees of freedom. Confidence intervals/sets are frequently used in conjunction with point estimates to convey information about estimate uncertainty. Based on Theorem 1, the EL confidence set for $(\beta_{1,0}, \beta_{2,0})$ at level τ can be obtained as follows:

$$\mathrm{CI}_\tau = \{(\beta_1, \beta_2) : \tilde{l}_\beta(\beta_1, \beta_2) \leq \chi^2_{2,1-\tau}\}.$$

Therefore, the implementation for constructing confidence sets is straightforward without estimating any additional quantity. Indeed, the function "*emplik*" in the R package (see Zhou, 2015) can be employed to compute $\tilde{l}_1(\beta_1)$ and $\tilde{l}_2(\beta_2)$ as easily as in the simulation study below.

2.2 Testing predictability allowing for a structural break

If the joint null hypothesis $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$ is rejected, then it is likely that at least one of the regression coefficients (β_1, β_2) is nonzero, which supports stock return predictability in a certain regime.

To test the null hypothesis of no predictability, allowing for a structural change on an unknown date T_1^0, we can rewrite the predictive regression (3) as follows:

$$y_t = z'_{t-1}\gamma_1 \mathbf{1}_{t \leq T_1^0} + z'_{t-1}\gamma_2 \mathbf{1}_{t > T_1^0} + u_t, \quad t = 1, \ldots, T, \tag{8}$$

where $z_{t-1} = (1, x_{t-1})'$ and $\gamma_i = (\alpha_i, \beta_i)'$ for $i = 1, 2$ otherwise. The magnitude of the change is denoted by $\|\gamma_2 - \gamma_1\| = \|\delta\| \neq 0$ where the notation $\|\cdot\|$ denotes the Euclidean norm; that is, $\|k\| = (\sum_{i=1}^m k_i^2)^{1/2}$ for $k \in \mathbb{R}^m$. To establish the consistency of the estimated break fraction $\hat{\lambda}$, we require the following assumption.

Assumption 3. *The magnitude of the level shift can be expressed as* $\|\gamma_2 - \gamma_1\| = \|\delta\| = \delta_T T^{-1/2}$ *where* $\delta_T = O(T^\epsilon)$ *for some* $\epsilon \in (0, 1/2]$.

Although Assumption 3 is particularly well suited to an adequate approximation of the exact distribution when the predictor variable x_t is stationary, it remains adequate for non-stationary predictor variables (see Bai and Perron, 1998; Bai et al., 1998; Kurozumi and Arai, 2006). As explained in Elliott and Müller (2007), the break is sufficiently large to be detected with probability one with any reasonable test for breaks and to estimate its date in terms of the fraction of the sample consistently under Assumption 3. They considered a small break, $|\beta_2 - \beta_1| = dT^{-1/2}$ with a constant d in linear regression models and found that λ_0 is not consistently estimable when the regressors are stationary processes.

Remark 2. *If the predictor variable* x_t *is integrated, Assumption 3 can be relaxed to the case where* $|\beta_2 - \beta_1| = \delta_T T^{-1}$*; that is, the magnitude of the slope change shrinks to zero* *at a rate faster than* $T^{-1/2}$. *As addressed in Bai et al. (1998), Kurozumi and Arai (2006, 2007), the break fraction can be consistently estimated with a significantly smaller shift in the cointegrating coefficients. In this study, we consider testing procedures for stock return predictability regardless of whether the predictor variable is* $I(0)$ *or* $I(1)$, *thereby introducing Assumption 3.*

In matrix notation, we can rewrite predictive regression (8) as follows:

$$Y = \bar{X}_{T_1}\gamma + U.$$

For any matrix A, let A' denote its transpose. Let $Y' = [y_1, \ldots, y_T]$, $U' = [u_1, \ldots, u_T]$, $\gamma = (\gamma_1', \gamma_2')'$, and $\bar{X}_{T_1} = [X_1(\lambda), X_2(\lambda)]$ where $X_1(\lambda)_t = z_{t-1}'$ if $t \leq T_1$ and zero otherwise while $X_2(\lambda)_t = z_{t-1}'$ if $t > T_1$ and zero otherwise. The break date can be estimated using the global least-squares criterion:

$$\hat{T}_1 = \underset{T_1 \in \Lambda}{\arg\min}\, Y'(I - P_{T_1})Y,$$

where P_{T_1} is the matrix that projects on the range space of $\bar{X}_{T_1}$, i.e., $P_{T_1} = \bar{X}_{T_1}(\bar{X}_{T_1}'\bar{X}_{T_1})^{-1}\bar{X}_{T_1}'$. The OLS estimate of the regression coefficient γ is $\hat{\gamma} = (\bar{X}_{\hat{T}_1}'\bar{X}_{\hat{T}_1})^{-1}\bar{X}_{\hat{T}_1}'Y$ where $\bar{X}_{\hat{T}_1}$ is constructed with the estimate of the break date $\hat{T}_1$.

Proposition 1. *Suppose Assumptions 2 and 3 hold true. Subsequently,* $\hat{\lambda} \xrightarrow{p} \lambda_0$ *as* $T \to \infty$.

Proposition 1 asserts that the estimated break fraction remains consistent, even when the magnitude of the break decreases as the sample size increases, regardless of whether the predictor variable x_t is stationary or (nearly) integrated.

Remark 3. *Allowing more than one break in predictive regression models is not difficult; however, because our main interest is to construct EL-based tests for predictability under general assumptions on the predictor variable and errors, the single break model is effective for delineating the results.*

Remark 4. *There may be some concern about estimating a spurious break that does not exist in the DGP. Nunes et al. (1995) and Bai (1998) showed that the OLS estimate of the break date* $\hat{T}_1$ *is a spurious one when the error is an I(1) process. Kuan and Hsu (1998) and Hsu and Kuan (2008) considered a change in mean model for a fractionally integrated process and found that a spurious break can be estimated if memory parameter* $d^* \in (0, 1.5)$. *Recently, Chang and Perron (2016) considered a linear trend model with a change in slope with or without a concurrent level shift and showed that it is likely to estimate a spurious break when* *fractionally integrated errors have memory parameters in the interval* $(0, 1.5)$*, excluding the boundary case 0.5. In this study, the noise component* (u_t, v_t) *is assumed to be i.i.d., which excludes the risk of estimating a spurious break. Furthermore, Proposition 1 confirms that the estimate of the break fraction is consistent, regardless of whether the regressor* x_t *is either I(0) or I(1).*

Without a loss of generality, we assume that $\hat{T}_1 < T_1^0$. Let $\hat{m}_1 = [\hat{T}_1/2]$, and $\hat{m}_2 = [(T - \hat{T}_1)/2]$. The difference series $\{\ddot{y}_t\}$ is obtained as follows for each subsample: (i) In the pre-$\hat{T}_1$ subsample, for $t = 1, 2, \ldots, \hat{m}_1$, we define $\ddot{y}_t = y_t - y_{t+\hat{m}_1}$, $\ddot{x}_t = x_t - x_{t+\hat{m}_1}$, and $\ddot{u}_t = u_t - u_{t+\hat{m}_1}$.

$$\ddot{y}_t = \beta_1 \, \ddot{x}_{t-1} + \ddot{u}_t.$$

(ii) In the post-$\hat{T}_1$ subsample, for $t = \hat{T}_1 + 1, \ldots, \hat{T}_1 + \hat{m}_2$, we define $\ddot{y}_t = y_t - y_{t+\hat{m}_2}$, $\ddot{x}_t = x_t - x_{t+\hat{m}_2}$ and $\ddot{u}_t = u_t - u_{t+\hat{m}_2}$. Then,

$$y_t = \begin{cases} \alpha_1 + \beta_1 \, x_{t-1} + u_t, & \text{for } t = \hat{T}_1 + 1, \ldots, T_1^0, \\ \alpha_2 + \beta_2 \, x_{t-1} + u_t, & \text{for } t = T_1^0 + 1, \ldots, T, \end{cases}$$

and

$$y_{t+\hat{m}_2} = \alpha_2 + \beta_2 \, x_{t-1} + u_t, \quad t = \hat{T}_1 + 1, \ldots, \hat{T}_1 + \hat{m}_2,$$

from Proposition 1. Hence, we have

$$\ddot{y}_t = \beta_2 \, \ddot{x}_{t-1} + \ddot{u}_t,$$

where

$$\ddot{u}_t = \begin{cases} (u_t - u_{t+\hat{m}_2}) + (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2) x_{t-1}, & \text{for } t = \hat{T}_1 + 1, \ldots, T_1^0, \\ (u_t - u_{t+\hat{m}_2}) & \text{for } t = T_1^0 + 1, \ldots, \hat{T}_1 + \hat{m}_2. \end{cases}$$

