Econometric Theory, **31**, 2015, 953–980.
doi:10.1017/S0266466614000590

TESTING INSTABILITY IN A PREDICTIVE REGRESSION MODEL WITH NONSTATIONARY REGRESSORS

Zongwu Cai
University of Kansas
and
Xiamen University

Yunfei Wang
Sun Trust Bank

Yonggang Wang
First Tennessee Bank

It is well known that allowing the coefficients to be time-varying in a predictive model with possibly nonstationary regressors can help to deal with instability in predictability associated with linear predictive models. In this paper, an L_2-type test statistic is proposed to test the stability of the coefficient vector, and the asymptotic distributions of the proposed test statistic are developed under both null and alternative hypotheses. A Monte Carlo experiment is conducted to evaluate the finite sample performance of the proposed test statistic and an empirical example is examined to demonstrate the practical application of the proposed testing method.

1. INTRODUCTION

A standard predictive regression has the following linear structural model

$$y_t = \alpha_0 + \alpha_1 x_{t-1} + \varepsilon_t \quad \text{and} \quad x_t = \rho x_{t-1} + u_t \ \text{ for } \ 1 \le t \le n, \tag{1}$$

where y_t is the dependent variable, say excess asset return at time t, x_{t-1} is a financial variable, such as the log dividend–price (d–p) ratio or the log earnings–price (e–p) ratio at time $t-1$, which is commonly formulated by an autoregressive model with order 1 (denoted by AR(1))[1] as in the second equation in (1), and innovations $\{(\varepsilon_t, u_t)\}$ in (1) are usually assumed in the finance literature to be independently and identically distributed (i.i.d.) bivariate normal $N(0, \Sigma)$

We thank the editor, Professor Peter C.B. Phillips, and the anonymous referees for their insightful comments that greatly improved our paper. Cai's research was supported, in part, by the National Nature Science Foundation of China grant #71131008 (Key Project). Address correspondence to Zongwu Cai, Department of Economics, University of Kansas, Lawrence, KS 66045, USA; e-mail: zongwucai@gmail.com.

with $\Sigma = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon u} \\ \sigma_{\varepsilon u} & \sigma_u^2 \end{pmatrix}$. Model (1) is commonly called a predictive regression. Note that it is easy to generalize model (1) to the multiple regression case; see Amihud, Hurvich, and Wang (2009) and Phillips, Li, and Gao (2013) for further discussions. For simplicity, it is assumed that x_t is one-dimensional.

As discussed in Amihud and Hurvich (2004), Phillips and Lee (2013), and Cai and Wang (2014), the ordinary least squares (OLS) estimator of model (1) is asymptotically unbiased with stationary regressors. However, when regressors are persistent, the OLS estimator has uncorrectable bias and the limiting theory is nonstandard. Many studies have proposed solutions of these problems; see, for example, Amihud and Hurvich (2004), Phillips and Lee (2013), and Cai and Wang (2014).

Note that the correlation coefficient between ε_t and u_t in (1) is $\gamma = \sigma_{\varepsilon u}/\sigma_\varepsilon \sigma_u$, which, unfortunately, is nonzero in many empirical studies; see, for example, Table 4 in Campbell and Yogo (2006) and Table 1 in Torous, Valkanov, and Yan (2004) for some real applications. This nonzero correlation between two innovations creates the so-called *embedded endogeneity* problem where x_{t-1} and ε_t may be correlated which leads to biased estimates (see Campbell and Yogo, 2006).

In the last decade, many empirical studies have shown significant in-sample evidence of predictability in asset returns, while the evidence of out-of-sample predictability appears to be very weak. In particular, Paye and Timmermann (2006) attributed this strong in-sample predictability, but weak out-of-sample predictability, to different predictive relations over time with possible structural changes. To examine this instability embedded in a predictive model, they studied the following model:

$$y_t = \alpha_{0t} + \alpha_{1t} x_{t-1} + \varepsilon_t \quad \text{and} \quad x_t = \rho x_{t-1} + u_t \ \text{ for } 1 \le t \le n \tag{2}$$

by assuming that both α_{0t} and α_{1t} change over time with possible structural breaks, and that x_t is stationary ($|\rho| < 1$). In other words, both α_{0t} and α_{1t} in (2) are assumed to be piecewise constant. They concluded that there is evidence of instability for the vast majority of models for international equity indices. However, as pointed out by Hansen (2001), Cai (2007), and Chen and Hong (2012), this assumption might be inappropriate in some real applications due to some leading driving forces of structural changes usually exhibiting evolutionary changes in the long term, and it is more reasonable to allow smooth structural changes over a period of time rather than sudden structural changes. Therefore, one could expect that both the trend and the prediction coefficient are smooth functions of time. Therefore, Wang (2010) applied the linear projection between two innovations to remove embedded endogeneity from the model, i.e., $\varepsilon_t = \beta_{1t} u_t + v_t$, and then considered

$$\begin{aligned} y_t &= \beta_{0t} + \beta_{1t} u_t + \beta_{2t} x_{t-1} + v_t \equiv \beta_t^T X_t + v_t \quad \text{and} \\ x_t &= \rho x_{t-1} + u_t \ \text{ for } 1 \le t \le n, \end{aligned} \tag{3}$$

where β_{it}, $i = 0, 1, 2$ are smooth functions of t, $\beta_t = (\beta_{0t}, \beta_{1t}, \beta_{2t})^T$, $X_t = (1, u_t, x_{t-1})^T$, $\rho = 1 + c/n$ with $c \leq 0$, and the error term $\{(u_t, v_t)\}$ is a strictly stationary mixing process. When $c = 0$, x_t is a unit root, or integrated, process, denoted by I(1). When $c < 0$ and is fixed, x_t is a nearly unit root, or integrated, process, denoted by NI(1) (see Phillips, 1988). Indeed, model (3) incorporates several known models as special cases; see Robinson (1989, 1991), Cai (2007), and Chen and Hong (2012) for the case where x_t is stationary. There are some papers in the literature that discuss the time-varying coefficient model when regressors are persistent and nonstationary; see Park and Hahn (1999), Wang (2010), and Phillips, Li, and Gao (2013). Especially, Philips, Li, and Gao (2013) generalized model (3) to a nonstationary multivariate regression that includes predictive regression and nonlinear cointegration.

In applications of (3), it is important to determine if coefficient $\beta_t = (\beta_{0t}, \beta_{1t}, \beta_{2t})$ is constant, i.e., to determine if a parametric linear model is appropriate. Therefore, the hypotheses can be formulated as

$$H_0 : \beta_t = \beta \text{ for some } \beta = (\beta_0, \beta_1, \beta_2) \in \mathbb{R}^3, \text{ versus } H_a : \beta_t \neq \beta \text{ for all } t \geq 0. \quad \textbf{(4)}$$

Clearly, the testing hypothesis formulated in (4) covers the special case $H_0 : \beta_{2t} = 0$ in (3), which is similar to the well-known test $H_0 : \alpha_1 = 0$ associated with the linear predictive model (1). Therefore, the testing hypothesis in (4) is more general than test $H_0 : \alpha_1 = 0$ in the literature.

