Testing the Intercept of a Balanced Predictive Regression Model

Qijun Wang 1,2, Xiaohui Liu 1,2, Yawen Fan 1,2,* and Ling Peng 1,2

1 School of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China; 2202021883@stu.jxufe.edu.cn (Q.W.); liuxiaohui@jxufe.edu.cn (X.L.); pengling@jxufe.edu.cn (L.P)
2 Key Laboratory of Data Science in Finance and Economics, Jiangxi University of Finance and Economics, Nanchang 330013, China
* Correspondence: 2202010084@stu.jxufe.edu.cn

Abstract: Testing predictability is known to be an important issue for the balanced predictive regression model. Some unified testing statistics of desirable properties have been proposed, though their validity depends on a predefined assumption regarding whether or not an intercept term nevertheless exists. In fact, most financial data have endogenous or heteroscedasticity structure, and the existing intercept term test does not perform well in these cases. In this paper, we consider the testing for the intercept of the balanced predictive regression model. An empirical likelihood based testing statistic is developed, and its limit distribution is also derived under some mild conditions. We also provide some simulations and a real application to illustrate its merits in terms of both size and power properties.

Keywords: balanced predictive regression model; intercept; empirical likelihood; stationary; non-stationary

JEL Classification: C12; C22

1. Introduction

Prediction of some important variables based on the other available predictors is a practice of great interest in many applications in finance and economics. Various predictive regression models have been suggested in the last few decades. Among them, the simplest one is the following linear predictive regression model:

\[ \begin{align*}
Y_t &= \alpha + \beta X_{t-1} + U_t \\
X_t &= \mu + \phi X_{t-1} + \epsilon_t, \quad B(L)\epsilon_t = V_t
\end{align*} \]

where \(\alpha\) and \(\beta\) are unknown parameters, \(Y_t\) denotes the response, and \(X_t\) the predictor. In practice, \(Y_t\) may denote the stock return, and \(X_t\) may be some financial indicators. \(L^{(c)}\epsilon_t = \epsilon_{t-c}, \quad B(L) = 1 - (\sum_{i=1}^{p} b_i L^i), \quad B(L) \neq 0\) with all of its roots are fixed and less than one in absolute value. \(\{U_t, V_t\}_{t=1}^{T}\) are independent and identically distributed (hereafter i.i.d.) random vectors with zero mean. For the predictor \(X_t\), \(\mu\) denotes the intercept term, and \(\phi\) is the autoregressive coefficient. It is known that the limit distribution of the least squares estimators of \(\alpha\) and \(\beta\) differs from each other when \(\{X_t\}\) follows: (i) \(|\phi| < 1\) (stationary), (ii) \(\mu = 0\) and \(\phi = 1 + c/n\) for some constant \(c\), (iii) \(\mu \neq 0\) and \(\phi = 1 + c/n\) for some constant \(c\). When \(c = 0\), \(\{X_t\}\) is a unit root AR(1) process.

Predictive regression models are widely used in financial research (see [1–8], and references therein). It is noted that, when the predictor \(\{X_t\}\) is non-stationary, the response \(\{Y_t\}\) can not be stationary. This is well known in the literature as the unbalanced problem for traditional predictive regression models. Due to the non-stationarity of most financial variables (see [9]), the problem of unbalance often appears in practical applications. To cover this problem, Ref. [10] proposed a balanced predictive regression (BPR) model by...
adding a difference term of \( X_t \) to the conditional function part. In detail, it takes the form as follows:

\[
\begin{align*}
Y_t &= \alpha + \beta X_{t-1} + \sum_{i=1}^{p} \psi_i \Delta X_{t-i} + U_t \\
X_t &= \mu + \phi X_{t-1} + \epsilon_t, \quad B(L)\epsilon_t = V_t,
\end{align*}
\]

(2)

where \( \Delta X_{t-i} = X_{t-i} - X_{t-i-1} \), and \( \psi_i \)'s are unknown parameters needing to be estimated. Compared to the original predictive regression model (1), the existence of \( \Delta X_{t-i} \)'s in (2) can balance the unmatched problem between \( Y_t \) and \( X_t \) with \( \beta = 0 \) when \( \{ X_t \} \) is non-stationary.

Similar practices are also considered by [11], which proposes a balanced predictive regression model by adding an additional lag of the predictors to the mean function. Through simple transformation, this model is the same as that considered in this paper when \( p = 1 \) by noting that \( \beta X_{t-1} + \psi_1 \Delta X_{t-1} = (\beta + \psi_1) X_{t-1} + (-\psi_1) X_{t-2} \). Both for predictive regression and BPR models, an importation issue of great interest is to test the predictability, i.e., testing the null hypothesis \( H_0 : \beta = 0 \) or \( H_0 : \psi_1 = \cdots = \psi_p = 0 \) is non-stationary. This issue has received extensive attention in the literature. The conventional method for testing predictability is the so-called \( t \)-test. Unfortunately, the standard \( t \)-test suffers from the issue of obvious over-rejection or ineffective when predictors have strong persistence. See [12,13] for more details. Note that [14] investigated the estimation of predictive regressions when the explanatory variable is nearly integrated. However, the limiting distributions of these tests depend on which Case of (i)–(iii) the predictor \( X_t \) follows. Some unified predictability tests have been developed in the last decade to avoid this problem. Among others, Ref. [15] considered the predictability test for model (1) when macroeconomic data are non-stationary or heavy-tailed. For the BPR model, Ref. [10] extended the method of [15] and proposed a unified predictability test regardless of \( X_t \) being stationary, nearly integrated, or unit root.

However, two statistics were developed in [10] to test predictability, which depends on a predefined assumption on whether or not there exists an intercept term \( \alpha \). Although both testing statistics are shown to be chi-squared distributed asymptotically under mild conditions, their finite sample performances are quite different from each other because, when there is an intercept, their second testing statistic depends on the idea of data splitting in order to get rid of the effect of the existence of the intercept term \( \alpha \). Note that both tests are applicable when the underlying true model has no intercept. We found that the \( p \)-values of these two tests are different, which may cause confusion when different \( p \)-values are obtained based on the same data set.

Hence, it is necessary to do the predefined test to check whether or not there exists a non-zero intercept in model (2) before conducting a predictability test. This motivates us to consider the current research. We propose an empirical likelihood test based on the idea of data splitting and proved that the asymptotic distribution of the test statistic is \( \chi^2 \) distribution, regardless of \( X_t \) being stationary or nearly integrated or unit root. The empirical likelihood method was proposed by [16] and proved to have many excellent properties [17], including the distribution of the data that does not need to be assumed. Therefore, this method has been widely studied in the literature [18–20]. This also prompted this paper to propose a unified test of interception terms using the empirical likelihood method.

The rest of this article is organized as follows: Section 2 presents the methodologies and the main asymptotic results of our proposed test. Section 3 contains the finite sample simulation studies. An empirical application is discussed in Section 4. The detailed proofs of the main results are presented in Appendix A.
2. Methodology and Main Results

Suppose the random observations \( \{ X_i, Y_i \}_{i=1}^n \) are generated from the model (2). The main interest of this paper is to develop a unified test for checking the intercept for model (2), i.e., the following hypothesis

\[
H_0 : \alpha = a_0, \quad \text{versus} \quad H_1 : \alpha \neq a_0.
\]

Hereafter, denote \( \theta_0 := (a_0, \beta_0, \psi_1, \ldots, \psi_p) \) as the true value of \( \theta := (\alpha, \beta, \psi_1, \ldots, \psi_p) \) without confusion.