Based on the preceding equations, the EL function is defined as follows:

$$\ddot{L}_1(\beta_1) = \sup \left\{ \prod_{t=1}^{\hat{m}_1} (\hat{m}_1 p_t) : p_1 \geq 0, \ldots, p_{\hat{m}_1} \geq 0, \sum_{t=1}^{\hat{m}_1} p_t = 1, \sum_{t=1}^{\hat{m}_1} p_t \ddot{H}_t(\beta_1) = 0 \right\},$$

where $\ddot{H}_t(\beta_1) = (\ddot{y}_t - \beta_1 \, \ddot{x}_{t-1}) \, \ddot{x}_{t-1} / (1 + \ddot{x}_{t-1}^2)^{1/2}$. After applying the Lagrange multiplier technique, we obtain:

$$\ddot{l}_1(\beta_1) = -2 \log \ddot{L}_1(\beta_1) = 2 \sum_{t=1}^{\hat{m}_1} \log\{1 + \ddot{\lambda}_1 \ddot{H}_t(\beta_1)\},$$

where $\ddot{\lambda}_1 = \ddot{\lambda}_1(\beta_1)$ satisfies

$$\sum_{t=1}^{\hat{m}_1} \frac{\ddot{H}_t(\beta_1)}{1 + \ddot{\lambda}_1 \ddot{H}_t(\beta_1)} = 0.$$

Similarly,

$$\ddot{L}_2(\beta_2) = \sup\left\{ \prod_{t=\hat{T}_1+1}^{\hat{T}_1+\hat{m}_2} (\hat{m}_2 p_t) : p_{\hat{T}_1+1} \geq 0, \ldots, p_{\hat{T}_1+\hat{m}_2} \geq 0, \sum_{t=\hat{T}_1+1}^{\hat{T}_1+\hat{m}_2} p_t = 1, \sum_{t=\hat{T}_1+1}^{\hat{T}_1+\hat{m}_2} p_t \ddot{H}_t(\beta_2) = 0 \right\},$$

where $\ddot{H}_t(\beta_2) = (\ddot{y}_t - \beta_2\, \ddot{x}_{t-1})\, \ddot{x}_{t-1}/(1 + \ddot{x}_{t-1}^2)^{1/2}$, and

$$\ddot{l}_2(\beta_2) = -2 \log \ddot{L}_2(\beta_2) = 2 \sum_{t=\hat{T}_1+1}^{\hat{T}_1+\hat{m}_2} \log\{1 + \ddot{\lambda}_2 \ddot{H}_t(\beta_2)\},$$

where $\ddot{\lambda}_2 = \ddot{\lambda}_2(\beta_2)$ satisfies

$$\sum_{t=\hat{T}_1+1}^{\hat{T}_1+\hat{m}_2} \frac{\ddot{H}_t(\beta_2)}{1 + \ddot{\lambda}_2 \ddot{H}_t(\beta_2)} = 0.$$

The following theorem shows that Wilks's theorem holds for the proposed EL method. Proposition 1 is crucial for deriving the asymptotic results in Theorem 2.

Theorem 2. *Under Assumptions 2 and 3, suppose that model (8) holds with coefficients ψ_j, satisfying that the linear process $\sum_{j=0}^{\infty} \psi_j v_{t-j}$ is a strictly stationary process, and either (i) $|\phi| < 1$ independent of T (stationary case), (ii) $\phi = 1 - c/T$ for some $c > 0$ (NI(1) case), or (iii) $\phi = 1$ (I(1) case). Then, under Condition A, $\ddot{l}_1(\beta_{1,0})$ and $\ddot{l}_2(\beta_{2,0})$ converge in distribution to a chi-square limit with one degree of freedom as $T \to \infty$.*

According to Theorem 2, unified EL tests can be obtained for testing $\mathbb{H}_0 : \beta_1 = 0$ and $\mathbb{H}_0 : \beta_2 = 0$ for model (8) based on $\ddot{l}_1(0)$ and $\ddot{l}_2(0)$, respectively. Again, a unified interval set can be obtained using Theorem 2.

3 Monte Carlo Simulation

In this section, we assess the finite sample properties of this procedure. We initially focus on the size properties of the test statistics. The DGP is given by (3), with $\beta_1 = \beta_2 = 0$; that is,

$$y_t = \alpha_1\, \mathbf{1}_{t \leq T_1^0} + \alpha_2\, \mathbf{1}_{t > T_1^0} + u_t,\ (t = 1, \ldots, T),$$

as defined in (4). The predictor variable, x_t, is specified by

$$x_t = \phi\, x_{t-1} + \sum_{j=0}^{\infty} \psi_j v_{t-j},$$

where the noise components (u_t, v_t) are contemporaneously correlated as follows:

$$u_t = \rho\, v_t + \sqrt{1 - \rho^2}\, \epsilon_t, \quad \rho = -0.95. \tag{9}$$

We choose either (i) $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$ or (ii) $\psi_j = 0.5^j$ for $j \geq 0$ in $\sum_{j=0}^{\infty} \psi_j v_{t-j}$ and consider the standard normal distribution for the distributions of (v_t, ϵ_t) in (9). We set the various parameters at the following values: $\alpha_1 = 0$, $\alpha_2 = aT^{-1/2}$ with $a \in \{4, 8, 12, 16\}$, $\lambda_0 \in \{0.5, 0.7\}$, and $\phi \in \{0.9, 1-2T^{-1}, 1\}$. In this configuration, the size is the rejection probability of the joint null hypothesis $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$. Three sample sizes, $T = 200, 500, 1000$, are considered to cover various-sized structural changes (see, for example, Elliott and Müller, 2007; Chang and Perron, 2018). The number of simulation replications is 5000 for each parameter combination. We use the R package "*emplik*" in Zhou (2015) to compute the EL function.

Tables 1-2 present the size properties of EL methods with a nominal size of 5%. The results regarding to the empirical size in the case of (i) $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$ are in Table 1. We observe that, for sample sizes $T \geq 500$, the EL tests have excellent size control across all values of ϕ. This is particularly true for medium and large level shifts. For $T = 200$, EL methods appear to be oversized for small and medium level shifts. When $\phi = 1$, mild size distortion remains even for a large level shift ($a = 16$). Table 2 presents the results regarding the empirical size in the case of (ii) $\psi_j = 0.5^j$ for $j \geq 0$ in $\sum_{j=0}^{\infty} \psi_j v_{t-j}$. We find that the EL statistic shows size close to the nominal rate 5% apart from some mild oversizing for $T = 200$ and $\lambda_0 = 0.7$, particularly for $\phi = 1$.

Next, we examine the power performance of the test against the deviation from the null hypothesis $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$. The DGP is given by (3):

$$y_t = (\alpha_1 + \beta_1 x_{t-1})\, \mathbf{1}_{t \leq T_1^0} + (\alpha_2 + \beta_2 x_{t-1})\, \mathbf{1}_{t > T_1^0} + u_t,\ (t = 1, \ldots, T),$$

Regarding the intercept terms, we set $\alpha_1 = 0$ and $\alpha_2 = 16T^{-1/2}$ (a single large level shift). We consider two cases for the slope coefficients: (b1) $\beta_1 = 0$, $\beta_2 = bT^{-1/2}$, and (b2) $\beta_1 = \beta_2 = bT^{-1/2}$ where $b \in \{1, 2, 4, 8\}$. Note that the DGP is reduced to the conventional predictive regression without a structural break; that is, $y_t = \alpha + \beta x_{t-1} + u_t$, $(t = 1, \ldots, T)$ if $\alpha_1 = \alpha_2$ and (b2) holds (see Zhu et al. (2014) for the finite sample performance of the EL-based tests). Tables 34 present the power properties of the EL method. The results for the empirical power in the case of (i) $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$ are listed in Table 3. In Panel (i), we consider the cases of (b1) $\beta_1 = 0$ and $\beta_2 = bT^{-1/2}$. It is clear that the powers of the EL tests increase with the sample size T and/or magnitude of the slope change b. In Panel (ii), we consider the case of (b2) $\beta_1 = \beta_2 = bT^{-1/2}$ where the predictor variable shows the same predictability in the pre- and post-break subsamples. The proposed EL tests have overall satisfactory power, and their powers increase as the alternative hypothesis is far from the null hypothesis. The test for the case of the NI(1) or I(1) is more powerful than that for the stationary case. Table

Table 1: Finite sample sizes for the test based on Theorem 1 and model (3) with $\theta = 0$ and normally distributed errors for testing $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$. We take $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$ in $\sum_{j=0}^{\infty} \psi_j v_{t-j}$.