The motivation of this study is to apply (3) and (4) to an analysis of the monthly S&P 500 excess return to test if it is predictable by some particular financial variables, such as the log e–p ratio and the log d–p ratio. If predictability exists, then this study will also test its stability. The detailed analysis of this dataset is reported in Section 4.3.

In the literature, there are several discussions about hypothesis $H_0 : \alpha_1 = 0$ based on (1). Stambaugh (1999) proved that the OLS estimation of the predictive slope coefficient in (1) is biased in finite samples, which renders the traditional t-test invalid. Using the bootstrap method in Kothari and Shanken (1997) to test the slope coefficient in (1) avoids the unknown distribution problem. Meanwhile, Lewellen (2004) not only proposed a hypothesis testing method based on the empirical observation that the autoregressive coefficient ρ in (1) is close to one, but also showed that the proposed method can dramatically improve the power of the test under the assumption that $\rho \approx 1$. The method in Lewellen (2004) was modified by Amihud and Hurvich (2004) to remove the assumption that $\rho \approx 1$. Also, a new method, called the augmented regression method, was proposed by Amihud et al. (2009) to test the predictive coefficients in (1). A comparison of the aforementioned methods can be found in Amihud et al. (2009). Recently, for nonstationary and/or heavy-tailed regressors, Cai and Wang (2014) derived the asymptotic theory and suggested using a Monte Carlo simulation to find the appropriate critical values for testing predictability, whereas Zhu, Cai, and Peng (2014) proposed novel empirical likelihood methods based on some weighted score equations to test predictability. In addition, an earlier work of Cavanagh, Elliott, and Stock (1995) proposed a Bonferroni t-test derived under local-to-unity asymptotics to obtain a valid inference for the coefficient α_1 and their method was revised by Campbell and Yogo (2006) to find a more powerful test. However, as demonstrated in Phillips (in press), the approach proposed by Campbell and Yogo (2006), based on confidence intervals, has zero coverage probability asymptotically as the sample size goes to infinity in the stationary regressor case with spurious predictability. These results may raise potential problems for practical applications when regressors are stationary. In fact, all of the aforementioned methods are based on the assumption that the predictive coefficients are simple constants instead of smooth functions of time t.

To test the instability of time-varying parameters in a nonparametric regression, Cai (2007) and Chen and Hong (2012) studied two consistent tests for smooth structural changes as well as for abrupt structural breaks by assuming that the predictor is stationary and uncorrelated to the innovation in the model. However, to the best of our knowledge, nothing in the literature discusses the hypothesis testing problem in (4), when regressors are highly persistent, possibly endogenous, and even nonstationary.

Here are two methods to consider the testing problem in (4), based on the nonparametric estimation procedure. The first is the conditional moment test proposed by Fan and Li (1996, 1999) and Zheng (1996), which may not be suitable for the testing problem in (4) because it is not formulated in a conditional moment, and the second is based on the integrated squared difference of the estimated function and the true function (see Li, Huang, Li, and Fu, 2002; Sun, Cai, and Li, 2008). We use the second method to construct the test statistic and to derive its asymptotic distribution.

Using alternative hypotheses to ours, other tests have been developed to test a linear/nonlinear cointegrating model, including the nonparametric specification tests of Sun et al. (2008), Xiao (2009), Wang and Phillips (2012), and Wu (2013). Without requiring $\{x_{t-1}\}$ to be independent of $\{u_t\}$, Wang and Phillips (2012) considered the problem of testing a linear cointegration model, $y_t = \theta_0 + \theta_1 x_{t-1} + u_t$, against a nonlinear cointegration model, $y_t = g(x_{t-1}) + u_t$, where $g(\cdot)$ is an unknown function and $\{x_t\}$ is a random walk process, while Wu (2013) tested a nonlinear cointegration model, $y_t = g_0(x_{t-1}, \theta) + u_t$, against a nonparametric cointegration model, $y_t = g(x_{t-1}) + u_t$, where $g_0(\cdot)$ is a known function and $\{x_t\}$ is a nearly integrated process. See Sun et al. (2008), Wang and Phillips (2012), and Wu (2013) for a summary of these tests.

The rest of this paper is organized as follows. Section 2 is devoted to deriving the test statistic, and its asymptotic results are presented in Section 3. We conduct a Monte Carlo simulation and present its results in Section 4. Also, an application of predictive regression with nonstationary regressors to an empirical example is reported in Section 4 to highlight the practical usefulness of the proposed testing procedure. Section 5 concludes the paper. All theoretical proofs of the asymptotic results are given in the Appendix.

2. TEST STATISTIC

Following Robinson (1989, 1991) and Cai (2007), we assume that $\beta_t = \beta(s_t)$, where $s_t = t/T$, and also that the second order derivative of $\beta(\cdot)$ is continuous in $[0,1]$; see Robinson (1989, 1991) and Cai (2007) for detailed discussions. Then, (4) becomes

$$H_0 : \beta(s) = \beta \text{ for some } \beta = (\beta_0, \beta_1, \beta_2) \in \mathbb{R}^3, \text{ versus}$$
$$H_a : \beta(s) \neq \beta \text{ for all } 0 \leq s \leq 1. \tag{5}$$

Now, following Li et al. (2002), we construct a test statistic based on the L_2-type test statistic as follows

$$\int_0^1 [\widehat{\beta}(s) - \widehat{\beta}]^T \omega(s) [\widehat{\beta}(s) - \widehat{\beta}] ds,$$

where $\widehat{\beta}$ is the OLS estimator of β under the null hypothesis, $\omega(s)$ is a weighting function, and $\widehat{\beta}(s)$ is a nonparametric estimator of $\beta(s)$ for any $0 \leq s \leq 1$; for example, the local linear or local constant estimator. If $\widehat{\beta}(s)$ is the local linear estimator, then

$$\begin{pmatrix} \widehat{\beta}_{ll}(s) \\ \widehat{\beta}_{ll}^{(1)}(s) \end{pmatrix} = \operatorname{argmin}_{\widehat{\beta}, \widehat{\beta}^{(1)}} \sum_{t=1}^{n} [y_t - X_t^T \widehat{\beta} - (s_t - s) X_t^T \widehat{\beta}^{(1)}]^2 K_h(s_t - s), \tag{6}$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function and $h = h(n)$ is the bandwidth satisfying $h \to 0$ and $nh \to \infty$ as $n \to \infty$. It is easy to show that the minimizer in (6) is given by

$$\begin{pmatrix} \widehat{\beta}_{ll}(s) \\ \widehat{\beta}_{ll}^{(1)}(s) \end{pmatrix} = \begin{pmatrix} S_{n,0}(s) & S_{n,1}^T(s) \\ S_{n,1}(s) & S_{n,2}(s) \end{pmatrix}^{-1} \begin{pmatrix} T_{n,0}(s) \\ T_{n,1}(s) \end{pmatrix} \equiv S_n^{-1}(s) T_n(s),$$

where $S_{n,j}(s) = \sum_{t=1}^n (s_t - s)^j K_h(s_t - s) X_t X_t^T$ for $0 \leq j \leq 2$ and $T_{n,j}(s) = \sum_{t=1}^n (s_t - s)^j K_h(s_t - s) X_t y_t$ for $0 \leq j \leq 1$. Alternatively, if $\widehat{\beta}(s)$ is the local constant estimator, then $\widehat{\beta}_{lc}(s) = S_{n,0}(s)^{-1} T_{n,0}(s)$. To avoid the random denominator problem, we consider the following. When $\widehat{\beta}(s)$ is the local linear estimator, we have

$$S_n(s) \begin{pmatrix} \widehat{\beta}_{ll}(s) - \widehat{\beta} \\ \widehat{\beta}_{ll}^{(1)}(s) - 0 \end{pmatrix} = \sum_{t=1}^{n} K_h(s_t - s) X_t \widehat{v}_t \begin{pmatrix} 1 \\ s_t - s \end{pmatrix},$$

where $\widehat{v}_t = y_t - X_t^T \widehat{\beta}$ is the parametric residual. When $\widehat{\beta}(s)$ is the local constant estimator, we have

$$S_{n,0}(s) \left(\widehat{\beta}_{lc}(s) - \widehat{\beta} \right) = \sum_{t=1}^{n} K_h(s_t - s) X_t \widehat{v}_t.$$