Since the empirical likelihood method enjoys many desirable properties as discussed in the literature, in the sequel, we are interested in developing a testing method for \( H_0 \) based on this method. Note that, when \( \theta = \theta_0 \), we have

\[
E \left( \left( Y_i - \alpha - \beta X_i - 1 - \sum_{i=1}^{p} \psi_i \Delta X_{i-1} \right) \tilde{\xi}_i \right) = 0,
\]

where \( \tilde{\xi}_i := (1, X_{i-1}, \Delta X_{i-1}, \ldots, \Delta X_{i-p})^\top \). Then, following [16], one may construct an empirical likelihood function by using the auxiliary vectors \( \{ W_i \} \), where

\[
W_i(\theta) := \left( Y_i - \alpha - \beta X_i - 1 - \sum_{i=1}^{p} \psi_i \Delta X_{i-1} \right) \tilde{\xi}_i.
\]

Unfortunately, this test will suffer from the trouble of the quantity \( \frac{1}{\sqrt{n}} \sum_{i=p+1}^{n} \left( D^{-1}_n W_i(\theta) \right)^\top D^{-1}_n \) not converging in probability as \( n \to \infty \) for Case (i), i.e., with \( \mu = 0 \) and \( \phi = 1 + c/n \) for some constant \( c \), where \( D_n = \text{diag}\{1, \sqrt{n}, 1, \ldots, 1\} \). Consequently, the quantity

\[
\left( \frac{1}{\sqrt{n}} \sum_{i=p+1}^{n} D^{-1}_n W_i(\theta) \right)^\top \left( \frac{1}{\sqrt{n}} \sum_{i=p+1}^{n} D^{-1}_n W_i(\theta) \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=p+1}^{n} D^{-1}_n W_i(\theta) \right)
\]

does not converge in distribution to a chi-squared distributed variable, which in turn results in the related log-empirical likelihood function having a non-standard limit distribution.

As an improvement, although it is possible to construct a new empirical likelihood-based test statistic, which converges in a chi-squared distribution, for \( \theta \) relying on the idea of data splitting as did in [21], the main interest of this paper is to test the intercept term, i.e., hypothesis \( H_0 \), and we need to handle the redundant parameters \( \beta, \psi_1, \ldots, \psi_p \) by using the profile empirical likelihood method as in [17]. Once again, this profile method still does not work; see Theorem 3 in [21] for a detailed discussion.

This motivates us to consider the following testing procedure. Let \( \psi := (\psi_1, \ldots, \psi_p)^\top \), \( Z_i(\alpha, \beta, \psi) = (Z_{i,1}(\alpha, \beta, \psi), Z_{i,2}(\alpha, \beta, \psi), \ldots, Z_{i,2+p}(\alpha, \beta, \psi)) \) for \( t = p+1, 2, \ldots, m \), where \( m = \lceil n/2 \rceil \) with \( \lceil \cdot \rceil \) being the floor function, and

\[
\begin{align*}
Z_{i,1}(\alpha, \beta, \psi) &= Y_i - \alpha - \beta X_{i-1} - \sum_{i=1}^{p} \psi_i \Delta X_{i-1} \\
Z_{i,2}(\alpha, \beta, \psi) &= (Y_{i+m} - \alpha - \beta X_{i+m-1} - \sum_{i=1}^{p} \psi_i \Delta X_{i+m-1}) w_{i+m} + \delta_i \\
Z_{i,2+k}(\alpha, \beta, \psi) &= (Y_i - \alpha - \beta X_{i-1} - \sum_{i=1}^{p} \psi_i \Delta X_{i-1}) \Delta X_{i-k}, \quad k = 1, 2, \ldots, p,
\end{align*}
\]

with \( w_i = \frac{X_{i-1}}{\sqrt{1 + X_{i-1}^2 \log(e + |X_{i-1}|)}} \), and \( \delta_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \delta_{ij} \) with \( N = 2000 \) and \( \delta_{ij} \sim N(0, \sigma^2) \), which are independent of the random observations. Through data splitting technology, we use the first half of the data to estimate redundant parameters, and the second half of the data to construct a likelihood function to reduce the challenges posed by noise in the data. To balance the size and the power, in practice, one may take \( \tilde{\sigma}^2 = \tilde{\sigma}_U^2 / 2 \) with \( \tilde{\sigma}_U^2 \) being the estimated error variance. Note that here \( \log(e + |X_{i-1}|) \) is designed in order to
let the quantity \( \frac{1}{m} \sum_{t=p+1}^{m} \{ (Y_{t+m} - \alpha - \beta X_{t+m-1} - \sum_{i=1}^{p} \psi_i \Delta X_{t+m-i}) w_{t+m} \} \) to vanish when \( \{ X_t \} \) follows the non-stationary Cases (ii). A similar weighting technique has also been used in [22], which uses the profile empirical likelihood to consider the intercept test of the predictive regression model.

Based on \( \{ Z_t(\alpha, \beta, \psi) \} \) above, a new empirical likelihood function for \( \theta \) can be defined as follows:

\[
L(\alpha, \beta, \psi) = \sup \left\{ \prod_{t=p+1}^{m} (p_t) : p_1 \geq 0, \cdots, p_m \geq 0, \sum_{t=p+1}^{m} p_t = 1, \sum_{t=p+1}^{m} p_t Z_t(\alpha, \beta, \psi) = 0 \right\}.
\]

Since we are only interested in \( \alpha \), we further define a profile empirical likelihood function for \( \alpha \) as

\[
\ell(\alpha) = \sup_{(\beta, \psi)} L(\alpha, \beta, \psi).
\]

The following theorem shows that Wilks’ theorem holds for the above proposed profile empirical likelihood method, which depends on some regular conditions as follows:

(C1) \( \{ (U_t, e_t) \} \) is a sequence of i.i.d. random vectors with mean zero, and \( E(|U_t|^{2+\epsilon_0}) + E(|e_t|^{2+\epsilon_0}) < \infty \) for some arbitrarily small positive constant \( \epsilon_0 \);

(C2) The predictive variable \( X_t \), with the initial value of \( X_0 \) being constant or random variable of order \( o_p(\sqrt{m}) \), belongs to one of the persistence classes following Cases (i)–(iii);

(C3) All roots of \( 1 - \sum_{i=1}^{p} \Psi_{i,0} x^i = 0 \) with respect to \( x \) are outside the unit circle.

**Theorem 1.** Suppose that Conditions C1–C3 hold. Then, under the null hypothesis \( \mathcal{H}_0 : \alpha = a_0 \), we have that \( -2 \log \ell(a_0) \) converges in distribution to a chi-squared distributed variable \( \chi^2_1 \) with one degree of freedom as \( n \to \infty \).

Based on the above theorem, we can reject \( \mathcal{H}_0 \) once

\[
-2 \log \ell(a_0) \geq \chi^2_1 (1 - \tau),
\]

at the significance level \( \tau \in (0, 1) \), where \( \chi^2_1 (1 - \tau) \) denotes the \( (1 - \tau) \)-th quantile of a chi-squared distribution with one degree of freedom. In particular, when one is interested in testing \( \mathcal{H}_0 : \alpha = 0 \), one may think that there is no intercept in model (2) if \( -2 \log \ell(0) \geq \chi^2_1 (1 - \tau) \) at the significance level \( \tau \).

Note that, similar to [23], one may take \( w_t = \frac{X_{t-1}}{(1 + X_{t-1}^2)^{\gamma_1}} \) to construct the test. Here, we however use another weight \( w_t = \frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2 \log(e + |X_{t-1}|)}} \) in \( Z_{t,2}(\alpha, \beta, \psi) \) in order to increase the local power of the proposed test for non-stationary cases. The following theorem states the power property of our proposed empirical likelihood-based test.

**Theorem 2.** Suppose the same conditions of Theorem 1 hold. Then, we have as \( n \to \infty \) that:

- For Case (i), under the local alternative hypothesis \( \mathcal{H}_1 : \alpha = a_0 - \frac{d_0}{\sqrt{m}} \) for some constant \( d_0 \in R \),

\[
-2 \log \ell(a_0) \xrightarrow{d} \chi^2_1 (v_1^2),
\]

where ‘\( \xrightarrow{d} \)’ denotes the convergence in distribution, and \( \chi^2_1 (v_1^2) \) a non-central chi-squared distributed variable with non-central parameter \( v_1^2 \), where \( v_1 \) is the second component of \( \Sigma^{-1/2} \gamma_1 \) with \( \gamma_1 = d_0 (1, \lim_{t \to \infty} E(w_{t+m}), \mu + (\phi - 1) \lim_{t \to \infty} E(X_{t-2}), \cdots, \mu + (\phi - 1) \lim_{t \to \infty} E(X_{t-p-1}))^T \), and

\[
\Sigma = \lim_{t \to \infty} E(Z_t(a_0, \beta_0, \psi_0)Z_t(a_0, \beta_0, \psi_0)^T);\]
• For Case (ii), under the local alternative hypothesis $H_{1}: \alpha = \alpha_0 - \frac{d_0 \log(m)}{\sqrt{m}}$ for some constant $d_0$,

$$-2 \log \ell(a_0) \rightarrow^{d} \left( \xi + \frac{2d_0}{\alpha} \int\frac{\text{sgn}(J_c(s))ds}{\alpha} \right)^2,$$

where $\text{sgn}(\cdot)$ denotes the sign function, $\xi$ a standard normally distributed variable, which is independent of $\int_{1}^{n} \text{sgn}(J_c(s))ds$, and $J_c(s) := \int_{s}^{n} e^{-(s-r)c}dW(r)$ with $W(r)$ being a Gaussian process with covariance function $(s,t) = 2E(c^2) \min(s,t)$;

• For Case (iii), under the local alternative hypothesis $H_{1}: \alpha = \alpha_0 - \frac{d_0 \log(m)}{\sqrt{m}}$ for some constant $d_0$,

$$-2 \log \ell(a_0) \rightarrow^{d} \chi^2_1(v_2),$$

where $v_2 = d_2 \int_{1}^{n} \text{sgn}(\mu \int_{0}^{n} e^{-(s-r)c}dr)ds / \sigma$.