	$T = 200$		$T = 500$		$T = 1000$	
(a, ϕ)	$\lambda_0 = 0.5$	$\lambda_0 = 0.7$	$\lambda_0 = 0.5$	$\lambda_0 = 0.7$	$\lambda_0 = 0.5$	$\lambda_0 = 0.7$
(4, 0.9)	0.069	0.074	0.061	0.061	0.056	0.050
(8, 0.9)	0.070	0.073	0.059	0.056	0.052	0.060
(12, 0.9)	0.059	0.056	0.050	0.054	0.054	0.053
(16, 0.9)	0.057	0.062	0.054	0.062	0.051	0.057
$(4,\ 1 - 2T^{-1})$	0.070	0.091	0.065	0.080	0.058	0.068
$(8,\ 1 - 2T^{-1})$	0.064	0.076	0.056	0.054	0.050	0.054
$(12,\ 1 - 2T^{-1})$	0.062	0.058	0.049	0.049	0.046	0.049
$(16,\ 1 - 2T^{-1})$	0.053	0.061	0.050	0.049	0.042	0.046
(4, 1)	0.085	0.093	0.075	0.076	0.067	0.076
(8, 1)	0.074	0.085	0.067	0.068	0.061	0.068
(12, 1)	0.070	0.077	0.065	0.072	0.058	0.064
(16, 1)	0.075	0.072	0.064	0.065	0.052	0.064

Note: The DGP is $y_t = \alpha_2 \mathbf{1}_{t > [\lambda_0 T]} + u_t$, where $x_t = \phi\, x_{t-1} + \sum_{j=0}^{\infty} \psi_j v_t$ with $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$, $\alpha_2 = a\, T^{-1/2}$ with $a \in \{4, 8, 12, 16\}$, and $\lambda_0 \in \{0.5, 0.7\}$. For the noise components, $v_t \sim i.i.d.\, N(0,1)$ and $u_t = \rho\, v_t + \sqrt{1-\rho^2}\, \eta_t$ with $\eta_t \sim i.i.d.\, N(0,1)$. Rejection frequencies are reported for testing $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$ with level 5% for the proposed empirical likelihood test $\tilde{l}_\beta(\beta_1, \beta_2)$.

Table 2: Finite sample sizes for the test based on Theorem 1 and model (3) with $\theta = 0$ and normally distributed errors for testing $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$. We take $\psi_j = 0.5^j$ for $j \geq 0$ in $\sum_{j=0}^{\infty} \psi_j v_{t-j}$.

	$T = 200$		$T = 500$		$T = 1000$	
(a, ϕ)	$\lambda_0 = 0.5$	$\lambda_0 = 0.7$	$\lambda_0 = 0.5$	$\lambda_0 = 0.7$	$\lambda_0 = 0.5$	$\lambda_0 = 0.7$
(4, 0.9)	0.069	0.072	0.060	0.066	0.054	0.056
(8, 0.9)	0.054	0.070	0.057	0.058	0.050	0.055
(12, 0.9)	0.058	0.060	0.051	0.062	0.056	0.050
(16, 0.9)	0.051	0.062	0.055	0.058	0.059	0.053
$(4,\ 1 - 2T^{-1})$	0.076	0.086	0.063	0.070	0.055	0.063
$(8,\ 1 - 2T^{-1})$	0.062	0.066	0.057	0.058	0.046	0.057
$(12,\ 1 - 2T^{-1})$	0.057	0.065	0.051	0.053	0.044	0.047
$(16,\ 1 - 2T^{-1})$	0.051	0.061	0.047	0.044	0.042	0.044
(4, 1)	0.082	0.092	0.080	0.072	0.071	0.078
(8, 1)	0.074	0.087	0.068	0.071	0.061	0.069
(12, 1)	0.074	0.077	0.064	0.063	0.062	0.058
(16, 1)	0.066	0.076	0.059	0.061	0.054	0.053

Note: The DGP is $y_t = \alpha_2 \mathbf{1}_{t > [\lambda^0 T]} + u_t$, where $x_t = \phi\, x_{t-1} + \sum_{j=0}^{\infty} \psi_j v_{t-j}$ with $\psi_j = 0.5^j$ for $j \geq 0$. The notes of Table 1 apply.

4 reports the results regarding the empirical power in the case of (ii) $\psi_j = 0.5^j$ for $j \geq 0$ and shows patterns similar to those obtained in Table 3, indicating that our method is robust against the dependent or independent errors in modeling the predicting variable.

Table 3: Finite sample powers for the test based on Theorem 1 and model (3) with $\theta = 0$ and normally distributed errors for testing $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$. We take $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$ in $\sum_{j=0}^{\infty} \psi_j v_{t-j}$.

		$T = 200$		$T = 500$		$T = 1000$	
		$\lambda_0 = 0.5$	$\lambda_0 = 0.7$	$\lambda_0 = 0.5$	$\lambda_0 = 0.7$	$\lambda_0 = 0.5$	$\lambda_0 = 0.7$
(i) (b1) holds							
$\phi = 0.9$	$b = 1$	0.187	0.192	0.164	0.145	0.153	0.130
	$b = 2$	0.478	0.361	0.470	0.325	0.478	0.304
	$b = 4$	0.991	0.872	0.999	0.931	1.000	0.923
	$b = 8$	1.000	0.999	1.000	1.000	1.000	1.000
$\phi = 1 - 2T^{-1}$	$b = 1$	0.563	0.414	0.868	0.599	0.977	0.807
	$b = 2$	0.952	0.737	0.999	0.949	1.000	0.996
	$b = 4$	0.999	0.986	1.000	0.999	1.000	1.000
	$b = 8$	1.000	0.999	1.000	1.000	1.000	1.000
$\phi = 1$	$b = 1$	0.681	0.505	0.916	0.676	0.986	0.847
	$b = 2$	0.975	0.795	0.999	0.962	1.000	0.997
	$b = 4$	1.000	0.990	1.000	1.000	1.000	1.000
	$b = 8$	1.000	1.000	1.000	1.000	1.000	1.000
(ii) (b2) holds							
$\phi = 0.9$	$b = 1$	0.262	0.305	0.261	0.266	0.249	0.266
	$b = 2$	0.813	0.830	0.802	0.815	0.803	0.803
	$b = 4$	1.000	1.000	1.000	1.000	1.000	1.000
	$b = 8$	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 1 - 2T^{-1}$	$b = 1$	0.857	0.882	0.994	0.997	1.000	1.000
	$b = 2$	0.999	0.999	1.000	1.000	1.000	1.000
	$b = 4$	1.000	1.000	1.000	1.000	1.000	1.000
	$b = 8$	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 1$	$b = 1$	0.944	0.951	0.998	0.998	1.000	1.000
	$b = 2$	1.000	1.000	1.000	1.000	1.000	1.000
	$b = 4$	1.000	1.000	1.000	1.000	1.000	1.000
	$b = 8$	1.000	1.000	1.000	1.000	1.000	1.000

Note: The DGP is $y_t = (\alpha_1 + \beta_1 x_{t-1})\,\mathbf{1}_{t \leq T_1^0} + (\alpha_2 + \beta_2 x_{t-1})\,\mathbf{1}_{t > T_1^0} + u_t$, where $x_t = \phi\, x_{t-1} + \sum_{j=0}^{\infty} \psi_j v_t$ with $\psi_0 = 1$ and $\psi_j = 0$ for $j \geq 1$. Regarding the intercept terms, we set $\alpha_1 = 0$ and $\alpha_2 = 16T^{-1/2}$ (a single large level shift). As for slope coefficients, we consider two cases: (b1) $\beta_1 = 0$, $\beta_2 = bT^{-1/2}$, and (b2) $\beta_1 = \beta_2 = bT^{-1/2}$ where $b \in \{1, 2, 4, 8\}$. See notes to Table 1.

Moreover, simulation experiments are devised to compare the finite sample performance of the EL tests with the IVX-based test. In contrast to the proposed EL method, the IVX-based test has no test statistic, and its limiting distribution for the null hypothesis $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$ in (4). For a fair comparison, we consider the case with an unknown break date and test the null hypothesis $\mathbb{H}_0 : \beta_1 = 0$; that is, we test for stock return predictability in the pre-break subsample. We report the finite sample rejection frequencies of the IVX-based test by Kostakis et al. (2015) along with those of the EL test. Table 5 presents the

Table 4: Finite sample powers for the test based on Theorem 1 and model (3) with $\theta = 0$ and normally distributed errors for testing $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$. We take $\psi_j = 0.5^j$ for $j \geq 0$ in $\sum_{j=0}^{\infty} \psi_j v_{t-j}$.

		$T = 200$		$T = 500$		$T = 1000$	
		$\lambda_0 = 0.5$	$\lambda_0 = 0.7$	$\lambda_0 = 0.5$	$\lambda_0 = 0.7$	$\lambda_0 = 0.5$	$\lambda_0 = 0.7$
(i) (b1) holds							
$\phi = 0.9$	$b = 1$	0.275	0.246	0.269	0.203	0.263	0.196
	$b = 2$	0.832	0.586	0.863	0.589	0.853	0.588
	$b = 4$	0.999	0.990	1.000	0.994	1.000	0.999
	$b = 8$	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 1 - 2T^{-1}$	$b = 1$	0.807	0.575	0.987	0.841	1.000	0.965
	$b = 2$	0.994	0.915	1.000	0.996	1.000	1.000
	$b = 4$	1.000	0.999	1.000	1.000	1.000	1.000
	$b = 8$	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 1$	$b = 1$	0.875	0.631	0.993	0.867	0.999	0.977
	$b = 2$	0.997	0.938	1.000	0.999	1.000	1.000
	$b = 4$	1.000	0.999	1.000	1.000	1.000	1.000
	$b = 8$	1.000	1.000	1.000	1.000	1.000	1.000
(ii) (b2) holds							
$\phi = 0.9$	$b = 1$	0.457	0.483	0.468	0.477	0.465	0.486
	$b = 2$	0.999	0.998	0.998	0.999	0.996	0.997
	$b = 4$	1.000	1.000	1.000	1.000	1.000	1.000
	$b = 8$	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 1 - 2T^{-1}$	$b = 1$	0.991	0.994	1.000	1.000	1.000	1.000
	$b = 2$	1.000	1.000	1.000	1.000	1.000	1.000
	$b = 4$	1.000	1.000	1.000	1.000	1.000	1.000
	$b = 8$	1.000	1.000	1.000	1.000	1.000	1.000
$\phi = 1$	$b = 1$	0.995	0.998	1.000	1.000	1.000	1.000
	$b = 2$	1.000	1.000	1.000	1.000	1.000	1.000
	$b = 4$	1.000	1.000	1.000	1.000	1.000	1.000
	$b = 8$	1.000	1.000	1.000	1.000	1.000	1.000