This implies that by taking an appropriate $\omega(s)$, the test statistic has the form

$$\int [\widehat{\beta}(s)-\widehat{\beta}]^T \omega(s)[\widehat{\beta}(s)-\widehat{\beta}]ds = \sum_{t=1}^{n}\sum_{r=1}^{n} X_t^T X_r \widehat{v}_t \widehat{v}_r \int K_h(s_t-s)K_h(s_r-s)ds.$$

Similar to Li et al. (2002), by removing the global center, i.e., the sum where $t=r$ in the above equation, and by replacing convolution kernel function $\int K_h(s_t-s) K_h(s_r-s)ds$ with $K_{tr} \equiv K((s_t-s_r)/h)$, we obtain the final test statistic

$$\widehat{I}_n = \frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_t^T X_r \widehat{v}_t \widehat{v}_r K_{tr}. \qquad \textbf{(7)}$$

In the next section, we discuss the asymptotic properties of $\widehat{I}_n$ under H_0 and H_a, respectively. Note that u_t is unknown in practice so that X_t in (7) should be $(1, \widehat{u}_t, x_{t-1})^T$, where $\widehat{u}_t = x_t - \widehat{\rho}\, x_{t-1}$.

3. ASYMPTOTIC THEORY

Before presenting the asymptotic distribution of the test statistic in (7), we need to list some assumptions.

Assumptions:

C.1 The error term $\{u_t\}$ has a zero mean and is a strictly stationary α-mixing sequence of size $-r/(r-2)$ with $r = 2+\delta$ for some $\delta > 0$ and $E|u_t|^r < \infty$.

C.2 Functions $\beta(\cdot)$ are twice continuously differentiable on $[0,1]$ and $\|\beta(\cdot)\|_{2q} < M < \infty$ for some $q > 1$.

C.3 The kernel function $K(\cdot)$ is symmetric and has a compact support $[-1,1]$. Also, $K(\cdot)$ satisfies $|K(u)-K(u^{'})| \le M|u-u^{'}|$ for some $M < \infty$.

C.4 The bandwidth h satisfies that $h \to 0$, $nh \to \infty$, and $n h^4 \to 0$.

C.5 Error term v_t satisfies $E(v_t)=0$, $E(v_t^2)=\sigma_v^2$ with σ_v^2 a positive constant, and $E(v_t^4) < \infty$. In addition, $\{v_t\}$ is independent of $\{x_t\}$ for all $1 \le t \le n$.

C.6 Error term $\{(u_t, v_t)\}$ is a strictly stationary β-mixing process with the mixing coefficient satisfying $\beta_k = O\left(\rho_1^{-k}\right)$ for some $\rho_1 > 1$.

Further, we assume the uniform consistency from Theorem 3.2 in Wang (2010), where

$$\sup_{|s|\le 1}|\widehat{\beta}_i(s)-\beta_i(s)| = O_p(q_n), \quad i=0,1 \ \text{ and}$$
$$\sup_{|s|\le 1}|\widehat{\beta}_2(s)-\beta_2(s)| = O_p(n^{-1/2} q_n)$$

under some regularity conditions, where $q_n = \sqrt{\ln n/nh} + h^2$. If $nh^4 \to 0$ as $n \to \infty$, it follows that $q_n \to 0$ so that

$$\sup_{|s|\le 1}|\widehat{\beta}_i(s)-\beta_i(s)| = o_p(1), \quad i=0,1 \quad \text{and}$$
$$\sup_{|s|\le 1}|\widehat{\beta}_2(s)-\beta_2(s)| = o_p(n^{-1/2}), \qquad \textbf{(8)}$$

which will be used in the proof of the asymptotic properties of the test statistic $\widehat{I}_n$. Also, the conditions given in Lemma A1 (see Appendix) and Assumption C.6 imply that

$$\xi_n(r) \equiv x_{[nr]}/\sqrt{n} \;\Rightarrow\; K_c(r), \tag{9}$$

where $K_c(r) = \int_0^r \exp((r-s)c)dW_u(s)$ is a diffusion process and $W_u(\cdot)$ is a one-dimensional Brownian motion with variance $\sigma_u^2 = \mathrm{Var}(u_t) + 2\sum_{k=2}^{\infty}\mathrm{Cov}(u_1, u_k)$. Clearly, $K_c(r)$ becomes $W_u(r)$ when $c = 0$. Here, and in what it follows, $\Rightarrow$ represents weak convergence, and $\xrightarrow{d}$ denotes convergence in distribution.

THEOREM 1. *Suppose that Assumptions C.1–C.6 hold. Then, under H_0, we have*

$$J_n = n\sqrt{h}\widehat{I}_n \xrightarrow{d} MN(0, \sigma^2), \tag{10}$$

where $MN(0,\sigma^2)$ is a mixed normal distribution with zero mean and conditional variance

$$\sigma^2 = 2\sigma_v^2\sigma_u^2 R^2(K)\int_0^1\int_0^r K_c^2(r)^T K_c^2(s)dsdr, \tag{11}$$

where $R^2(K) = \int_{-1}^1\int_0^r K^2(s)dsdr$ and $K_c(\cdot)$ is given in (9). In addition, if (10) and (11) hold, then

$$\tilde{\sigma}^2 = \frac{1}{n^4 h}\sum_{t=1}^n\sum_{r\neq t}^n \tilde{v}_t^2\tilde{v}_r^2\left(X_t^T X_r\right)^2 K_{tr}^2 \xrightarrow{p} \sigma^2, \tag{12}$$

where $\tilde{v}_t = y_t - X_t^T\widehat{\beta}_t$ is the nonparametric residual.

Remark 1. In the consistent estimate of σ^2, $\tilde{v}_t$ must be nonparametric residual. If a parametric residual is used, say, $\widehat{v}_t \equiv y_t - X_t^T\widehat{\beta}$, then $\widehat{v}_t = O(\sqrt{n})$ under H_a, and $\widehat{\sigma}^2$ may not be a consistent estimate of σ^2. See Phillips (1988) for the definition of a mixed normal distribution.