Theorem 2 indicates that the local power of the proposed test for Cases (ii) and (iii) is of order $O\left( \frac{\log(m)}{\sqrt{m}} \right)$, which does not achieve the normally parametric order $O(1/\sqrt{m})$ but is better than $O(n^{-a})$ for any $a \in (0, 1/2)$.

**Remark 1.** Although we only consider in Theorems 1 and 2 the case that $\{U_t\}$ is a sequence of i.i.d. errors, it is possible to obtain similar results when $U_t$ follows a dependent process. In fact, when $U_t$ follows the AR($q$) process, one can obtain similar limits as in Theorems 1 and 2 by taking into account the AR structure of $U_t$.

**Remark 2.** When $e_t$ follows a strictly stationary GARCH process or a linear process, Theorems 1 and 2 still hold.

### 3. Simulation Results

In this section, we investigate the finite sample performance of the proposed profile empirical likelihood test using simulated data sets and compare it with the traditional $t$-test. The random observations $\{X_t, Y_t\}$ are generated from the predictive regression model (2), where $\alpha = 0$, $\beta = \{\beta_0, \beta_1\} = \{0.5, 0.3\}$. $\mu \in \{0, 0.01\}$ indicates that the model has no intercept item and intercept item, respectively. $\phi \in \{0.5, 1, \frac{3}{2}\}$, where $0.5$ indicates that $X_t$ is a stationary process, $1$ indicates a unit root process, and $1 - \frac{3}{2}$ indicates a near unit root process. The innovations $\{U_t, V_t\}_{t=1}^{n}$ were considered in two cases.

**Case (i):** $\{U_t, V_t\}_{t=1}^{n}$ are from a bivariate Gaussian Copula $C(F_{1}(U_t), F_{2}(V_t), \rho)$, whose marginal distribution $F_{i}$ is a Student’s $t$-distribution with degrees of freedom $df \in \{10, 15\}$. The dependence parameter $\rho$ is set to be $\{-0.2, -0.4, -0.6\}$, respectively, to explore the performance of two innovations under different degrees of dependence.

**Case (ii):** $\{V_t\}_{t=1}^{n}$ are from GARCH(1,1) structure, i.e.,

$$\begin{align*}
V_t &= \eta_v \sigma_{t-1}, \\
\sigma_{t-1}^2 &= \omega_v + a_v V_{t-1}^2 + b_v \sigma_{t-1}^2,
\end{align*}$$

where $\{\omega_v, a_v, b_v\} = \{0.1, 0.4, 0.2\}$. $\{U_t\}_{t=1}^{n}$ and $\{\eta_{V,t}\}_{t=1}^{n}$ are generated similarly to $\{U_t, V_t\}_{t=1}^{n}$ in Case (i). All simulating results are repeatedly carried 10,000 times with the sample size $n$ ranging from 200 to 1200.

Table 1 reports the size performance of the proposed method with different settings in Case (i) under the significance level $\tau = 0.05$. For $df = \{10, 15\}$, the results show that, when $X_t$ is a stationary or unit root process, regardless of whether $\mu$ is 0 or not, with the increase of sample $n$, size values are closer to $\tau$, and the performance of the method is robust under different $\rho$. However, when $X_t$ is a near unit root process, it can be found that, when $n$ is small, the method has the performance of oversize, especially when $\rho = -0.6,
and as \( n \) increases, it gradually converges to the significance level. For the \( t \)-test when \( X_t \) is a stable process, the \( t \)-test has a good size performance, and its performance is relatively stable even with high dependency. However, when \( X_t \) is a unit root or near unit root process, it gradually shows over-rejecting with the increase of \( \rho \), no matter whether there is an intercept or not. The \( df = (15,15) \) behaves similarly to \( df = (10,10) \). Table 2 reports the size performance in Case (ii), in which the innovation of \( X_t \) has a GARCH structure. The results show that the performances of the profile empirical likelihood and \( t \)-test are similar to that in Case (i).

Table 1. The size performance of \( \{U_t, V_t\}_{t=1}^n \) under Case (i) and \( \tau = 0.05 \).

| \( \phi \) | \( \mu \) | \( \rho \) | 200 | 400 | 800 | 1000 | 1200 | \( EL \) | 200 | 400 | 800 | 1000 | 1200 | \( T \)-Test |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 0.01 | 0.0526 | 0.0511 | 0.0483 | 0.0503 | 0.0475 | 0.0586 | 0.0615 | 0.0559 | 0.0575 | 0.0574 |
| | 0.0501 | 0.0508 | 0.0502 | 0.0469 | 0.0457 | 0.0827 | 0.0836 | 0.0814 | 0.0881 | 0.0917 |
| | 0.0508 | 0.0508 | 0.0506 | 0.0472 | 0.0519 | 0.0552 | 0.0587 | 0.0574 | 0.0607 | 0.0594 |
| | 0.0506 | 0.0506 | 0.0473 | 0.0486 | 0.0498 | 0.0827 | 0.0805 | 0.0843 | 0.0867 | 0.0825 |
| | 0.0488 | 0.0493 | 0.0493 | 0.0432 | 0.0443 | 0.1277 | 0.1208 | 0.1184 | 0.1292 | 0.1229 |
| 0.6 | 0.0484 | 0.0468 | 0.0465 | 0.0502 | 0.0512 | 0.0533 | 0.0513 | 0.0544 | 0.0550 | 0.0543 |
| | 0.0475 | 0.0507 | 0.0453 | 0.0467 | 0.0440 | 0.0683 | 0.0707 | 0.0679 | 0.0634 | 0.0703 |
| | 0.0463 | 0.0389 | 0.0366 | 0.0422 | 0.0403 | 0.0911 | 0.0893 | 0.0845 | 0.0876 | 0.0846 |
| | 0.0507 | 0.0481 | 0.0489 | 0.0496 | 0.0487 | 0.0513 | 0.0578 | 0.0526 | 0.0522 | 0.0505 |
| | 0.0467 | 0.0501 | 0.0458 | 0.0454 | 0.0470 | 0.0695 | 0.0682 | 0.0677 | 0.0690 | 0.0658 |
| | 0.0386 | 0.0409 | 0.0403 | 0.0397 | 0.0415 | 0.0869 | 0.0898 | 0.0898 | 0.0835 | 0.0868 |
| 1–0.5 | 0.0421 | 0.0497 | 0.0495 | 0.0486 | 0.0486 | 0.0471 | 0.0495 | 0.0495 | 0.0529 | 0.0495 |
| 0.5 | 0.0427 | 0.0455 | 0.0463 | 0.0455 | 0.0499 | 0.0463 | 0.0528 | 0.0490 | 0.0516 | 0.0505 |
| | 0.0361 | 0.0439 | 0.0429 | 0.0519 | 0.0486 | 0.0549 | 0.0480 | 0.0531 | 0.0477 | 0.0527 |
| | 0.0446 | 0.0498 | 0.0453 | 0.0497 | 0.0496 | 0.0499 | 0.0541 | 0.0523 | 0.0514 | 0.0517 |
| | 0.0435 | 0.0462 | 0.0460 | 0.0455 | 0.0471 | 0.0512 | 0.0457 | 0.0499 | 0.0528 | 0.0504 |
| | 0.0417 | 0.0444 | 0.0460 | 0.0451 | 0.0468 | 0.0477 | 0.0516 | 0.0513 | 0.0508 | 0.0528 |
| 0.6 | 0.0494 | 0.0511 | 0.0519 | 0.0497 | 0.0448 | 0.0566 | 0.0617 | 0.0567 | 0.0584 | 0.0572 |
| | 0.0542 | 0.0496 | 0.0505 | 0.0527 | 0.0497 | 0.0839 | 0.0881 | 0.0914 | 0.0862 | 0.0798 |
| | 0.0506 | 0.0492 | 0.0454 | 0.0464 | 0.0432 | 0.1226 | 0.1266 | 0.1282 | 0.1244 | 0.1248 |
| | 0.0518 | 0.0469 | 0.0514 | 0.0533 | 0.0498 | 0.0632 | 0.0568 | 0.0607 | 0.0587 | 0.0588 |
| | 0.0533 | 0.0481 | 0.0486 | 0.0486 | 0.0472 | 0.0802 | 0.0818 | 0.0829 | 0.0823 | 0.0821 |
| | 0.0483 | 0.0463 | 0.0468 | 0.0468 | 0.0435 | 0.1284 | 0.1275 | 0.1224 | 0.1317 | 0.1304 |
| 1–0.6 | 0.0511 | 0.0468 | 0.0502 | 0.0511 | 0.0508 | 0.0502 | 0.0546 | 0.0534 | 0.0525 | 0.0514 |
| 0.6 | 0.0476 | 0.0476 | 0.0547 | 0.0449 | 0.0619 | 0.0657 | 0.0635 | 0.0670 | 0.0685 | 0.0677 |
| | 0.0436 | 0.0412 | 0.0435 | 0.0422 | 0.0408 | 0.0908 | 0.0831 | 0.0890 | 0.0866 | 0.0886 |
| | 0.0539 | 0.0500 | 0.0481 | 0.0472 | 0.0492 | 0.0559 | 0.0545 | 0.0561 | 0.0506 | 0.0565 |
| | 0.0467 | 0.0474 | 0.0427 | 0.0476 | 0.0428 | 0.0678 | 0.0681 | 0.0659 | 0.0641 | 0.0702 |
| | 0.0428 | 0.0445 | 0.0438 | 0.0410 | 0.0417 | 0.0855 | 0.0844 | 0.0864 | 0.0857 | 0.0840 |
Table 2. The size performance of $\{U_i, V_i\}_{i=1}^n$ under Case (ii) and $\tau = 0.05$.