Note: The DGP is $y_t = (\alpha_1 + \beta_1 x_{t-1})\, \mathbf{1}_{t \leq T_1^0} + (\alpha_2 + \beta_2 x_{t-1})\, \mathbf{1}_{t > T_1^0} + u_t$, where $x_t = \phi\, x_{t-1} + \sum_{j=0}^{\infty} \psi_j v_{t-j}$ with $\psi_j = 0.5^j$ for $j \geq 0$. See notes to Tables 1 and 3.

size and power properties of the aforementioned tests. The DGP is specified as follows: $y_t = (\alpha_1 + \beta_1 x_{t-1}) + u_t$, $x_t = \phi\, x_{t-1} + v_t$ with $\lambda_0 = 0.5$, $\alpha_1 = aT^{-1/2}$ with $a \in \{4, 8, 12, 16\}$, and $\beta_1 = bT^{-1/2}$ with $b \in \{0, 1, 2, 4, 8\}$. We first consider the case of $b = 0$ for size comparison. For the noise components, $v_t \sim N(0, 1)$ and $u_t = \rho v_t + (1 - \rho^2)^{1/2} \epsilon_t$ with $\epsilon_t \sim N(0, 1)$. The sample sizes are $T = 200, 500, 1000$. The proposed EL methods have empirical sizes close to the nominal level of 5%.

The IVX-based test reveals several interesting features. If the predictor variable x_t is (nearly) integrated, the IVX-based test can be somewhat liberal because the finite sample rejection frequencies are above the nominal level of 5%. Size distortions for the IVX-based

Table 5: Comparison of finite sample sizes and powers

(a, ϕ)	$b=0$ IVX	$b=0$ EL	$b=1$ IVX	$b=1$ EL	$b=2$ IVX	$b=2$ EL	$b=4$ IVX	$b=4$ EL	$b=8$ IVX	$b=8$ EL
					$T=200$					
(4, 0.9)	0.073	0.050	0.193	0.124	0.404	0.387	0.412	0.838	0.479	0.993
(8, 0.9)	0.060	0.060	0.235	0.136	0.632	0.482	0.586	0.925	0.516	0.993
(12, 0.9)	0.045	0.059	0.245	0.133	0.778	0.495	0.780	0.968	0.592	0.996
(16, 0.9)	0.044	0.059	0.245	0.131	0.833	0.517	0.928	0.988	0.705	0.996
$(4, 1-2T^{-1})$	0.093	0.055	0.376	0.319	0.482	0.719	0.640	0.955	0.772	0.999
$(8, 1-2T^{-1})$	0.072	0.054	0.497	0.395	0.536	0.754	0.652	0.961	0.773	0.998
$(12, 1-2T^{-1})$	0.066	0.058	0.622	0.476	0.631	0.794	0.692	0.958	0.786	0.998
$(16, 1-2T^{-1})$	0.070	0.053	0.720	0.513	0.729	0.835	0.721	0.966	0.791	0.999
(4, 1)	0.090	0.054	0.520	0.460	0.665	0.824	0.802	0.972	0.886	0.999
(8, 1)	0.076	0.066	0.554	0.486	0.683	0.837	0.800	0.972	0.887	0.999
(12, 1)	0.079	0.067	0.633	0.543	0.704	0.841	0.818	0.972	0.899	0.999
(16, 1)	0.080	0.069	0.683	0.578	0.723	0.843	0.813	0.974	0.891	0.999
					$T=500$					
(4, 0.9)	0.059	0.055	0.161	0.141	0.388	0.442	0.340	0.947	0.388	1.000
(8, 0.9)	0.041	0.050	0.229	0.150	0.692	0.508	0.520	0.992	0.405	1.000
(12, 0.9)	0.042	0.054	0.258	0.143	0.835	0.517	0.820	0.999	0.473	1.000
(16, 0.9)	0.041	0.057	0.254	0.147	0.858	0.533	0.963	1.000	0.605	1.000
$(4, 1-2T^{-1})$	0.076	0.046	0.409	0.567	0.532	0.894	0.701	0.996	0.824	1.000
$(8, 1-2T^{-1})$	0.068	0.046	0.512	0.627	0.575	0.902	0.728	0.997	0.832	1.000
$(12, 1-2T^{-1})$	0.054	0.050	0.639	0.690	0.614	0.911	0.724	0.997	0.834	1.000
$(16, 1-2T^{-1})$	0.066	0.046	0.736	0.761	0.692	0.924	0.724	0.998	0.836	1.000
(4, 1)	0.078	0.045	0.552	0.700	0.699	0.939	0.834	0.997	0.909	1.000
(8, 1)	0.063	0.055	0.590	0.702	0.721	0.928	0.840	0.998	0.903	1.000
(12, 1)	0.063	0.060	0.621	0.724	0.717	0.924	0.823	0.996	0.909	1.000
(16, 1)	0.068	0.057	0.667	0.767	0.731	0.943	0.835	0.997	0.910	1.000
					$T=1000$					
(4, 0.9)	0.061	0.052	0.149	0.138	0.379	0.454	0.302	0.981	0.329	1.000
(8, 0.9)	0.043	0.484	0.228	0.165	0.680	0.511	0.501	0.997	0.341	1.000
(12, 0.9)	0.038	0.052	0.252	0.149	0.820	0.531	0.838	0.999	0.362	1.000
(16, 0.9)	0.037	0.043	0.254	0.168	0.848	0.543	0.979	1.000	0.506	1.000
$(4, 1-2T^{-1})$	0.086	0.049	0.450	0.763	0.590	0.972	0.748	1.000	0.848	1.000
$(8, 1-2T^{-1})$	0.068	0.043	0.503	0.776	0.595	0.977	0.752	1.000	0.844	1.000
$(12, 1-2T^{-1})$	0.061	0.043	0.581	0.804	0.629	0.978	0.753	1.000	0.858	1.000
$(16, 1-2T^{-1})$	0.057	0.045	0.687	0.851	0.674	0.980	0.772	1.000	0.857	1.000
(4, 1)	0.077	0.051	0.590	0.823	0.743	0.984	0.860	1.000	0.922	1.000
(8, 1)	0.052	0.058	0.608	0.828	0.739	0.985	0.863	1.000	0.923	1.000
(12, 1)	0.059	0.061	0.649	0.832	0.749	0.980	0.858	1.000	0.921	1.000
(16, 1)	0.065	0.063	0.663	0.850	0.766	0.980	0.862	1.000	0.925	1.000

Note: The DGP is $y_t = (\alpha_1 + \beta_1 x_{t-1})\,\mathbf{1}_{t \le T_1^0} + u_t$ and $x_t = \phi\, x_{t-1} + v_t$, where $\lambda_0 = 0.5$, $\alpha_1 = a\,T^{-1/2}$ with $a \in \{4, 8, 12, 16\}$, and $\beta_1 = b\,T^{-1/2}$ with $b \in \{0, 1, 2, 4, 8\}$. Note $b = 0$ for size comparison. For the noise components, $v_t \sim i.i.d.\,N(0,1)$ and $u_t = \rho\, v_t + \sqrt{1-\rho^2}\,\eta_t$ with $\eta_t \sim i.i.d.\,N(0,1)$. Rejection frequencies are reported for testing $\mathbb{H}_0 : \beta_1 = 0$ with level 5%.

test remain, even when the sample size and magnitude of the level shift are relatively large. Overall, EL methods are remarkable in terms of size control, which confirms their validity, allowing a structural break in predictive regression. For $b > 0$, we examine the power properties of the aforementioned tests without size adjustment because there is no oversizing in the proposed EL tests. Surprisingly, the IVX-based test shows non-monotonic power against the alternative hypothesis of predictability such that the power of the test can decrease as the magnitude of change increases. For example, when $\phi = 0.9$, the power of the IVX-based test decreases as the magnitude of break b becomes greater than two, regardless of the sample size. This finding highlights that the IVX-based test statistic may not be reliable when there exists at least a structural break in the predictive regressions.

In simulation experiments, we consider predictive regressions under practical assumptions on the predictor variable and errors: (i) there exists a structural break in the predictive regression model, (ii) the predictor variable can be stationary or nonstationary, and (iii) the errors (u_t, v_t) are contemporaneously correlated. The simulation results appear to support the proposed EL methods having good finite sample properties in terms of both size and power. The results shed new light on testing for stock return predictability because the EL methods can help address various time-series characteristics of the predictor variable and incorporate a structural break on an unknown date in predictive regressions. Because the IVX-based test requires a complicated implementation and, more importantly, does not provide a theoretical justification for the joint null hypothesis of no predictability; that is, $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$, the proposed EL method can be a useful complement to testing for asset return predictability.