Next, to examine the power of test statistic J_n, we introduce $O_e(a_n)$ to denote an *exact* probability order of $a_n \to +\infty$. For example, $A_n = O_e(a_n)$ means that $A_n = O_p(a_n)$ but $A_n \neq o_p(a_n)$. Theorem 2 shows that J_n is a consistent test.

THEOREM 2. *Suppose that Assumptions C.1–C.6 hold. Then, under H_a, we have*

$$J_n = O_e(n^2\sqrt{h}), \qquad \text{i.e.,} \quad P(J_n > C_n) \to 1$$

as $n \to \infty$ for any nonstochastic sequence $C_n = o(n^2\sqrt{h})$.

Theorem 2 shows that, under H_a, the test statistic J_n diverges to $+\infty$ at the same rate whether or not $\beta_2(s)$ is constant.

4. APPLICATIONS

In this section, we conduct a simulation to examine the finite sample performance of the proposed test statistic. We first describe the simulation procedure and then report the simulation results under both null and alternative hypotheses.

4.1. Simulation Procedure

We use the test statistic given by (7) and its standard form, according to (10) in Theorem 1, can be constructed as

$$\widehat{H}_n = n\sqrt{h}\widehat{I}_n/\sqrt{\tilde{\sigma}^2}, \tag{13}$$

where $\tilde{\sigma}^2$ is the estimator of the conditional variance of J_n in Theorem 1, based on a nonparametric residual. From Section 3, we know that the asymptotic distribution of $\widehat{H}_n$ is conditional normal with zero mean and unit variance. The critical values of the proposed test statistic are simulated to compute the empirical size and power of the test. The simulation procedure is briefly described as follows.

(1) Under H_0, estimate the coefficients in model $y_t = X_t^T\beta + \upsilon_t$ using the OLS method and obtain residual $\widehat{\upsilon}_t$. Also, run regression $x_t = \rho x_{t-1} + u_t$ to obtain residual $\widehat{u}_t$.
(2) Apply the proposed two-stage estimation procedure on the simulated sample in Step (1) to obtain estimates of the time-varying coefficients, calculate the estimated nonparametric residual $\tilde{\upsilon}_t$, and then find the estimate of the conditional variance of J_n in Theorem 1, say $\tilde{\sigma}^2$.
(3) Replace with $\widehat{\upsilon}_t$ and $\tilde{\sigma}^2$ in (13) and calculate the test statistic $\widehat{H}_n$.
(4) Perform a large number of iterations, say $1,000$, to find the empirical distribution of $\{\widehat{H}_n\}$. The critical value at significance level α is given by the $(1-\alpha)$th quantile.

4.2. Simulation Results

To find the empirical size and power of our proposed test statistic, we consider the data generating process

$$y_t = \beta_0(s_t) + \beta_1(s_t)u_t + \beta_2(s_t)x_{t-1} + \upsilon_t,\ s_t \in [0,1];\ x_t = \rho x_{t-1} + u_t,\ \rho = 1 + c/n.$$

To estimate the size of this model, we choose $\beta_0(s_t) = 0$, $\beta_1(s_t) = \delta$ with $\delta = -0.75$ or -0.95, and $\beta_2(s_t) = 0$ for $s_t \in [0,1]$. We also choose different values for the persistency parameter c, i.e., $c = 0, -2$, or -20, corresponding to the state variable x_t being an I(1), NI(1), or stationary process. The two innovations, u_t and υ_t, are generated from the AR(1) models

$$u_t = 0.3u_{t-1} + e_{1t} \quad \text{and} \quad \upsilon_t = 0.3\upsilon_{t-1} + e_{2t},$$

where e_{1t} and e_{2t} are independently generated from normal distributions with zero mean and variances $\sigma_{e_1}^2 = 0.91$ and $\sigma_{e_2}^2 = 0.4$ when $\delta = -0.75$ or $\sigma_{e_2}^2 = 0.09$ when $\delta = -0.95$, respectively, which guarantee that both u_t and v_t have a standard normal distribution.

We choose four sample sizes of $n = 250$, 500, 750, and $1,000$ and repeat the simulation $m = 1,000$ times for each sample size. When applying the proposed two-stage estimation procedure, we need to specify the bandwidths at both stages. For the first stage, the bandwidth is $h_1 = d_1 n^{-2/5}$ with $d_1 = 1.0$, 2.5, and 5.0. Based on the mean absolute deviation error for estimating all three coefficients, $d_1 = 2.5$ is chosen for our test since $d_1 = 1.0$ and $d_1 = 5.0$ generate a bandwidth that is either too small or too large. At stage two, the bandwidth is selected by the cross-validation method. We also consider hypothesis testing at different nominal sizes of 10%, 5%, and 1% to check the effectiveness of the proposed test statistic at different significance levels. The simulated sizes at different bandwidths are listed in Tables 1–3. Note that we also conducted simulations for $\delta = 0.75$ and the results were similar to those in Tables 1–3. To save space, the results are not reported here. However, they are available from the authors upon request.

From Table 1, we observe clearly the convergence of the simulated size to the corresponding nominal size as the sample size increases, no matter how persistent the state variable is or which setting is chosen. In addition, there is no significant difference among the sizes when δ takes different values, no matter which bandwidth, d_1, is selected. Similar conclusions can be drawn from Tables 2 and 3. By comparing the results in Tables 1–3, we see that $d_1 = 2.5$ (Table 2) performs the best among the three bandwidth values. This shows that the proposed test statistic is consistent under the null hypothesis so that it delivers a right test size.

TABLE 1. The empirical test size for testing constant predictability at $d_1 = 1.0$

		$\delta = -0.75$			$\delta = -0.95$		
		$c = 0$	$c = -2$	$c = -20$	$c = 0$	$c = -2$	$c = -20$
10%	$n = 250$	0.188	0.184	0.162	0.184	0.180	0.164
	$n = 500$	0.136	0.130	0.126	0.130	0.130	0.126
	$n = 750$	0.116	0.116	0.110	0.114	0.111	0.112
	$n = 1000$	0.100	0.100	0.101	0.100	0.100	0.100
5%	$n = 250$	0.110	0.108	0.094	0.108	0.104	0.096
	$n = 500$	0.074	0.072	0.066	0.070	0.070	0.069
	$n = 750$	0.058	0.060	0.054	0.060	0.058	0.058
	$n = 1000$	0.050	0.052	0.050	0.050	0.050	0.050
1%	$n = 250$	0.036	0.034	0.028	0.032	0.032	0.030
	$n = 500$	0.020	0.019	0.016	0.020	0.018	0.018
	$n = 750$	0.014	0.014	0.012	0.014	0.014	0.014
	$n = 1000$	0.012	0.012	0.010	0.012	0.012	0.012

TABLE 2. The empirical test size for testing constant predictability at $d_1 = 2.5$