| $\phi$ | $\mu$ | $\rho$ | 200 | 400 | 800 | 1000 | 1200 | 200 | 400 | 800 | 1000 | 1200 | $t$-Test |
|-------|-------|-------|-----|-----|-----|------|------|-----|-----|-----|------|------|-------|
| 0     | 0.0533 | 0.437 | 0.0504 | 0.0507 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 |
| 1     | 0.0533 | 0.437 | 0.0504 | 0.0507 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 |
| 0.01  | 0.0523 | 0.437 | 0.0504 | 0.0507 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 |
| 0.01  | 0.0523 | 0.437 | 0.0504 | 0.0507 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 |
| 0.01  | 0.0523 | 0.437 | 0.0504 | 0.0507 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 |
| 0.01  | 0.0523 | 0.437 | 0.0504 | 0.0507 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 |
| 0.01  | 0.0523 | 0.437 | 0.0504 | 0.0507 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 |
| 0.01  | 0.0523 | 0.437 | 0.0504 | 0.0507 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 |
| 0.01  | 0.0523 | 0.437 | 0.0504 | 0.0507 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 | 0.0509 |

We also generate data with $\alpha = \frac{d_0}{\sqrt{n}}$ (stationary) or $\alpha = \frac{d_0 \log(n)}{\sqrt{n}}$ (unit root or near the unit root), $d_0 = \{0, 1, 2, \cdots, 7\}$, and other parameters are set the same as Case (i) and Case (ii), in order to investigate the power performance of the proposed method. The results are shown in Figure 1 ($\mu = 0$) and Figure 2 ($\mu = 0.01$). Here, we only show the case where $n = 1000$. The abscissa represents the constant $d_0$ value, and the ordinate is the power value. Different colors represent different processes of $X_t$ and tests. We can see that, under all settings, Figures 1 and 2 show the same performance. Firstly, as the value of $d_0$ increases, the power converges quickly to 1. Secondly, when $X_t$ is a near unit root, the convergence rate is better than the unit root, and the unit root is better than the stationary. Last but not least, when $d_0 = 0$, it can be seen that the value of the $t$-test is significantly higher than the significance level of 0.05 under all settings. This also indicates the poor size performance of the $t$-test, even though it has a good convergence rate in power performance.

In summary, the proposed method has good size and power performance when checking whether intercept $a$ exists and also has good robustness for different innovative distributions. The traditional $t$-test shows over-reject, which makes it easy to misjudge
real data analysis; therefore, the profile empirical likelihood method proposed in this paper is more recommended when real data are endogenous.

Figure 1. The power performance with $\mu = 0$. 
4. A Real Data Application

In this section, we illustrate our testing method on U.S. equity data, which has been explored by numerous scholars, such as [10,12,24], etc. We collected the data coming from the Center for Research in Security Prices (CRSP), and the sample period is January 1952 to December 2015 with monthly data (n = 1068). Among them, the value-weighted excess returns of the S&P 500 index are used as the predicted variable $Y_t$, and the other 10 financial
variables are dividend–price ratio, dividend yield, earnings–price ratio, dividend payout ratio, book-to-market value ratio, T-bill rate, default yield spread, long-term yield, term spread, and net equity expansion, which are used as predicting variables, respectively.

We know that predictability is one of the critical concerns in various fields, including finance, climatology, tourism, sociology, and so on (see, e.g., [25–27]). However, as [10] shows, a small intercept can change the predictability dramatically when the predicting variable is nearly integrated. Hence, the intercept term test is an important pre-test before the statistical inference. On the other hand, Refs. [12,28] showed that the innovation of financial data generally contains endogeneity, which will bring bias to relevant estimates. Therefore, it is undoubtedly very vital to obtain a unified test for the intercept test, no matter whether the predictive variables are stationary, nearly integrated, unit root, or innovation have endogeneity.

Firstly, we use the adf.test function in the R package to test the stationarity of each predictor variable. Then, the least square estimation was used to fit the model (2), in which \( V_t \) is estimated based on the \( V_t = e_t - \sum_{i=1}^{12} b_i e_{t-i} \), in order to make \( V_t \) uncorrelated. In addition, the correlation coefficient between \( U_t \) and \( V_t \) was tested. Before performing our test, the ArchTest function was also used to test whether \( V_t \) has a GARCH structure. Finally, the method proposed in this paper is used to test \( H_0: \alpha = 0 \). Table 3 reports all the results for each predictor variable. The first column is the predictive variables. The second column is the estimators of intercept terms, and the third column is the estimators of the autoregressive coefficient of predictive variables. The fourth column is the \( p \)-values of the ADF test. The fifth column is the correlation coefficient between the two innovations \( U_t \) and \( V_t \). The sixth column is the \( p \)-values of the GARCH test of \( V_t \) sequence. The last two columns are the \( p \)-values of the test proposed in this paper and the traditional \( t \)-test.