It is noteworthy that the powers of the EL tests are very close to 1 if the magnitude of the level shift is sufficiently large, $\alpha_2 = 16T^{-1/2}$, in the predictive regression. This result indicates that the magnitude of the change in predictive regression plays an important role in the finite sample properties of the proposed EL tests. For further investigation, we allowed the values of β_1 to range from 1 to 1 in increments of 0.1. For other parameters, we set $\beta_2 = 0$, $\rho = -0.95$, $\theta = 0$, $\phi = 0.99$, and $T = 150$. The noise components $(v_t, \epsilon_t) \sim i.i.d.\, N(0, 1)$. The break date can be estimated by minimizing the SSR, as described in (5). For the null hypothesis of $\beta_2 = 0$, rejection frequencies are shown in Figure 1. Because we set $\alpha_1 = \alpha_2 = 0$ in the DGP, the EL method with known α, which is free from splitting the data into two parts, is well defined (labeled EL1) whereas the EL method with unknown α, $\ddot{l}_2(\beta_2)$ allows for an irrelevant intercept (EL2). We simulate the following statistics: EL1 with $\hat{T}_1$, EL1 with T_1^0, EL2 with $\hat{T}_1$, and EL2 with T_1^0. The tests EL1 with $\hat{T}_1$ and EL2 with $\hat{T}_1$ suffer

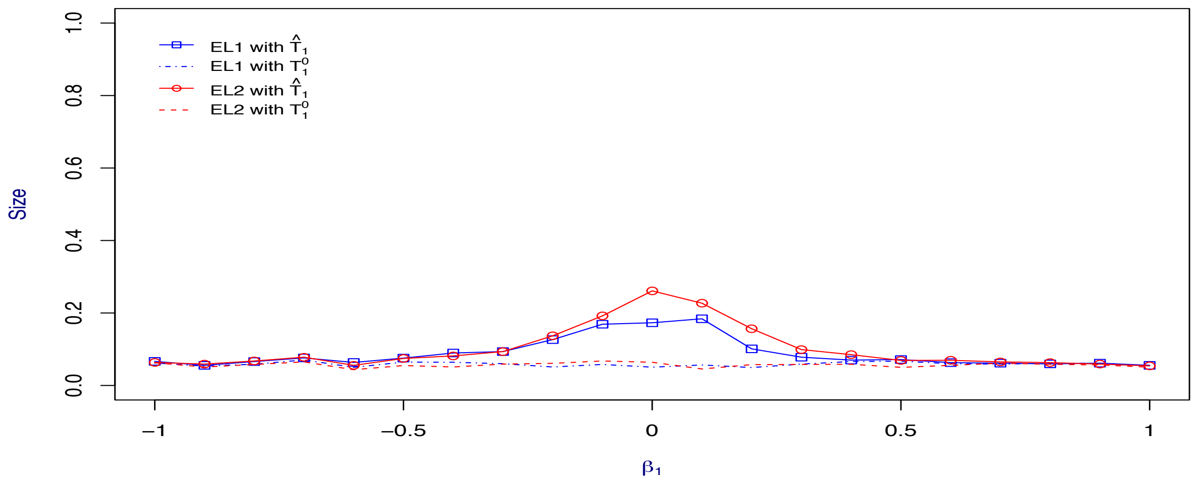

Figure 1: Rejection Probabilities of EL tests for $\mathbb{H}_0 : \beta_2 = 0$

Note: The DGP is $y_t = \beta_1 \, x_{t-1} \mathbf{1}_{t \leq [\lambda_0 T]} + u_t$ with $\lambda_0 = 0.5$, $x_t = 0.99 \, x_{t-1} + v_t$, $u_t = \rho \, v_t + (1 - \rho^2)^{1/2} \, \epsilon_t$, $\rho = -0.95$, $(v_t, \epsilon_t) \sim i.i.d. \, N(0, 1)$, and $T = 150$. Rejection frequencies are reported for testing $\mathbb{H}_0 : \beta_2 = 0$ as β_1 varies from -1 to 1.

from some liberal size distortions unless $|\beta_1|$ is large. This suggests that the estimates $\hat{T}_1$ are variable when $|\beta_1|$ is not sufficiently large. This randomness translates into distributions of EL tests constructed with $\hat{T}_1$, which are far from the true break date case, and liberal size distortions occur. Compared to test EL2 with $\hat{T}_1$, test EL1 with $\hat{T}_1$, which uses the fact that the intercept terms α_1 and α_2 are known, is less affected by this problem, although some size distortions remain. The finite sample properties of EL methods emphasize that testing for and estimating the break date in the predictive regression is crucial to make EL tests statistically reliable.

4 Empirical Analysis

In this section, we apply the EL method to test stock return predictability. Specifically, the predictable variable y_t is the monthly S&P 500 value-weighted log excess return, and the predictor variable x_t includes the log dividendprice (d/p) ratio, log earningsprice (e/p) ratio, dividend yield (dy), book-to-market (b/m) ratio, treasury bill rates (tbl), default yield spread (dfy), log dividend payout (d/e) ratio, net equity expansion ($ntis$), and term spread (tms). The dataset contains monthly data and spans 1927:01 to 2012:12; hence, the sample size is $T = 1032$. The monthly returns are computed by the difference between the S&P 500 index, including dividends, and the one-month Treasury bill rate. Data from Goyal and Welch (2008) have been widely used in the predictive regression literature, such as Cenesizoglu and

Timmermann (2008), Maynard et al. (2011), Kostakis et al. (2015), Liu et al. (2019), and Lee (2016) and Fang and Lee (2019) in a quantile regression framework. Distinguished from existing methodologies such as Lewellen (2004), Campbell and Yogo (2006), Cai and Wang (2014), it is unnecessary for EL methods to evaluate the persistence of the predictor variable x_t. In fact, based on unit root tests, both *d/p* and *e/p* ratios are I(1) or NI(1) processes (see Cai et al. (2015) for details).

Table 6 presents the EL test results where the p-values for testing $\mathbb{H}_0 : \beta = 0$ and $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$ are reported. The EL2 statistic in Zhu et al. (2014) is designed to test

Table 6: p-values of predictability tests on the S&P 500 excess returns (1927:01 – 2012:12)

	EL2 $\mathbb{H}_0 : \beta = 0$	$\tilde{l}_\beta(\beta_1, \beta_2)$ $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$
d/p	0.1495	0.4484
e/p	**0.0256**	0.7036
dy	**0.0798**	0.3410
b/m	**0.0844**	0.6071
tbl	0.6702	**0.0725**
dfy	0.5680	0.6288
d/e	0.9900	0.3750
ntis	0.4951	0.7736
tms	0.9293	0.3199

Note: This table reports p-values of EL tests for the null hypothesis of no predictability. The predictor variables are the log dividend-price (*d/p*), log earnings-price (*e/p*) ratios, dividend yeild (*dy*), book-to-market (*b/m*) ratio, Treasury-bill rate (*tbl*), default yield spread (*dfy*), log dividend-payout (*d/e*) ratio,net equity expansion (*ntis*), and term spread (*tms*). The EL2 statistics in Zhu et al. (2014) is designed to test the null hypothesis of no predictability, $\mathbb{H}_0 : \beta = 0$ with an unknown intercept α whereby non level shift is allowed. The $\tilde{l}_\beta(\beta_1, \beta_2)$ statistic also tests the null hypothesis of no predictability, $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$, allowing for a level shift. We estimate the date of a level shift, 1942:04, and use it to implement $\tilde{l}_\beta(\beta_1, \beta_2)$. The results in bold indicate the rejection of the null hypothesis of no predictability at the 10% significance level.

the null hypothesis of no predictability $\mathbb{H}_0 : \beta = 0$ with an unknown intercept α whereby no level shift is allowed. The $\tilde{l}_\beta(\beta_1, \beta_2)$ statistic also tests the null hypothesis of no predictability, $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$, allowing for a level shift. As the former null hypothesis can be rejected because of different intercepts, we can infer whether the valuation ratios considered induce episodic predictability.

Specifically, we estimate the date of a level shift in the monthly S&P 500 value-weighted log excess return, 1942:04, and use it to implement $\tilde{l}_\beta(\beta_1, \beta_2)$. The results in bold indicate the rejection of the null hypothesis of no predictability at the 10% significance level. Focusing first on the *dy* series, we find that, based on the EL2 statistic, the null hypothesis of no predictability induced by *dy* is rejected with a p-value of 0.0798. On the other hand, our new test statistic leads to a non-rejection of the null hypothesis $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$, which implies that predictability over the full sample is likely to occur because of unequal intercepts arising from macroeconomic policy changes rather than the *dy* predictor. This is consistent with the preceding literature, which claims that the predictability of dividend yield has declined because of greater dividend smoothing. Moreover, we find similar patterns for the *e/p* and *b/m* predictors, suggesting there is very little evidence of regime-specific predictability, consistent with what has been documented in the predictive regression literature (see Kostakis et al. (2015), Gonzalo and Pitarakis (2017), and Liu et al. (2019)).