		$\delta = -0.75$			$\delta = -0.95$		
		$c = 0$	$c = -2$	$c = -20$	$c = 0$	$c = -2$	$c = -20$
10%	$n = 250$	0.108	0.108	0.138	0.108	0.108	0.138
	$n = 500$	0.100	0.100	0.112	0.100	0.100	0.110
	$n = 750$	0.100	0.100	0.104	0.100	0.100	0.107
	$n = 1000$	0.100	0.100	0.100	0.100	0.100	0.100
5%	$n = 250$	0.054	0.053	0.072	0.052	0.054	0.072
	$n = 500$	0.050	0.050	0.056	0.050	0.050	0.056
	$n = 750$	0.050	0.050	0.052	0.050	0.050	0.053
	$n = 1000$	0.050	0.050	0.050	0.050	0.050	0.050
1%	$n = 250$	0.012	0.012	0.018	0.010	0.012	0.018
	$n = 500$	0.010	0.010	0.012	0.010	0.010	0.012
	$n = 750$	0.010	0.010	0.012	0.010	0.010	0.012
	$n = 1000$	0.010	0.010	0.010	0.010	0.010	0.010

TABLE 3. The empirical test size for testing constant predictability at $d_1 = 5.0$

		$\delta = -0.75$			$\delta = -0.95$		
		$c = 0$	$c = -2$	$c = -20$	$c = 0$	$c = -2$	$c = -20$
10%	$n = 250$	0.052	0.062	0.104	0.052	0.064	0.102
	$n = 500$	0.072	0.076	0.086	0.072	0.080	0.084
	$n = 750$	0.092	0.092	0.092	0.090	0.094	0.092
	$n = 1000$	0.102	0.100	0.100	0.100	0.100	0.100
5%	$n = 250$	0.022	0.028	0.052	0.024	0.028	0.050
	$n = 500$	0.034	0.036	0.042	0.032	0.038	0.040
	$n = 750$	0.046	0.046	0.048	0.044	0.046	0.046
	$n = 1000$	0.050	0.050	0.050	0.050	0.050	0.050
1%	$n = 250$	0.004	0.006	0.012	0.004	0.006	0.010
	$n = 500$	0.006	0.008	0.008	0.007	0.008	0.008
	$n = 750$	0.010	0.010	0.010	0.010	0.010	0.010
	$n = 1000$	0.010	0.010	0.010	0.010	0.010	0.010

Next, to compute the power of the proposed test, we consider a sequence of alternatives indexed by $b \in [0, 1]$ as

$$H_a : \beta(s_t) = \beta + b(\beta^0(s_t) - \beta), \quad s_t \in [0, 1],$$

where $\beta^0(\cdot) \in \mathcal{R}^3$ with $\beta_0^0(s_t) = 0.02\exp(-0.8 + 4s_t) - 0.12$, $\beta_1^0(s_t) = 0.2\sin(4s_t + 20)$, and $\beta_2^0(s_t) = 0.3s_t - 0.3\exp(-7s_t^2) - 0.05$. Notice that when $b = 0$, the specified alternative collapses to the null hypothesis and the power becomes the test size. Similar to the test size investigation, here we also consider three different bandwidth values. For sample size, we choose $n = 250$, 500, and 750. Figure 1 reports the results for $d_1 = 2.5$ and nominal size 5%. The results for $d_1 = 1.0$ and $d_1 = 5.0$ and nominal sizes 1% and 10% are similar and so are not presented here to save space.

The plots on the left-hand side (when $\delta = -0.75$) of Figure 1 illustrate the power curves when the correlation between two innovations is relatively weak. From these plots, we see that the test becomes more powerful as the sample size increases. Also, the power curves in the top two panels approach to one much

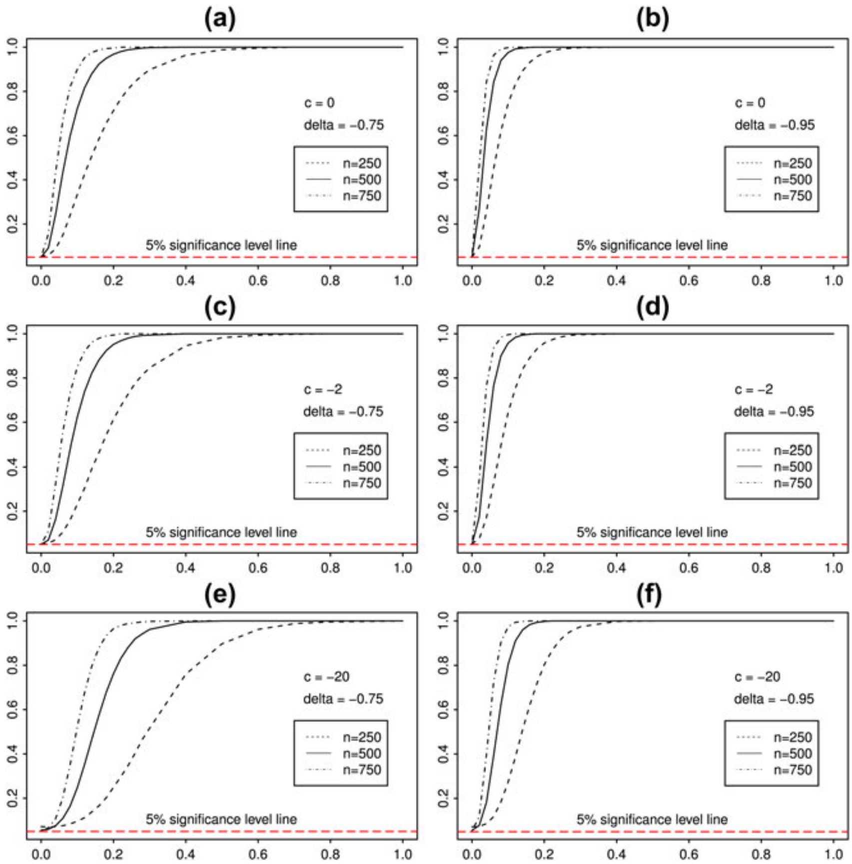

FIGURE 1. Plots of the empirical power curves against b for the test hypothesis when $d_1 = 2.5$ with nominal size 5% in Section 4.2. The dashed, solid, and dashed-dotted lines represent $n = 250$, 500, and 750, respectively. (a) $c = 0$ and $\delta = -0.75$; (b) $c = 0$ and $\delta = -0.95$; (c) $c = -2$ and $\delta = -0.75$; (d) $c = -2$ and $\delta = -0.95$; (e) $c = -20$ and $\delta = -0.75$; and (f) $c = -20$ and $\delta = -0.95$.

faster when the state variable is nonstationary ($c = 0$ or $c = -2$). This is reasonable since the proposed test should be more powerful when the state variable in (3) is highly persistent, or even nonstationary. From the plots on the right-hand side (when $\delta = -0.95$) of Figure 1, similar conclusions can be made, except that the power curves converge to one quicker than in those on the left-hand side of Figure 1. Finally, for $c = 0$ and $c = -2$ with $n = 500$ and $n = 750$, when b is greater than or equal to 0.2, the power is approximately one. For $c = -20$, the power is approximately one when $b \geq 0.3$ for $n = 500$ and $n = 750$. This implies that the proposed test is quite powerful.