| Predictor                       | \( \hat{\alpha} \)  | \( \hat{\phi} \)  | Adf.Test     | Cor(\( U_t, V_t \)) | GARCH-V          | EL          | \( t \)-Test  |
|---------------------------------|----------------------|------------------|--------------|----------------------|-----------------|-------------|--------------|
| Dividend–price ratio            | 0.0319               | 0.9928           | 0.0157 **    | -0.9618              | 9.6402 \( \times 10^{-37} \) *** | 0.4200      | 0.0102 **    |
| Dividend yield                  | 0.0357               | 0.9929           | 0.0239 **    | -0.0791              | 5.4329 \( \times 10^{-34} \) *** | 0.4626      | 0.0042 ***   |
| Earnings–price ratio            | 0.0308               | 0.9870           | 0.0112 **    | -0.7966              | 7.6498 \( \times 10^{-35} \) *** | 0.2730      | 0.0058 ***   |
| Dividend payout ratio           | 0.0072               | 0.9913           | 0.0100 ***   | -0.0407              | 1.7194 \( \times 10^{-41} \) *** | 0.0828*     | 0.0487 **    |
| Book-to-market value ratio      | -0.0050              | 0.9858           | 0.0100 ***   | -0.8031              | 1.1404 \( \times 10^{-84} \) *** | 0.1322      | 0.2146       |
| T-bill rate                     | 0.0097               | 0.9934           | 0.5266       | -0.0762              | 8.1443 \( \times 10^{-68} \) *** | 0.2945      | 0.0001 ***   |
| Default yield spread            | 0.0013               | 0.9752           | 0.0100 ***   | -0.2433              | 3.4015 \( \times 10^{-72} \) *** | 0.2027      | 0.6769       |
| Long-term yield                 | 0.0105               | 0.9965           | 0.9568       | -0.1098              | 2.0058 \( \times 10^{-44} \) *** | 0.2722      | 0.0033 **    |
| Term spread                     | 0.0033               | 0.9608           | 0.0100 ***   | -0.0070              | 8.4860 \( \times 10^{-43} \) *** | 0.2126      | 0.2350       |
| Net equity expansion            | 0.0090               | 0.9805           | 0.0100 ***   | -0.0619              | 3.1793 \( \times 10^{-39} \) *** | 0.7741      | 1.43 \( \times 10^{-5} \) *** |

Significance levels: * \( p \leq 0.1 \), ** \( p \leq 0.05 \), *** \( p \leq 0.01 \).

From Table 3, we can see that the \( \hat{\phi} \) of all variables are very close to 1, and the adf.test shows that only T-bill rate and long-term yield do not reject the null hypothesis of unit root, indicating that these variables are non-stationary. The results of the correlation coefficient show that \( U_t \) and \( V_t \) have different degrees of correlation under different predictive variables, and we have also carried out corresponding settings in the simulation. In addition, the results of the ArchTest test show that \( V_t \) under all predictor variables has a significant ARCH effect. The \( \hat{\alpha} \) of ten variables obtained by the least square estimation are close to 0. The \( t \)-test showed that the intercept of Book-to-market, Default yield spread, and Term spread are 0. However, the results of the EL test are quite different from \( t \)-test, which shows that only the intercept of the Dividend payout ratio is not 0.

Overall, the \( t \)-test rejected the original hypothesis of alpha = 0 under most variables, which is consistent with the performance in the simulation. Notice that Ref. [12] mentioned that the \( t \)-test does not perform well when the predictor variable is persistent and its innovations are highly correlated. Furthermore, all the variables have a GARCH structure,
which also will affect the t-test. In addition, Table 1 in [29] indicates the size distortion of the t-test for plausible parameter values. Hence, we recommend the EL test proposed in this paper over the traditional test.

5. Conclusions

Predictability testing is a common problem in the analysis of real data such as in the economic and financial fields. However, the validity of most existing tests depends on the assumption that the intercept term of the model exists. Therefore, the intercept term test is an important pre-test in practical application. However, it is worth noting that heteroscedasticity and endogeneity are common characteristics of financial data, which makes the existing intercept term test not perform well. In this paper, we develop a unified intercept test based on the empirical likelihood method under the balanced prediction regression model. The asymptotic distribution of the EL test is proved to be chi-squared, and the local power under stationary and non-stationary conditions is also proved. Simulation results verify the unified and effectiveness of the test and are superior to the t-test.

Furthermore, in the empirical application, we performed an intercept term test on U.S. equity data. The results showed that the t-test rejected the null hypothesis in most variables, but the EL test shows the opposite. The relevant performance is consistent with the simulation. Therefore, when considering the intercept term test of financial data, the EL test is recommended.

Author Contributions: Conceptualization, X.L.; data curation, Y.F.; methodology, X.L.; investigation, Y.F.; writing—original draft preparation, Y.F., Q.W.; writing—review and editing, X.L., Y.F., Q.W., L.P.; visualization, Q.W.; supervision, L.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NNSF of China grant numbers (11971208, 11601197), NSSF of China grant number (21BTJ035), National Major Social Science Project of China grant number (21&ZD152), the 17th student research project of Jiangxi University of Finance and Economics grant number (20220929165358818).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data used in applications can be found in the Center for Research in Security Prices.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Detailed Proofs of the Main Results

In this appendix, we provide detailed proofs for the main results presented in Section 2. Before proceeding further, we specify some useful lemmas as follows:

Lemma A1. Under the same conditions of Theorem 1, we have as $n \to \infty$ that:

(a) $\max_{p+1 \leq i \leq m} \sup_{(\beta, \psi) \in B} \|Z_i(\alpha_0, \beta, \psi)\| = o_p(\sqrt{m})$,

(b) $\frac{1}{\sqrt{m}} \sum_{i=p+1}^{m} Z_i(\alpha_0, \beta, \psi) = \frac{1}{\sqrt{m}} \sum_{i=p+1}^{m} Z_i(\alpha_0, \beta_0, \psi_0) + O_p(1)$ holds uniformly for $(\beta, \psi) \in B$,

(c) $\frac{1}{m} \sum_{i=p+1}^{m} \{Z_i(\alpha_0, \beta, \psi)Z_i^\top(\alpha_0, \beta, \psi)\} = \frac{1}{m} \sum_{i=p+1}^{m} \{Z_i(\alpha_0, \beta_0, \psi_0)Z_i^\top(\alpha_0, \beta_0, \psi_0)\} + o_p(1)$ holds uniformly for $(\beta, \psi) \in B$,

where $B = \{(\hat{\beta}, \hat{\psi}) : m^{(k-1)/2} |\hat{\beta} - \beta_0| + |\hat{\psi} - \psi_0| \leq \frac{C}{\sqrt{m}} \}$ for some constant $C > 0$, and $k = 1, 2, 3$ correspond to Cases (i)-(iii), respectively.
Proof of Lemma A1. In the sequel, we only prove Case (ii), as the rest of the cases can be proved similarly. We first prove Part (a). Note that, by [30], we have for any $s \in (0, 2]$ that
\[
\frac{1}{\sqrt{m}} X_{[m]} \overset{D}{\to} J_c(s) := \int_0^s e^{-(s-r)}c dW(r),
\]
for Case (ii) as $n \to \infty$, where \( \overset{D}{\to} \) stands for the convergence in space, and $D(0, 1]$ is the space of real-valued functions on the interval $(0, 1]$ that are right continuous and have finite left limits. Since
\[
\max_{p+1 \leq l \leq m} \|Z_l(a_0, \beta, \psi)\| \leq \sum_{k=1}^{p+2} \max_{p+1 \leq l \leq m} |Z_{l,k}(a_0, \beta, \psi)| \text{ for } k \in \{0, 1\},
\]
we need only to check the upper bound for each component $Z_{l,k}(a_0, \beta, \psi)$. 

For the first component, rewrite $Z_{1,1}(a_0, \beta, \psi) = \epsilon_{1} - \sqrt{m} (\beta - \beta_0 ) \frac{1}{\sqrt{m}} X_{t-1} - \sum_{i=1}^{\beta} (\psi_i - \psi_{i,0}) \Delta X_{t-i-1}$. 