We find an interesting feature of Treasury-bill rates (*tbl*) as a predictor variable. EL2 cannot reject the null hypothesis $\mathbb{H}_0 : \beta = 0$ while the $\tilde{l}_\beta(\beta_1, \beta_2)$ statistic rejects the null hypothesis $\mathbb{H}_0 : \beta_1 = \beta_2 = 0$ with a p-value of 0.0725. This result strongly supports the idea that predictability is driven by the *tbl* predictor rather than a level shift. Recently, Kostakis et al. (2015) considered the sub-period 1952:012012:12 and found predictability for the predicting variables of T-bill (*tbl*) and term spread (*tms*) at the 10% significance level. The Q-test of Campbell and Yogo (2006) finds that *tbl*, *tms*, and *dy* are marginally significant at the 10% level. As evidence that predictability has diminished in recent sample periods (see, for example, Campbell and Yogo, 2006), it is noteworthy that *tbl* has more information on stock return predictability. By contrast, there is no evidence of predictability for the *d/p*, *dfy*, *d/e*, *ntis*, and *tms* from the $\tilde{l}_\beta(\beta_1, \beta_2)$ statistics. For the period 1927:012012:12, the Q-test of Campbell and Yogo (2006) showed that *dy*, *bm*, and *ntis* are significant at the 10% level while the IVX-based test of Kostakis et al. (2015) found that the null of no predictability can be rejected at the 5% level only for *e/p*, *b/m*, and *ntis* and at the 10% level for *dy*. The Q-test and IVX-based test are designed for persistent predictor variables without allowing for the possibility of a level shift in the predictive regression. In particular, the Bonferroni-type test (Q-test) is valid if the predictor variable is as persistent as an NI(1) process with strong assumptions, such as normality and known covariance of innovations. Hence, the test results based on existing methods may be misleading if a level shift exists. The empirical findings reveal the merits of EL methods that complement the existing methods.

Before finishing this section, we must briefly discuss the stationarity assumption for the predictor variables. To understand the implications of this assumption, we examine it in two different ways. From an economic perspective, the predictor variables should be stationary; that is, $|\phi| < 1$. Otherwise, an explosive bubble may exist in stock prices. In other words, the predicted variable cannot be stationary if the predictor variable is nonstationary, which is still a controversial topic in empirical studies. As noted in Lettau and Van Nieuwerburgh (2008), a level shift in the predictor variable is also a form of nonstationarity. From a statistical perspective, it is preferable to make tests reliable regardless of whether the predictor variables are stationary or nonstationary. However, the test results cannot identify the time-series characteristic of the predictor variable. Because it makes little sense to predict asset returns with an integrated process, further investigation of the predictor variable might be required (see also Liu et al. (2019)).

5 Concluding Remarks

In this study, we consider predictive regression models in which model parameters exhibit a structural break on an unknown date and the predictor variable is allowed to be either stationary or (nearly) integrated. As addressed by Viceira (1997), Paye and Timmermann (2006), and Rapach and Wohar (2006), parameter instability in predictive regressions is a long-standing problem. The main contribution of this study to the literature is to provide a unified approach that is valid regardless of the time-series characteristics of the predictor variable when the model parameters are unstable in predictive regressions. We established novel testing procedures for asset return predictability using EL methods based on weighted score equations and studied their asymptotic distributions. The proposed methods are particularly useful because they allow us to distinguish predictability generated by a certain predictor variable from that induced by a permanent shift in level, and they require no prior knowledge as to (i) whether the predictor variable is stationary or nonstationary and (ii) whether the noise components are contemporaneously correlated or not, common cases in practice. Simulations have demonstrated the usefulness of EL methods and demonstrated clear improvements over existing tests. Furthermore, an extension of practical interest is to generalize our model to analyze a large global dataset, which is useful for capturing the predictable component of stock returns (see, for example, Paye and Timmermann, 2006; Hjalmarsson, 2010). Such investigations, among others, are objects of ongoing research.

References

Amihud, Y. and C. Hurvich (2004). Predictive regressions: A reduced-bias estimation method. *Journal of Financial and Quantitative Analysis 39*, 813–841.

Amihud, Y., C. Hurvich, and Y. Wang (2009). Multiple-predictor regressions: Hypothesis testing. *Review of Economic Studies 22*, 413–434.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*, 821–856.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*, 1383–1414.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis 15*, 453–472.

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics 79*, 551–563.

Bai, J. (1998). A note on spurious break. *Econometric Theory 14*, 663–669.

Bai, J., R. L. Lumsdaine, and J. H. Stock (1998). Testing for and dating common breaks in multivariate time series. *Review of Economic Studies 65*, 395–432.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*, 1–22.

Bai, J. and P. Perron (2006). Multiple structural change models: A simulation analysis. *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, 212–237. in: Corbae, D., Durlauf, S.N., Hansen, B.E. (Eds.), Cambridge Press.

Breitung, J. and M. Demetrescu (2015). Instrumental variable and variable addition based inference in predictive regressions. *Journal of Econometrics 187*, 358–375.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (2nd ed.). New York: Springer-Verlag.

Cai, Z. and Y. Wang (2014). Testing predictive regression models with nonstationary regressors. *Journal of Econometrics 178*, 4–14. Corrigendum, *Journal of Econometrics 181*, 194.

Cai, Z., Y. Wang, and Y. Wang (2015). Testing instability in a predictive regression model with nonstationary regressors. *Econometric Theory 31*, 953–980.

Campbell, J. Y. (2008). Viewpoint: Estimating the equity premium. *Canadian Journal of Economics 41*, 1–21.

Campbell, J. Y., A. W. Lo, and A. C. Mackinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press.

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

Cavanagh, C. L., G. Elliott, and J. H. Stock (1995). Inference in models with nearly integrated regressors. *Econometric Theory 11*, 1131–1147.

Cenesizoglu, T. and A. Timmermann (2008). Is the distribution of stock returns predictable? *Unpublished Working Paper*. HEC Montreal and UCSD. Available at SSRN 1107185, 2008 - papers.ssrn.com.

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

Chan, N. H. and L. Peng (2005). Weighted least absolute deviations estimation for an AR(1) process with ARCH(1) errors. *Biometrika 92*, 477–484.

Chan, N. H. and C. Z. Wei (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics 15*, 1050–1063.

Chang, S. Y. and P. Perron (2016). Inference on a structural break in trend with fractionally integrated errors. *Journal of Time Series Analysis 37*, 555–574.

Chang, S. Y. and P. Perron (2018). A comparison of alternative methods to construct confidence intervals for the estimate of a break date in linear regression models. *Econometric Reviews 37*, 577–601.

Chen, W. W. and R. S. Deo (2009). Bias reduction and likelihood-based almost exactly sized hypothesis testing in predictive regressions using the restricted likelihood. *Econometric Theory 25*, 1143–1179.

Demetrescu, M. and P. Rodrigues (2020). Residual-augmented ivx predictive regression. *Journal of Econometrics DOI: http://dx.doi.org/10.1016/j.jeconom.2020.11*.

Elliott, G. (2011). A control function approach for testing the usefulness of trending variables in forecast models and linear regression. *Journal of Econometrics 164*, 79–91.

Elliott, G. and U. K. Müller (2006). Efficient tests for general persistent time variation in regression coefficients. *Review of Economic Studies 73*, 907–940.

Elliott, G. and U. K. Müller (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics 141*, 1196–1218.

Elliott, G. and J. H. Stock (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric Theory 10*, 672–700.

Fang, R. and J. H. Lee (2019). Predictive quantile regressions under persistence and conditional heteroskedasticity. *Journal of Econometrics 213*, 261–280.

Gonzalo, J. and J.-Y. Pitarakis (2012). Regime-specific predictability in predictive regressions. *Journal of Business & Economic Statistics 30*, 229–241.

Gonzalo, J. and J.-Y. Pitarakis (2017). Inferring the predictability induced by a persistent regressor in a predictive threshold model. *Journal of Business & Economic Statistics 35*, 202–217.

Goyal, A. and I. Welch (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*, 1455–1508.

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

Hjalmarsson, E. (2010). Predicting global stock returns. *Journal of Financial and Quantitative Analysis 45*, 49–80.

Hosseinkouchack, M. and M. Demetrescu (2021). Finite-sample size control of ivx-based tests in predictive regressions. *Econometric Theory 37*, 769–793.

Hsu, Y.-C. and C.-M. Kuan (2008). Change-point estimation of nonstationary I(d) processes. *Economic Letters 98*, 115–121.

Jansson, M. and M. J. Moreira (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica 74*, 681–714.

Kostakis, A., T. Magdalinos, and M. P. Stamatogiannis (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies 28*, 1506–1553.

Kuan, C.-M. and C.-C. Hsu (1998). Change-point estimation of fractionally integrated processes. *Journal of Time Series Analysis 19*, 693–708.

Kurozumi, E. and Y. Arai (2006). Efficient estimation and inference in cointegrating regressions with structural change. *Journal of Time Series Analysis 28*, 545–575.

Kurozumi, E. and Y. Arai (2007). Testing for the null hypothesis of cointegration with a structural break. *Econometric Reviews 26*, 705–739.

Lee, J. H. (2016). Predictive quantile regression persistent covariates: IVX-QR approach. *Journal of Econometrics 192*, 105–118.