4.3. An Empirical Application

In this section, we apply the proposed test procedure to analyze the monthly S&P 500 excess return. We use the log e–p ratio as the univariate predictor during the time period 1938:12 to 1998:12 and use the log d–p ratio to predict asset return from 1945:12 to 2005:12. The monthly S&P 500 stock price, dividends, and earnings data were obtained from Professor Robert Shiller's website.[2] The d–p ratio is calculated as the ratio of average dividends during the last year over the current stock price. The e–p ratio is computed as the average earnings over the past ten years divided by the current price. The one-month T-bill from the CRSP is used to calculate the excess return. See Cai and Wang (2014) for details on variable definitions and calculations. We investigate the predictability of these two time series on asset return. Both sample periods selected have a length of 721 months, which makes the sample size possibly large enough to obtain a reliable test result.

First, we examine the persistency of the predictors. To this end, we consider the AR(1) model with an intercept term as the working model for both predictors; that is

$$x_t = \theta + \rho x_{t-1} + u_t,$$

where x_t stands for the log e–p ratio or the log d–p ratio for the two time periods described above. Then, we consider testing the null hypothesis that $\theta = 0$ and $\rho = 1$. The least squares method is applied to find the estimates $\widehat{\theta}_{OLS}$ and $\widehat{\rho}_{OLS}$. Based on Hamilton (1994), neither has a limiting Gaussian distribution. By using a Monte Carlo simulation as suggested in Hamilton (1994), we find the 95% confidence intervals of the coefficients. We summarize the OLS estimates and their confidence intervals in Table 4. For both predictors, the confidence interval

TABLE 4. OLS estimates (95% C.I.) for the AR(1) models

	log e–p ratio (1938:12–1998:12)	log d–p ratio (1945:12–2005:12)
$\widehat{\theta}_{OLS}$	−0.0039 (−0.0076, 0.0084)	−0.0070 (−0.0077, 0.0082)
$\widehat{\rho}_{OLS}$	0.9992 (0.9841, 1.0005)	0.9983 (0.9839, 1.0005)

for θ covers zero and for ρ covers one. Thus, we cannot reject the random walk null hypothesis for either time series.

Next, we apply the least squares method to rerun the AR(1) model without the constant term, $x_t = \rho x_{t-1} + u_t$, for both predictors to get $\widehat{u_t}$. Then, we apply the two-step estimation method proposed in Cai and Wang (2014) to estimate the constant coefficients in

$$r_t = \beta_0 + \beta_1 \widehat{u_t} + \beta_2 x_{t-1} + v_t. \tag{14}$$

The estimates of the coefficients, together with their 95% confidence intervals, are listed in Table 5. From these results, we see that the log d–p ratio can be used to predict asset return during its sample period since zero is not covered by the confidence interval for β_2. However, the log e–p ratio may not have predictive power.

One can conclude from the above testing results that the constancy of the coefficient is uncertain. So, we now apply the proposed test procedure to test if the coefficient vector is a constant vector. To this end, we propose the following varying coefficient model

$$r_t = \beta_0(s_t) + \beta_2(s_t) x_{t-1} + \varepsilon_t, \quad x_t = \rho x_{t-1} + u_t, \quad \rho = 1 + c/n,$$

where $\varepsilon_t = \beta_1(s_t)\widehat{u_t} + v_t$, and $s_t \in [0, 1]$.

We apply a two-stage estimation procedure to estimate the functional coefficients. To find an appropriate bandwidth in the first stage, we set $h_1 = d_1 n^{-2/5}$ with d_1 selected from 0.2 to 10 in increments of 0.2. The bandwidth which generates the lowest mean square error is used for testing. Hence, we select $d_1 = 1.2$ and $d_1 = 2.2$ for the log e–p ratio and the log d–p ratio, respectively. Following the proposed test procedure, we calculate the values of the test statistic and obtain their corresponding p-values by repeating the simulation $m = 1{,}000$ times. The results are listed in Table 6. For both predictors, the p-values are close to zero which implies that we have a strong evidence to reject the null of constant coefficient vectors. Together with the results in Table 5, we conclude that both the log e–p ratio and the log d–p ratio have time-varying predictability to forecast asset return, although their predictability may be weak at some points during the sample periods.

TABLE 5. Two-step estimates (95% C.I.) for constant coefficients in (14)

	log e–p ratio (1938:12–1998:12)	log d–p ratio (1945:12–2005:12)
$\widehat{\beta_0}$	0.0042 (0.0037, 0.0047)	0.0098 (0.0091, 0.0104)
$\widehat{\beta_1}$	−1.0045 (−1.0048, −1.0042)	−0.9903 (−0.9907, −0.9900)
$\widehat{\beta_2}$	0.0005 (−0.0004, 0.0014)	0.0022 (0.0009, 0.0036)

TABLE 6. The value of test statistic and its p-value for the empirical example

State variable	Test statistic	p-Value
log e–p ratio	6.179	0.007
log d–p ratio	14.286	<0.001

5. CONCLUSION

In this paper, we propose a consistent nonparametric test for testing the null hypothesis of constant coefficients against the alternative of nonparametric smooth coefficients in a varying coefficient predictive regression model. We show that the proposed test statistic converges to a mixed normal distribution under the null hypothesis.

There are several issues related to our paper which should be addressed. First, we only consider the case where x_{t-1} might be correlated with ε_t through u_t since ε_t and u_t are correlated. It would be of interest to generalize the results of this paper to where x_{t-1} and ε_t are directly correlated, similar to Wang and Phillips (2012) for regression model. Additionally, it would be of further interest to consider testing problems such as whether structural changes exist as argued by Paye and Timmermann (2006); that is, to test if nonparametric coefficients are piecewise constant (model (3) versus the model considered in Paye and Timmermann, 2006). Furthermore, when x_t is a multiple predictor, testing whether a portion of varying coefficients are constant is challenging since the model under the null hypothesis becomes a semiparametric varying-coefficient model. This is similar to Sun et al. (2008) for regression model. Finally, in practice, selecting the bandwidth in the proposed test statistic is of great importance. These issues are left as future research topics.

NOTES

1. A higher order of AR model for x_t is possible. For simplicity, our focus here is on the AR(1) model.
2. http://www.econ.yale.edu/shiller/data.htm.

REFERENCES

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

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

Cai, Z. (2007) Trending time varying coefficient time series models with serially correlated errors. *Journal of Econometrics* 137, 163–188.

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

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

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

Chen, B. & Y. Hong (2012) Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* 80, 1157–1183.

Fan, Y. & Q. Li (1996) Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* 65, 865–890.

Fan, Y. & Q. Li (1999) Central limit theorem for degenerate U-statistic of absolutely regular process: With an application to model specification testing. *Journal of Nonparametric Statistics* 10, 245–271.

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

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.

Hansen, B.E. (2001) The new econometrics of structural change: Dating breaks in U.S. labor productivity. *Journal of Economic Perspectives* 15, 117–128.

Kothari, S.P. & J. Shanken (1997) Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics* 44, 169–203.

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

Li, Q., C.J. Huang, D. Li, & T. Fu (2002) Semiparametric smooth coefficient model. *Journal of Business and Economics Statistics* 20, 412–422.

Park, J.Y. & S.B. Hahn (1999) Cointegrating regressions with time varying coefficients. *Econometric Theory* 15, 664–703.

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

Phillips, P.C.B. (1988) Regression theory for near-integrated time series. *Econometrica* 56, 1021–1043.

Phillips, P.C.B. (2014) On confidence intervals for autoregressive roots and predictive regression. *Econometrica* 82, 1177–1195.