We have
\[
\max_{p+1 \leq l \leq m} |\epsilon_{l}| = O(\sqrt{m}).
\]
Trivially, based on Conditions C1–C2, we have
\[
\max_{p+1 \leq l \leq m} |\epsilon_{l}| = O(\sqrt{m}).
\]
Next, using (A1), it is easy to check for any $\epsilon > 0$ that
\[
P \left( \max_{p+1 \leq l \leq m} \left| \frac{1}{\sqrt{m}} X_{t-1} \right| \geq \epsilon \sqrt{m} \right) \leq \frac{2}{\epsilon^2 + \beta m} \frac{1}{2m} \left( \sum_{l=1}^{\beta} \right)^{2+\delta} \rightarrow 0,
\]
which implies
\[
\max_{1 \leq l \leq 2m} \left| \frac{1}{\sqrt{m}} X_{t-1} \right| = O(\sqrt{m}), \text{ and in turn } \max_{1 \leq l \leq 2m} \left| (\beta - \beta_0 ) X_{t-1} \right| = O(1) \text{ uniformly for } (\frac{\beta}{m}) \in B.
\]
Using this, by $\Delta X_{t} = X_{t} - X_{t-1} = \frac{1}{m} X_{t-1} + \epsilon_{t}$, we have as $n \to \infty$ that
\[
\max_{1 \leq l \leq 2m} \left| \Delta X_{l} \right| \leq \frac{c \sqrt{m}}{n} \max_{1 \leq l \leq 2m} \left| \frac{1}{\sqrt{m}} X_{l} \right| + \max_{1 \leq l \leq 2m} |\epsilon_{l}| = \max_{1 \leq l \leq 2m} |\epsilon_{l}| \cdot \{1 + o_p(1)\},
\]
and hence
\[
\max_{p+1 \leq l \leq 2m} \left| \sum_{i=1}^{p} (\psi_i - \psi_{i,0}) \Delta X_{t-i-1} \right| = o_p(1),
\]
uniformly for $(\frac{\beta}{m}) \in B$. This shows
\[
\max_{p+1 \leq l \leq m} \sup_{B} |Z_{l,1}(a_0, \beta, \psi)| \leq \max_{1 \leq l \leq m} |\epsilon_{l}| \cdot \{1 + o_p(1)\}.
\]
Similarly, we can show that
\[
\max_{p+1 \leq l \leq m} \sup_{B} |Z_{l,1}(a_0, \beta, \psi)| \leq \max_{1 \leq l \leq m} |\epsilon_{l}| \cdot \{1 + o_p(1)\},
\]
for $l = 2, \ldots, p + 2$. This shows Part (a).

For Part (b), note that
\[
Z_{l}(a_0, \beta, \psi) = Z_{l}(a_0, \beta_0, \psi_0) - (\beta - \beta_0) \bar{X}_{t-1} - \sum_{i=1}^{p} (\psi_i - \psi_{i,0}) \bar{X}_{t-i}^*, \quad (A2)
\]
where
\[
\bar{X}_{t-1} = (X_{t-1}, X_{t+m-1}, \ldots, X_{t-1}, \bar{X}_{t-1})^T, \text{ and } \bar{X}_{t-i}^* = (\Delta X_{t-j-i}, \Delta X_{t+j-m-i}, \Delta X_{t-i}^*, \ldots, \Delta X_{t-i} \bar{X}_{t-p})^T.
\]
Then, by using a similar proof to that of Part (a), we have
\[
m(\beta - \beta_0) \cdot \frac{1}{m} \sum_{l=1}^{\beta} \bar{X}_{l-1} = O_p(1), \text{ and }
\]
\[
\sum_{i=1}^{\beta} \sqrt{m}(\psi_i - \psi_{i,0}) \cdot \frac{1}{m} \sum_{l=1}^{\beta} \bar{X}_{l-i}^* = O_p(1),
\]
uniformly for \((\hat{\beta}_t) \in \mathcal{B}\) as \(n \to \infty\).

The proof of Part (c) is similar to that of Part (b) based on (A2). We omit the details. This completes the proof of Lemma A1. □

Lemma A2. Under the same conditions of Theorem 1, we have as \(n \to \infty\) that:

(a) \[ \frac{1}{\sqrt{m}} \sum_{t=p+1}^m Z_t(\alpha_0, \beta_0, \psi_0) \xrightarrow{d} N(0, \Sigma), \]

(b) \[ \frac{1}{\sqrt{m}} \sum_{t=p+1}^m \{ Z_t(\alpha_0, \beta_0, \psi_0)Z_t^\top(\alpha_0, \beta_0, \psi_0) \} \xrightarrow{p} \Sigma, \]

where ‘\( \xrightarrow{p} \)’ denotes the convergence in probability. For Case (i), \( \Sigma \) is specified in Theorem 2, and Cases (ii) and (iii),

\[ \Sigma = \left( \begin{array}{c} E(U_1^2) \\ \sigma^2 \\ \Gamma \end{array} \right) \quad \text{with} \quad \Gamma = E(U_1^2) \left( \begin{array}{ccc} \mu^2 + \lim_{t \to \infty} E(e_{t-1}^2) & \cdots & \mu^2 + \lim_{t \to \infty} E(e_{t-p}^2) \end{array} \right). \]

Note that \( \mu = 0 \) for Case (ii).

Proof of Lemma A2. We only prove Part (a), as Part (b) follows a similar fashion. Put

\[ Z_t(\alpha_0, \beta_0, \psi_0) = (U_t, U_tw_t + \delta_t, U_t\Delta X_{t-1}, \cdots, U_t\Delta X_{t-p})^\top. \]

Denote \( \mathcal{F}_t \) as the sigma field generated by \( \{(U_s, e_s) : 1 \leq s \leq t, m + 1 \leq s \leq t + m\} \). Then, it is easy to check that \( \{Z_t(\alpha_0, \beta_0, \psi_0)\} \) is a Martingale difference array with respect to the filtration \( \mathcal{F}_t \).

For Case (i), noting that \( \{X_t\} \) is stationary, the normality of \( \frac{1}{\sqrt{m}} \sum_{t=p+1}^m Z_t(\alpha_0, \beta_0, \psi_0) \)

follows immediately under Conditions C1–C3 by the center limit theory for Martingale differences of [31] and the Cramér–Wold device.

For Case (ii), trivially, we have

\[ \frac{1}{\sqrt{m}} \sum_{t=p+1}^m Z_{t,1}(\alpha_0, \beta_0, \psi_0) \xrightarrow{d} N(0, E(U_1^2)), \quad \text{as} \quad n \to \infty. \]

Next, note that \( |X_t| \xrightarrow{p} \infty \), which results in \( \frac{1}{\sqrt{m}} \sum_{t=p+1}^m w_t^2 \xrightarrow{p} 0 \) as \( n \to \infty \). Hence,

\[ \frac{1}{\sqrt{m}} \sum_{t=p+1}^m Z_{t,2}(\alpha_0, \beta_0, \psi_0) = \frac{1}{\sqrt{m}} \sum_{t=p+1}^m U_{t+1}w_{t+m-1} + \frac{1}{\sqrt{m}} \sum_{t=p+1}^m \delta_t \]

\[ = \frac{1}{\sqrt{m}} \sum_{t=p+1}^m \delta_t + o_p(1), \]

which implies \( \frac{1}{\sqrt{m}} \sum_{t=p+1}^m Z_{t,2}(\alpha_0, \beta_0, \psi_0) \xrightarrow{d} N(0, \sigma^2). \)
For \( \frac{1}{\sqrt{m}} \sum_{t=p+1}^{m} Z_{t,2+i}(\alpha_0, \beta_0, \psi_0), k = 1, 2, \cdots, p \), by using a similar proof to [30], we have

\[
\frac{1}{m} \sum_{t=p+1}^{m} E(Z_{t,2+i}^2(\alpha_0, \beta_0, \psi_0)|F_{t-1}) \tag{A3}
\]

\[
= E(U_{t}^2) \frac{1}{m} \sum_{t=p+1}^{m} (\Delta X_{t-i})^2
\]

\[
= E(U_{t}^2) \frac{1}{m} \sum_{t=p+1}^{m} \left( \frac{c}{n} X_{t-i-1} + \epsilon_{t-i} \right)^2
\]

\[
= E(U_{t}^2) \left\{ \frac{c^2}{n} \frac{1}{m} \sum_{t=p+1}^{m} \left( \frac{1}{\sqrt{n}} X_{t-i-1} \right)^2 + \frac{2c}{\sqrt{n}} \frac{1}{m} \sum_{t=p+1}^{m} X_{t-i-1} \epsilon_{t-i} + \frac{1}{m} \sum_{t=p+1}^{m} \epsilon_{t-i}^2 \right\}
\]

\[
= E(U_{t}^2) \left\{ \frac{c^2}{n} \int_{0}^{1} f^2(s) ds + \frac{2c}{\sqrt{n}} \int_{0}^{1} f(s) dW(s) + \lim_{t \to \infty} E(\epsilon_{t-i}^2) \right\}
\]