Lettau, M. and S. Van Nieuwerburgh (2008). Reconciling the return predictability evidence. *Review of Financial Studies 21*, 1607–1652.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics 74*, 209–235.

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

Liao, X., Z. Cai, and H. Chen (2018). A perspective on recent models for testing predictability of asset returns. *Applied Mathematics – A Journal of Chinese Universities, Series B 33*, 127–144.

Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society: Series B 67*, 381–393.

Liu, X., B. Yang, Z. Cai, and L. Peng (2019). A unified test for predictability of asset returns regardless of properties of predicting variables. *Journal of Econometrics 208*, 141–159.

Maynard, A., K. Shimotsu, and Y. Wang (2011). Inference in predictive quantile regressions. *Unpublished Working Paper*. University of Guelph.

Nunes, L. C., C.-M. Kuan, and P. Newbold (1995). Spurious break. *Econometric Theory 11*, 736–749.

Owen, A. B. (2001). *Empirical Likelihood.* Chapman & Hall, New York.

Paye, B. S. and A. Timmermann (2006). Instability of return prediction models. *Journal of Empirical Finance 13*, 274–315.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica 57*, 1361–1401.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrica 74*, 535–547.

Phillips, P. C. B. (2015). Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics 13*, 521–555.

Phillips, P. C. B. and J. H. Lee (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics 177*, 250–264.

Phillips, P. C. B. and T. Magdalinos (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics 136*, 115–130.

Phillips, P. C. B. and T. Magdalinos (2009). Econometric inference in the vicinity of unity. *CoFie Working Paper No.7*. Singapore Management University.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating functions. *Annals of Statistics 22*, 300–325.

Rapach, D. E. and M. E. Wohar (2006). Structural breaks and predictive regression models of aggregate U.S. stock returns. *Journal of Financial Econometrics 4*, 238–274.

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

Stambaugh, R. (1999). Predictive regressions. *Journal of Financial Economics 54*, 375–421.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*, 11–30.

Torous, W., R. Valkanov, and S. Yan (2005). On predicting stock returns with nearly integrated explanatory variables. *Journal of Business 77*, 937–966.

Viceira, L. M. (1997). Testing for structural change in the predictability of asset returns. *Unpublished Working Paper*. Harvard University.

Yang, B., X. Liu, L. Peng, and Z. Cai (2021). Unified tests for a dynamic predictive regression. *Journal of Business & Economic Statistics 39*, 684–699.

Yang, B., W. Long, L. Peng, and Z. Cai (2020). Testing the predictability of us housing price index returns based on an ivx-ar model. *Journal of The American Statistical Association 115*, 1598–1619.

Zhou, M. (2015). Empirical likelihood ratio for censored/truncated data. *R package version 1.0*, http://CRAN.R–project.org/package=emplik.

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

Zhu, F., M. Liu, S. Ling, and Z. Cai (2021). Testing for structural change of predictive regressionmodel to threshold predictive regressionmodel. *Journal of Business & Economic Statistics*, DOI: https://doi.org/10.1080/07350015.2021.2008406.

Appendix: Mathematical Proofs

To prove Theorem 1, we consider the following predictive regression model with a structural break on an unknown date T_1^0 for $t = 1, \ldots, T$,

$$\begin{cases} y_t = (\alpha_1 + \beta_1 x_{t-1})\, \mathbf{1}_{t \leq T_1^0} + (\alpha_2 + \beta_2 x_{t-1})\, \mathbf{1}_{t > T_1^0} + u_t, \\ .x_t = \theta + \phi\, x_{t-1} + \sum_{j=0}^{\infty} \psi_j v_{t-j}, \end{cases}$$

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

$$\theta = 0, \quad \phi = 1 - c\, T^{-1} \quad \text{for some} \ \ c \in \mathbb{R} \tag{A.1}$$

because the proofs for the other cases are straightforward. In particular, when x_t is stationary, we can prove Theorem 1 based on the weak law of large numbers, the central limit theorem for martingales in Hall and Heyde (1980), and the standard arguments for establishing the profile EL method based on estimating equations in Qin and Lawless (1994). We state the following lemmas before proving Theorem 1: All limit statements are considered as $T \to \infty$ whenever there is no confusion.

The estimate of the break date $\hat{T}_1$ can be obtained using a global least squares criterion, as explained in (5). Here, we consider the case in which $\hat{T}_1 < T_1^0$; that is, the estimated break date is located in the pre-T_1^0 subsample. Owing to symmetry, the same arguments can be applied to the case where $\hat{T}_1 > T_1^0$. We consider two subsamples: the pre-$\hat{T}_1$ subsample and the post-$\hat{T}_1$ subsample. We define this set as follows:

$$V(\epsilon) = \{T_1 : |T_1 - T_1^0| < \epsilon, \, \forall \epsilon > 0\},$$

From Proposition 1, $\Pr(\hat{T}_1 \in V(\epsilon)) \to 1$, regardless of whether x_t is stationary or (nearly) integrated. Hence, it suffices to consider the behavior of EL methods for all $T_1 \in V(\epsilon)$. Let $m_1 \equiv [T_1/2]$ and $m_2 \equiv [(T - T_1)/2]$.

Lemma 1. *Under the conditions of Theorem 1 and (A.1), we have*

$$\frac{1}{\sqrt{T}} \sum_{t=2}^{m_1} \tilde{H}_t(\beta_{1,0}) = \left(\frac{1}{\sqrt{T}} \sum_{t=2}^{m_1} \tilde{u}_t \right) \frac{\tilde{x}_{m_1 - 1}}{\sqrt{1 + \tilde{x}_{m_1-1}^2}} + o_p(1)$$

and

$$\frac{1}{\sqrt{T}} \sum_{t=T_1+2}^{T_1+m_2} \tilde{H}_t(\beta_{2,0}) = \left(\frac{1}{\sqrt{T}} \sum_{t=T_1+2}^{T_1+m_2} \tilde{u}_t \right) \frac{\tilde{x}_{T_1+m_2-1}}{\sqrt{1 + \tilde{x}_{T_1+m_2-1}^2}} + o_p(1),$$

where $\tilde{H}_t(\beta_{j,0}) = (\tilde{y}_t - \beta_{j,0}\, \tilde{x}_{t-1})\, \tilde{x}_{t-1}/(1 + \tilde{x}_{t-1}^2)^{1/2}$ *and* $\beta_{j,0}$ *denotes the true value of* β_j *for* $j = 1, 2$.

Proof of Lemma 1. It follows from Phillips (1987) under (A.1),

$$T^{-1/2}x_{[Tr]} \Rightarrow K_c(r), \tag{A.2}$$

where $K_c(r) = \int_0^r e^{-(r-s)c}dW_u(s)$ is a diffusion process and $W_u(s)$ is a one-dimensional Brownian motion with variance $\sigma_u^2 = \mathrm{Var}(u_t) + 2\Omega_1$, and $\Omega_1 = \sum_{k=2}^{\infty} \mathbf{E}[u_1 u_k]$. Here, "$\Rightarrow$" denotes a weak convergence in the Skorohod topology. From (A.2), we obtain

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=2}^{m_1}\tilde{H}_t(\beta_{1,0}) &= \frac{1}{\sqrt{T}}\sum_{t=2}^{m_1}\left(\sum_{j=1}^{t}\tilde{u}_j - \sum_{j=1}^{t-1}\tilde{u}_j\right)\frac{\tilde{x}_{t-1}}{\sqrt{1+\tilde{x}_{t-1}^2}} \\
&= \frac{1}{\sqrt{T}}\sum_{j=1}^{m_1}\tilde{u}_j\frac{\tilde{x}_{m_1-1}}{\sqrt{1+\tilde{x}_{m_1-1}^2}} + \frac{1}{\sqrt{T}}\sum_{t=2}^{m_1-1}\left(\sum_{j=1}^{t}\tilde{u}_j\right)\frac{\tilde{x}_{t-1}}{\sqrt{1+\tilde{x}_{t-1}^2}} - \frac{1}{\sqrt{T}}\sum_{t=2}^{m_1}\left(\sum_{j=1}^{t-1}\tilde{u}_j\right)\frac{\tilde{x}_{t-1}}{\sqrt{1+\tilde{x}_{t-1}^2}} \\
&= \left(\frac{1}{\sqrt{T}}\sum_{t=2}^{m_1}\tilde{u}_t\right)\frac{\tilde{x}_{m_1-1}}{\sqrt{1+\tilde{x}_{m_1-1}^2}} + \frac{1}{\sqrt{T}}\sum_{t=2}^{m_1-1}\left(\sum_{j=1}^{t}\tilde{u}_j\right)\left(\frac{\tilde{x}_{t-1}}{\sqrt{1+\tilde{x}_{t-1}^2}} - \frac{\tilde{x}_t}{\sqrt{1+\tilde{x}_t^2}}\right) + o_p(1).
\end{aligned} \tag{A.3}$$