Phillips, P.C.B & 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., D. Li, & J. Gao (2013). Estimating Smooth Structural Change in Cointegration Models. Discussion paper no. 1910, Cowles Foundations, Yale University.

Rio, E. (1995) The functional law of the iterated logarithm for stationary strongly mixing sequences. *Annals of Probability* 23, 1188–1203.

Robinson, P.M. (1989) Nonparametric estimation of time-varying parameters. In P. Hackled & A.H. Westland (eds.), *Statistical Analysis and Forecasting of Economic Structural Change*, pp. 253–264. Springer-Verlag.

Robinson, P.M. (1991) Time-varying nonlinear regression. In P. Hackled & A.H. Westland (eds.), *Economic Structure Change, Analysis and Forecasting*, pp. 179–190. Springer-Verlag.

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

Sun, Y., Z. Cai, & Q. Li (2008) Consistent Nonparametric Test on Parametric Smooth Coefficient Model with Nonstationary Data. Working paper, Department of Economics, University of Guelph.

Torous, W., R. Valkanov, & S. Yan (2004) On predicting stock returns with nearly integrated explanatory variables. *Journal of Business* 77, 937–966.

Wang, Y. (2010) Essays on predictive regression models for asset returns. Ph.D. dissertation, University of North Carolina at Charlotte.

Wang, Q. & P.C.B. Phillips (2012) A specification test for nonlinear nonstationary models. *Annals of Statistics* 40, 727–758.

Wu, L. (2013) Consistent nonparametric test on nonlinear regression models with nearly integrated covariates. Ph.D. dissertation, University of North Carolina at Charlotte.

Xiao, Z. (2009) Functional coefficient co-integration models. *Journal of Econometrics* 152, 81–92.
Zheng, J.X. (1996) A consistent test of functional form via nonparametric estimation technique. *Journal of Econometrics* 75, 263–289.
Zhu, F., Z. Cai, & L. Peng (2014) Predictive regressions for macroeconomics data. *The Annals of Applied Statistics* 8, 577–594.

APPENDIX: Mathematical Proofs

To prove the theorems in Section 3, we need the following two lemmas.

LEMMA A1. *Let $a_n^* = n^{-\theta_*}(\ln n)^{\lambda_*}$, where $\theta_* = 1/2 - 1/(2+\delta_*)$ with $0 < \delta_* \le 2$ and $\lambda_* > 0$ is a function of δ_*. Under the assumption that u_t is a stationary α-mixing sequence with mixing coefficient $\alpha(n)$ that satisfies*

$$E|u_t|^r < \infty, \quad \text{and} \quad \sum_{n=1}^{\infty} \alpha(n)^{1/(2+\delta_*)-1/r} < \infty$$

for some $r > 2+\delta_*$, it follows that the nearly I(1) process $\xi_n(r) = x_{[nr]}/\sqrt{n}$ for $0 \le r \le 1$ has the following strong approximation

$$\sup_{0\le r\le 1} |\xi_n(r) - K_c(r)| = O_{as}(a_n^*),$$

where $K_c(\cdot)$ is the diffusion process given in (9) and $O_{as}(1)$ means almost surely (a.s.).

Proof of Lemma A1. See Lemma 3.3 in Wang (2010). ■

LEMMA A2. *Let $\{S_{ni}, \mathcal{F}_{ni}, 1 \le i \le k_n, n \ge n\}$ be a zero-mean, square-integrable martingale array with differences X_{ni}, and let η^2 be an a.s. finite random variable (r.v.) Suppose that*

$$V_{nk_n}^2 = \sum E(X_{ni}^2|\mathcal{F}_{n,i-1}) \xrightarrow{p} \eta^2, \quad \sum_i E[X_{ni}^2 I(|X_{ni}| > \varepsilon)|\mathcal{F}_{n,i-1}] \xrightarrow{p} 0 \text{ for all } \quad \varepsilon > 0,$$

and the σ-fields are nested, that is $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$ for $1 \le i \le k_n$, $n \ge 1$, then

$$S_{n,k_n} = \sum_i X_{ni} \xrightarrow{d} Z,$$

where the r.v. Z has the characteristic function $E\left[\exp(-\frac{1}{2}\eta^2 t^2)\right]$.

Proof of Lemma A2. See Corollary 3.1 in Hall and Heyde (1980). ■

Proof of Theorem 1. Under H_0, $\widehat{\upsilon}_t = y_t - X_t^T \widehat{\beta} = \upsilon_t - X_t^T \left(\widehat{\beta} - \beta\right)$. Decompose $\widehat{I}_n$ as follows:

$$\widehat{I}_n = \frac{1}{n^3 h} \sum_{t=1}^{n} \sum_{r\ne t}^{n} X_t^T X_r \widehat{\upsilon}_t \widehat{\upsilon}_r K_{tr}$$

$$= \frac{1}{n^3 h} \sum_{t=1}^{n} \sum_{r\ne t}^{n} X_t^T X_r \left[\upsilon_t - X_t^T(\widehat{\beta} - \beta)\right]\left[\upsilon_r - (\widehat{\beta} - \beta)^T X_r\right] K_{tr}$$

$$= \frac{1}{n^3 h} \sum_{t=1}^{n} \sum_{r \neq t}^{n} X_t^T X_r \left[\upsilon_t \upsilon_r + (\widehat{\beta} - \beta)^T X_r X_t^T (\widehat{\beta} - \beta) - 2 \upsilon_r X_t^T (\widehat{\beta} - \beta) \right] K_{tr}$$

$$= I_{n,1} + (\widehat{\beta} - \beta)^T M_{n,2} (\widehat{\beta} - \beta) - 2 M_{n,3} (\widehat{\beta} - \beta), \tag{A.1}$$

where

$$I_{n,1} = \frac{1}{n^3 h} \sum_{t=1}^{n} \sum_{r \neq t}^{n} X_t^T X_r \upsilon_t \upsilon_r K_{tr}, \quad M_{n,2} = \frac{1}{n^3 h} \sum_{t=1}^{n} \sum_{r \neq t}^{n} X_t^T X_r X_r X_t^T K_{tr},$$

and $M_{n,3} = \frac{1}{n^3 h} \sum_{t=1}^{n} \sum_{r \neq t}^{n} X_t^T X_r \upsilon_r X_t^T K_{tr}.$