\[
= E(U_{t}^2) \cdot \lim_{t \to \infty} E(\epsilon_{t-i}^2) \cdot \{1 + o_p(1)\}, \text{ as } n \to \infty.
\]

and, for arbitrarily small \( d_1, d_2 \in (0, e_0) \),

\[
\frac{1}{m} \sum_{t=p+1}^{m} E(|Z_{t,2+i}(\alpha_0, \beta_0, \psi_0)|^2 I(|Z_{t,2+i}(\alpha_0, \beta_0, \psi_0)| \geq d_1 \sqrt{m})|F_{t-1})
\]

\[
\leq \frac{1}{d_1^2 m^2 / 2} \frac{1}{m} \sum_{t=p+1}^{m} E(|Z_{t,2+i}(\alpha_0, \beta_0, \psi_0)|^{2+d_2}|F_{t-1})
\]

\[
= \frac{E(U_{t}^2(2+d_2))}{d_1^2 m^2 / 2} \frac{1}{m} \sum_{t=p+1}^{m} \left| \frac{c}{n} X_{t-i-1} + \epsilon_{t-i} \right|^{2+d_2}
\]

\[
\leq \frac{E(U_{t}^2(2+d_2))}{d_1^2 m^2 / 2} \frac{2^{1+d_2}}{m} \sum_{t=p+1}^{m} \left| \frac{c}{n} X_{t-i-1} \right|^{2+d_2} + \frac{1}{m} \sum_{t=p+1}^{m} |\epsilon_{t-i}|^{2+d_2}
\]

\[
= \frac{E(U_{t}^2(2+d_2))}{d_1^2 m^2 / 2} \frac{2^{1+d_2}}{m} \sum_{t=p+1}^{m} |\epsilon_{t-i}|^{2+d_2} \cdot \{1 + o_p(1)\} \xrightarrow{p} 0, \text{ as } n \to \infty.
\]

Based on this and (A3), a simple application of the center limit theory for Martingale differences of [31] leads to the fact that, as \( n \to \infty \),

\[
\frac{1}{\sqrt{m}} \sum_{t=p+1}^{m} Z_{t,2+i}(\alpha_0, \beta_0, \psi_0) \xrightarrow{d} N(0, E(U_{t}^2) \cdot \lim_{t \to \infty} E(\epsilon_{t-i}^2)), \ k = 1, 2, \cdots, p.
\]

By noting that \( \{\delta_t\} \) is independent of the random observations, we can show that

\[
\frac{1}{m} \sum_{t=p+1}^{m} E(Z_{t,2}(\alpha_0, \beta_0, \psi_0)Z_{t,j}(\alpha_0, \beta_0, \psi_0)|F_{t-1}) \xrightarrow{p} 0,
\]

for \( j \neq 2 \). Similar to (A3), it is easy to check that

\[
\frac{1}{m} \sum_{t=p+1}^{m} E(Z_{t,j}(\alpha_0, \beta_0, \psi_0)Z_{t,j}(\alpha_0, \beta_0, \psi_0)|F_{t-1}) \xrightarrow{p} 0,
\]

for \( i, j \neq 2 \) as \( n \to \infty \). Hence, Part (a) follows immediately by using the Cramér–Wold device. This completes the proof of this lemma. \( \square \)
Proof of Theorem 1. We only prove Case (ii) as Cases (i) and (iii) are similar. For convenience, denote

\[ h(a_0, \beta, \psi, x) = \frac{1}{m} \sum_{t=p+1}^{m} \frac{Z_t(a_0, \beta, \psi)}{1 + x^\top Z_t(a_0, \beta, \psi)}, \text{ and} \]

\[ g_t(a_0, \beta, \psi) = \lambda^\top Z_t(a_0, \beta, \psi) \]

with \( \lambda \) being the solution to \( h(a_0, \beta, \psi, x) = 0 \) with respect to \( x \) for fixed \( \beta, \psi \).

The following proof is similar to that of Theorem 1 in [32]. Let \( \lambda = \varphi v \) with \( \varphi = ||\lambda|| \). Then, similar to [33], we have \( p_t = \frac{1}{m} \frac{1}{1 + \lambda^\top Z_t(a_0, \beta, \psi)}, t = p + 1, \cdots, m \), by using the Lagrange multipliers method. Note that

\[
0 = ||\lambda|| \frac{e^\top \frac{1}{m} \sum_{t=p+1}^{m} Z_t(a_0, \beta, \psi) Z_t(a_0, \beta, \psi)^\top e}{1 + \max_{p+1 \leq t \leq m} g_t(a_0, \beta, \psi)}
\]

\[
\leq \left| \frac{1}{m} \sum_{t=p+1}^{m} g_t(a_0, \beta, \psi) \right|
\]

\[
\leq \lambda^\top \frac{1}{m} \sum_{t=p+1}^{m} Z_t(a_0, \beta, \psi) \psi_0) + \lambda^\top \frac{1}{m} \sum_{t=p+1}^{m} (Z_t(a_0, \beta, \psi) - Z_t(a_0, \beta_0, \psi_0))
\]

\[
= ||\lambda|| \times O_p(1/\sqrt{m}),
\]

uniformly for \( (\hat{\beta}^\psi) \in B \). Hence,

\[ ||\lambda|| = O_p(1/\sqrt{m}), \]

uniformly for \( (\hat{\beta}^\psi) \in B \) by noting that \( \Sigma \) is positive definite. This, together with Part (a) of Lemma A1, shows

\[
\max_{p+1 \leq t \leq m} g_t(a_0, \beta, \psi) \leq ||\lambda|| \times \max_{p+1 \leq t \leq m} ||Z_t(a_0, \beta, \psi)|| = O_p(1/\sqrt{m})o_p(\sqrt{m}) = o_p(1),
\]

uniformly for \( (\hat{\beta}^\psi) \in B \). In turn,

\[
\left| \frac{1}{m} \sum_{t=p+1}^{m} Z_t(a_0, \beta, \psi) \frac{g_t^2(a_0, \beta, \psi)}{1 + g_t(a_0, \beta, \psi)} \right| \leq ||\lambda|| \cdot a_{\max} \cdot 2 \max_{p+1 \leq t \leq m} g_t(a_0, \beta, \psi) = o_p(1/\sqrt{m}),
\]

where \( a_{\max} \) denotes the maximum eigenvalue of \( \Sigma \).
Next, by Lemma 1, we have

\[ 0 = h(a_0, \beta, \psi, \lambda) \]

\[ = \frac{1}{m} \sum_{t=p+1}^m Z_t(a_0, \beta, \psi) \frac{1 - g_t^2(a_0, \beta, \psi) + g_t^2(a_0, \beta, \psi)}{1 + g_t(a_0, \beta, \psi)} \]

\[ = \frac{1}{m} \sum_{t=p+1}^m Z_t(a_0, \beta, \psi)(1 - g_t(a_0, \beta, \psi)) + \frac{1}{m} \sum_{t=p+1}^m Z_t(a_0, \beta, \psi) \frac{g_t^2(a_0, \beta, \psi)}{1 + g_t(a_0, \beta, \psi)} \]

\[ = \frac{1}{m} \sum_{t=p+1}^m Z_t(a_0, \beta, \psi)(1 - Z_t(a_0, \beta, \psi)\top \lambda) + o_p(1/\sqrt{m}) \]

\[ = \frac{1}{m} \sum_{t=p+1}^m Z_t(a_0, \beta, \psi) - \Sigma \lambda + o_p(1/\sqrt{m}), \]

uniformly for \((\hat{\beta}_1) \in B\). Then, we have

\[ \lambda = \Sigma^{-1} \frac{1}{m} \sum_{t=p+1}^m Z_t(a_0, \beta, \psi) + o_p(1/\sqrt{m}) \quad (A4) \]

uniformly for \((\hat{\beta}_1) \in B\) as \(n \to \infty\).