It follows from Taylor expansion that

$$\frac{\tilde{x}_{t-1}}{\sqrt{1+\tilde{x}_{t-1}^2}} - \frac{\tilde{x}_t}{\sqrt{1+\tilde{x}_t^2}} = (1+\xi_t^2)^{-3/2}(\tilde{x}_{t-1} - \tilde{x}_t), \tag{A.4}$$

where ξ_t lies between $\tilde{x}_{t-1}$ and $\tilde{x}_t$. From (A.2), we have $|\tilde{x}_{t-1}|/t^a \xrightarrow{p} \infty$, $|\tilde{x}_t|/t^a \xrightarrow{p} \infty$, and $|\tilde{x}_{t-1} - \tilde{x}_t|/t^a \xrightarrow{p} 0$ for any $a \in (0, 1/2)$ as $t \to \infty$, thereby

$$|\xi_t|/t^a \xrightarrow{p} \infty \quad \text{for any } a \in (0, 1/2) \text{ as } t \to \infty. \tag{A.5}$$

By (A.2), (A.3)-(A.5), we have

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{m_1}\tilde{H}_t(\beta_{1,0}) = \left(\frac{1}{\sqrt{T}}\sum_{t=2}^{m_1}\tilde{u}_t\right)\frac{\tilde{x}_{m_1-1}}{\sqrt{1+\tilde{x}_{m_1-1}^2}} + o_p(1). \tag{A.6}$$

The post-T_1 subsample is divided into two parts around T_1^0. More precisely,

$$y_t = \begin{cases} \alpha_1 + \beta_1\, x_{t-1} + u_t, & \text{for } t = T_1+1, \ldots, T_1^0, \\ \alpha_2 + \beta_2\, x_{t-1} + u_t, & \text{for } t = T_1^0+1, \ldots, T, \end{cases}$$

and

$$y_{t+m_2} = \alpha_2 + \beta_2\, x_{t-1} + u_t, \quad t = T_1+1, \ldots, T_1+m_2.$$

On the set $V(\epsilon)$, it holds that $(t+m_2) > T_1^0$ for $t = T_1+1, \ldots, T_1+m_2$. Hence, $\tilde{y}_t = \beta_2\, \tilde{x}_{t-1} + \tilde{u}_t$, where

$$\tilde{u}_t = \begin{cases} (u_t - u_{t+m_2}) + (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)x_{t-1}, & \text{for } t = T_1+1, \ldots, T_1^0, \\ (u_t - u_{t+m_2}), & \text{for } t = T_1^0+1, \ldots, T_1+m_2. \end{cases}$$

It is straightforward to show that

$$\frac{1}{\sqrt{T}}\sum_{t=T_1+2}^{T_1+m_2}\tilde{H}_t(\beta_{1,0}) = \frac{1}{\sqrt{T}}\sum_{t=T_1+2}^{T_1+m_2}\left(\sum_{j=1}^{t}\tilde{u}_j - \sum_{j=1}^{t-1}\tilde{u}_j\right)\frac{\tilde{x}_{t-1}}{\sqrt{1+\tilde{x}_{t-1}^2}}$$

$$= \frac{1}{\sqrt{T}}\sum_{j=1}^{T_1+m_2}\tilde{u}_j\frac{\tilde{x}_{T_1+m_2-1}}{\sqrt{1+\tilde{x}_{T_1+m_2-1}^2}} + \frac{1}{\sqrt{T}}\sum_{t=T_1+2}^{T_1+m_2-1}\left(\sum_{j=1}^{t}\tilde{u}_j\right)\frac{\tilde{x}_{t-1}}{\sqrt{1+\tilde{x}_{t-1}^2}}$$

$$- \frac{1}{\sqrt{T}}\sum_{t=T_1+2}^{T_1+m_2}\left(\sum_{j=1}^{t-1}\tilde{u}_j\right)\frac{\tilde{x}_{t-1}}{\sqrt{1+\tilde{x}_{t-1}^2}} + \frac{1}{\sqrt{T}}\sum_{t=T_1+2}^{T_1+m_2}[(\alpha_1-\alpha_2)+(\beta_1-\beta_2)x_{t-1}]\frac{\tilde{x}_{t-1}}{\sqrt{1+\tilde{x}_{t-1}^2}}$$

$$= \left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T_1+m_2}\tilde{u}_t\right)\frac{\tilde{x}_{T_1+m_2-1}}{\sqrt{1+\tilde{x}_{T_1+m_2-1}^2}} + \frac{1}{\sqrt{T}}\sum_{t=T_1+2}^{T_1+m_2-1}\left(\sum_{j=1}^{t}\tilde{u}_j\right)\left(\frac{\tilde{x}_{t-1}}{\sqrt{1+\tilde{x}_{t-1}^2}} - \frac{\tilde{x}_t}{\sqrt{1+\tilde{x}_t^2}}\right) + o_p(1), \tag{A.7}$$

where T_1 is in the set $V(\epsilon)$ with probability approaching 1; therefore, the number of terms in the summation $\sum_{T_1+2}^{T_1+m_2}$ is finite, which suggests that $T^{-1/2}\sum_{t=T_1+2}^{T_1+m_2}[(\alpha_1-\alpha_2)+(\beta_1-\beta_2)x_{t-1}]\tilde{x}_{t-1}/(1+\tilde{x}_{t-1}^2)^{1/2} \xrightarrow{p} 0$ because $m_2 = [(T-T_1)/2]$. Using (A.2), (A.4), (A.5) and (A.7), we have

$$\frac{1}{\sqrt{T}}\sum_{t=T_1+2}^{T_1+m_2}\tilde{H}_t(\beta_{2,0}) = \left(\frac{1}{\sqrt{T}}\sum_{t=T_1+2}^{T_1+m_2}\tilde{u}_t\right)\frac{\tilde{x}_{T_1+m_2-1}}{\sqrt{1+\tilde{x}_{T_1+m_2-1}^2}} + o_p(1). \tag{A.8}$$

Hence, the lemma follows from (A.6) and (A.8). ■

Lemma 2. *Under the conditions of Theorem 1 and (A.2), we have*

$$\frac{1}{T}\sum_{t=2}^{m_1}\tilde{H}_t^2(\beta_{1,0}) \xrightarrow{p} \Sigma \quad \text{and} \quad \frac{1}{T}\sum_{t=T_1+2}^{T_1+m_2}\tilde{H}_t^2(\beta_{2,0}) \xrightarrow{p} \Sigma,$$

where $\Sigma = \mathbf{E}[\tilde{u}_1^2]$.

Proof of Lemma 2. Using (A.2), (A.4) and (A.5), we have

$$\frac{1}{T}\sum_{t=2}^{m_1}\tilde{H}_t^2(\beta_{1,0}) = \frac{1}{T}\sum_{t=2}^{m_1}\tilde{u}_t^2\frac{\tilde{x}_{t-1}^2}{1+\tilde{x}_{t-1}^2} = \frac{1}{T}\sum_{t=2}^{m_1}\tilde{u}_t^2 + o_p(1) = \mathbf{E}[\tilde{u}_1^2] + o_p(1).$$

Similarly, we show that $T^{-1}\sum_{t=T_1+2}^{T_1+m_2}\tilde{H}_t^2(\beta_{2,0}) \xrightarrow{p} \Sigma$. Therefore, Lemma 2 holds. ■

Lemma 3. *Suppose that Condition A holds. Then,*

$$\max_{2\le t\le m_1}\|\tilde{H}_t(\beta_{1,0})\| = O_p(T^{1/2}) \quad \text{and} \quad \max_{T_1+2\le t\le T_1+m_2}\|\tilde{H}_t(\beta_{2,0})\| = O_p(T^{1/2}).$$

Proof of Lemma 3. Clearly, $\mathbf{E}[\tilde{H}_t(\beta_{1,0})] = 0$ and $\mathbf{E}[\tilde{H}_t(\beta_{1,0})^2] \leq \mathbf{E}[\tilde{u}_t^2] = O(1)$, implying that

$$\left(\mathbf{E}\left[\max_{2\leq t\leq m_1} \|\tilde{H}_t(\beta_{1,0})\|\right]\right)^2 \leq \mathbf{E}\left[\left(\max_{2\leq t\leq m_1} \|\tilde{H}_t(\beta_{1,0})\|\right)^2\right]$$
$$\leq \sum_{t=2}^{m_1} \mathbf{E}[\tilde{H}_t(\beta_{1,0})^2] = O(T),$$

where the first inequality is implied by Jensen's inequality. As a result,

$$\max_{2\leq t\leq m_1} \|\tilde{H}_t(\beta_{1,0})\| = O_p(T^{1/2}).$$

In a similar way, we can show that

$$\max_{T_1+2\leq t\leq T_1+m_1} \|\tilde{H}_t(\beta_{2,0})\| = O_p(T^{1/2}).$$

This completes the proof of Lemma 3. ■

Proof of Theorem 1. Using Lemmas 1-3 and the standard arguments in the proof of the EL method (Owen, 2001, Chapter 11), both $\tilde{l}_1(\beta_{1,0})$ and $\tilde{l}_2(\beta_{2,0})$ converge in distribution to a chi-square limit with one degree of freedom. Consequently, the EL statistic in each regime is independent of the others, suggesting that their summation goes to $\chi^2(2)$, which completes the proof of Theorem 1. ■

Proof of Proposition 1. For the proof, see Proposition 4 in Bai and Perron (1998) and Proposition 2 in Kurozumi and Arai (2006). ■

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