As did in the proof of Theorem 1 in \cite{32}, we can show by the Taylor expansion that

\[ \sup_B \left\| -2 \log L(a_0, \beta, \psi) - \xi_T^\top (\beta, \psi) \Sigma^{-1} \xi_n(\beta, \psi) \right\| = o_p(1), \quad \text{as } n \to \infty, \quad (A5) \]

where \(\xi_n(\beta, \psi) = \frac{1}{\sqrt{m}} \sum_{t=p+1}^m Z_t(a_0, \beta, \psi)\), based on \((A4)\) and Lemmas 1 and 2. Using this, we have that the minimizer \((\hat{\beta}, \hat{\psi}^\top)\) of \(-2 \log L(a_0, \beta, \psi) + 2 \log L(a_0, \beta_0, \psi_0)\) must lie in \(B\). Next, for any \((\hat{\beta}, \hat{\psi}) \in B\), write \((\sqrt{m}(\hat{\beta} - \beta_0)) = y\) with \(y = (y_1, \cdots, y_m)\) and \(\|y\| < C\) for some constant \(C\) given in \(B\). Then, by \((A2)\), a similar proof to that of Part (a) in Lemma \((A1)\) leads to the fact that

\[ \frac{1}{\sqrt{m}} \sum_{t=p+1}^m (Z_t(a_0, \beta, \psi) - Z_t(a_0, \beta_0, \psi_0)) \]

\[ = -\left\{ \frac{1}{m} \sum_{t=p+1}^m \left( \frac{1}{\sqrt{m}} \tilde{X}_{t-1}, \tilde{X}_{t-1}^2, \cdots, \tilde{X}_{t-1}^p \right) \right\} \left( \frac{m(\beta - \beta_0)}{\sqrt{m}(\psi - \psi_0)} \right) + o_p(1) \]

\[ = -\lim_{t \to \infty} \left( \int_0^1 \frac{1}{\sqrt{m}} L(s) ds \right) \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \begin{pmatrix} m(\beta - \beta_0) \sqrt{m}(\psi - \psi_0) \end{pmatrix} + o_p(1) \]

\[ := \mathbb{A} y + o_p(1), \]

where \(\tilde{X}_{t-1}, \tilde{X}_{t-1}^i, i = 1, 2, \cdots, p\), are defined in \((A2)\). Hence, we have

\[ \varphi(y) := -2 \log L \left( a_0, \beta_0 + \frac{y_1}{m}, \psi_0 + \frac{y_2}{\sqrt{m}} \right) + 2 \log L(a_0, \beta_0, \psi_0) \]

\[ = \xi_n^\top \left( \beta_0 + \frac{y_1}{m}, \psi_0 + \frac{y_2}{\sqrt{m}} \right) \Sigma^{-1} \xi_n(\beta_0, \psi_0) + o_p(1) \]

\[ = y^\top \mathbb{A}^\top \Sigma^{-1} \mathbb{A} y + 2 y^\top \mathbb{A}^\top \Sigma^{-1} \xi_n(\beta_0, \psi_0) + o_p(1), \]
uniformly for \((\hat{\beta}_\varphi) \in \mathcal{B}\). Then, the minimizer of \(\varphi(y)\) satisfies
\[
\hat{y} = - \left( \mathbb{A}^\top \Sigma^{-1} \mathbb{A} \right)^{-1} \mathbb{A}^\top \Sigma^{-1} \mathbb{A} \hat{\beta}_\varphi + \mathcal{O}_p(1).
\]
Using this, we further have
\[
-2 \log L \left( \alpha_0, \beta_0 + \frac{y_1}{m}, \psi_0 + \frac{y_2}{\sqrt{m}} \right)
= \mathbb{A}_n(\beta_0, \psi_0)^\top \left( \Sigma^{-1} - \mathbb{A}_n(\beta_0, \psi_0)^\top \mathbb{A}_n(\beta_0, \psi_0)^{-1} \right)^{-1} \mathbb{A}_n(\beta_0, \psi_0) + \mathcal{O}_p(1)
= \mathbb{A}_n(\beta_0, \psi_0)^\top \left( \text{diag}\{\sigma^2, 0, \cdots, 0\} \right)^{-1} \mathbb{A}_n(\beta_0, \psi_0) + \mathcal{O}_p(1)
= \sigma^{-2} \left( \frac{1}{\sqrt{m}} \sum_{i=1}^m Z_{i,2}(\alpha_0, \beta_0, \psi_0) \right)^2 + \mathcal{O}_p(1)
\]
\[
\xrightarrow{d} \chi^2_1,
\]
based on Lemmas 1 and 2 as \(n \to \infty\). This completes the proof of this Theorem. □

**Proof of Theorem 2.** We only show Case (i) as the other two cases can be proved in a similar fashion. Under the local alternative hypothesis \(\mathcal{H}_1\), we have
\[
Z_t(\alpha_0, \beta, \psi) = Z_t(\alpha_0 - \frac{C_0}{\sqrt{m}}, \beta, \psi) + \frac{C_0}{\sqrt{m}} \bar{X}_{t-1},
\]
where \(\bar{X}_{t-1} = (1, w_{t+m}, \Delta X_{t-1}, \cdots, \Delta X_{t-p})^\top\). Then, it is easy to show as \(n \to \infty\) that
\[
\frac{1}{\sqrt{m}} \sum_{i=1}^m Z_t(\alpha_0, \beta, \psi) = \frac{1}{\sqrt{m}} \sum_{i=1}^m Z_t(\alpha_0 - \frac{C_0}{\sqrt{m}}, \beta, \psi) + \frac{C_0}{\sqrt{m}} \bar{X}_{t-1}
= \frac{1}{\sqrt{m}} \sum_{i=1}^m Z_t(\alpha_0 - \frac{C_0}{\sqrt{m}}, \beta, \psi) + C_0 \gamma_1 + \mathcal{O}_p(1),
\]
and
\[
\frac{1}{m} \sum_{i=1}^m Z_t(\alpha_0, \beta, \psi) Z_t(\alpha_0, \beta, \psi)^\top = \frac{1}{m} \sum_{i=1}^m \left( Z_t(\alpha_0 - \frac{C_0}{\sqrt{m}}, \beta, \psi) + \frac{C_0}{\sqrt{m}} \bar{X}_{t-1} \right) \left( Z_t(\alpha_0 - \frac{C_0}{\sqrt{m}}, \beta, \psi) + \frac{C_0}{\sqrt{m}} \bar{X}_{t-1} \right)^\top
= \frac{1}{m} \sum_{i=1}^m Z_t(\alpha_0 - \frac{C_0}{\sqrt{m}}, \beta, \psi) Z_t(\alpha_0 - \frac{C_0}{\sqrt{m}}, \beta, \psi)^\top + \mathcal{O}_p(1).
\]
Based on the quantities \(\frac{1}{\sqrt{m}} \sum_{i=1}^m Z_t(\alpha_0 - \frac{C_0}{\sqrt{m}}, \beta, \psi)\) and \(\frac{1}{m} \sum_{i=1}^m Z_t(\alpha_0 - \frac{C_0}{\sqrt{m}}, \beta, \psi) Z_t(\alpha_0 - \frac{C_0}{\sqrt{m}}, \beta, \psi)^\top\) given in (A6) and (A7), respectively, the rest of the proof is similar to that of Lemmas 1 and 2 and Theorem 1. We omit the details. □

**References**
1. Amihud, Y. Illiquidity and stock returns cross-section and time-series effects. *J. Financ. Mark.* 2002, 5, 31–56. [CrossRef]
2. Baker, M.; Stein, J.C. Market liquidity as a sentiment indicator. *J. Financ. Mark.* 2004, 7, 271–299. [CrossRef]
3. Campbell, J.Y.; Shiller, R.J. The dividend-price ratio and expectations of future dividends and discount factors. *Rev. Financ. Stud.* 1988, 1, 195–228. [CrossRef]
4. Kaul, G. Predictable components in stock returns. *Handb. Stat.* 1996, 14, 269–296.
5. Keim, D.B.; Stambaugh, R.F. Predicting returns in the stock and bond markets. *J. Financ. Econ.* 1986, 17, 357–390. [CrossRef]
6. Fama, E.F.; French, K.R. Business conditions and expected returns on stocks and bonds. *J. Financ. Econ.* 1989, 25, 23–49. [CrossRef]
7. Zhu, F.; Liu, M.; Ling, S.; Cai, Z. Testing for structural change of predictive regression model to threshold predictive regression model. *J. Bus. Stat. 2022*. [CrossRef]
8. Cai, Z.; Jing, B.; Kong, X.; Liu, Z. Nonparametric regression with nearly integrated regressors under long-run dependence. *Econom. J.* 2017, 20, 118–138. [CrossRef]