We now discuss the above three parts in (A.1) separately. First, let us consider $I_{n,1}$. By the fact that $\widehat{u}_t = u_t - (\widehat{\rho} - \rho) x_{t-1} = u_t + O_p(n^{-1/2})$, $X_t = (1, u_t, x_{t-1})^T + O_p(n^{-1/2})$. By ignoring the higher order term, to abuse the notation, we still use X_t to denote $(1, u_t, x_{t-1})^T$ here, and in what it follows. Then, $I_{n,1}$ can be written as

$$I_{n,1} = \frac{1}{n^3 h} \sum_{t=1}^{n} \sum_{r \neq t}^{n} \upsilon_t \upsilon_r K_{tr} + \frac{1}{n^3 h} \sum_{t=1}^{n} \sum_{r \neq t}^{n} u_t u_r \upsilon_t \upsilon_r K_{tr} + \frac{1}{n^3 h} \sum_{t=1}^{n} \sum_{r \neq t}^{n} x_{t-1} x_{r-1} \upsilon_t \upsilon_r K_{tr}$$

$$\equiv I_{n,11} + I_{n,12} + I_{n,13}, \tag{A.2}$$

where the definitions of $I_{n,11}$, $I_{n,12}$, and $I_{n,13}$ are clearly apparent. For $I_{n,11}$, we have $E(I_{n,11}) = 0$ and

$$E(I_{n,11})^2 = E\left[\frac{2}{n^6 h^2} \sum_{t=2}^{n} \sum_{r=1}^{t-1} \upsilon_t^2 \upsilon_r^2 K_{tr}^2 \right] = \frac{4\sigma_\upsilon^4}{n^6 h} \sum_{t=2}^{n} \sum_{r=1}^{t-1} \frac{K_{tr}^2}{h}$$

$$= \frac{2\sigma_\upsilon^4}{n^4 h} \left[\int_0^1 \int_0^t K^2 \left(\frac{t-r}{h}\right) dr\, dt + o(1) \right]$$

$$= \frac{2\sigma_\upsilon^4}{n^4 h} \left[\int_0^1 \int_0^{t/h} \frac{1}{h} K^2(s)\, ds\, dt + o(1) \right]$$

$$= \frac{2\sigma_\upsilon^4}{n^4 h} \left[\int_0^1 \int_0^1 K^2(s)\, ds\, dt + o(1) \right] = O(n^{-4} h^{-1}). \tag{A.3}$$

Therefore,

$$I_{n,11} = O_p(n^{-2} h^{-1/2}). \tag{A.4}$$

Similarly, it is easy to obtain $E(I_{n,12}) = 0$. Based on the assumption that u_t and υ_t are uncorrelated, we have

$$E(I_{n,12})^2 = E\left[\frac{4}{n^6 h^2} \sum_{t=1}^{n} \sum_{r=1}^{t-1} u_t^2 u_r^2 \upsilon_t^2 \upsilon_r^2 K_{tr}^2 \right] = \frac{2\sigma_\upsilon^4}{n^6 h} \sum_{t=1}^{n} \sum_{r=1}^{t-1} \frac{K_{tr}^2}{h} E\left[u_t^2 u_r^2 \right]$$

$$= O(n^{-4} h^{-1}).$$

Hence, we obtain

$$I_{n,12} = O_p(n^{-2}h^{-1/2}). \tag{A.5}$$

As for $I_{n,13}$, let $z_t = n^{-2}h^{-1/2}x_{t-1}\upsilon_t \sum_{r=1}^{t-1} x_{r-1}\upsilon_r K_{tr}$. Then, $n\sqrt{h}I_{n,13} = 2\sum_{t=2}^{n} z_t$. Given $E(\upsilon_t|\mathcal{F}_{n,t}) = 0$, we can show that $E_{t-1}(z_t) = 0$, which implies that $\{n\sqrt{h}I_{n,13}, \mathcal{F}_{n,t}\}$ is a martingale. We use the martingale central limit theorem in Lemma A2 to establish its asymptotic distribution. More specifically, we need to show that

$$V_n^2 = \sum_{t=2}^{n} E_{t-1}\left(z_t^2\right) \xrightarrow{p} \sigma^2/2 \tag{A.6}$$

$$\text{and} \quad \sum_{t=2}^{n} E_{t-1}\left[z_t^2 I\left(|z_t| > \varepsilon\right)\right] \xrightarrow{p} 0 \quad \text{for all} \quad \varepsilon > 0. \tag{A.7}$$

First, we prove (A.6). Indeed,

$$\begin{aligned}
V_n^2 &= \sum_{t=2}^{n} E_{t-1}(z_t^2) = \frac{1}{n^4 h}\sum_{t=2}^{n} E_{t-1}\left[\left(x_{t-1}\upsilon_t \sum_{r=1}^{t-1} x_{r-1}\upsilon_r K_{tr}\right)^2\right] \\
&= \frac{1}{n^4 h}\sum_{t=2}^{n}\sum_{r=1}^{t-1}\sum_{s=1}^{t-1} \upsilon_r\upsilon_s x_{r-1}x_{s-1}K_{tr}K_{ts}x_{t-1}^2\upsilon_t^2 \\
&= \frac{\sigma_\upsilon^2}{n^4 h}\sum_{t=2}^{n}\sum_{r=1}^{t-1}\sum_{s=1}^{t-1} x_{r-1}x_{s-1}x_{t-1}^2 K_{ts}K_{tr}\upsilon_r\upsilon_s \\
&= \frac{\sigma_\upsilon^2}{n^4 h}\sum_{t=2}^{n}\sum_{r=s}^{t-1} x_{t-1}^2 x_{r-1}^2 K_{tr}^2\upsilon_r^2 + \frac{\sigma_\upsilon^2}{n^4 h}\sum_{t=2}^{n}\sum_{r\neq s}^{t-1} \upsilon_r\upsilon_s x_{t-1}^2 x_{r-1}x_{s-1}K_{ts}K_{tr} \\
&\equiv V_{n,1}^2 + V_{n,2}^2,
\end{aligned} \tag{A.8}$$

where the definitions of $V_{n,i}^2$, $i = 1, 2$, are obvious. The strong approximation in Lemma A1 and Condition C.3 give

$$\begin{aligned}
\frac{1}{n^4 h}\sum_{t=2}^{n} x_{t-1}^2 \sum_{r=1}^{t-1} x_{r-1}^2 K_{tr}^2\upsilon_r^2 &= \frac{1}{n}\sum_{t=2}^{n}\left(\frac{x_{t-1}}{\sqrt{n}}\right)^2 \frac{1}{n}\sum_{r=1}^{t-1}\left(\frac{x_{r-1}}{\sqrt{n}}\right)^2 \frac{K_{tr}^2}{h}\upsilon_r^2 \\
&\xrightarrow{p} \int_0^1\int_0^t K_c^2(t)K_c^2(r)K^2\left(\frac{t-r}{h}\right)\upsilon_r^2\,dr\,dt \\
&= \int_0^1\int_0^{t/h} K_c^2(t)K_c^2(t-sh)K^2(s)\upsilon_r^2\,ds\,dt \\
&= \int_0^1\int_0^1 K_c^4(t)K^2(s)\,ds\,dt\int_0^1 \upsilon_s^2\,ds + s.o. \\
&= \sigma_u^2\int_0^1\int_0^1 K_c^4(t)K^2(s)\,ds\,dt + s.o.,
\end{aligned}$$

where the last equation is obtained by applying a Taylor expansion and $\sigma_u^2 = \int_0^1 \upsilon_s^2 ds$. Combined with $E(\upsilon_r^2) = \sigma_\upsilon^2$, we have

$$V_{n,1}^2 \xrightarrow{p} \frac{\sigma^2}{2} \equiv \sigma_\upsilon^2 \sigma_u^2 \int_0^1 \int_0^1 K_c^4(t) K^2(s) ds dt.$$

$$= \sigma_\upsilon^2 \sigma_u^2 R^2(K) \int_0^1 \int_0^r K_c^2(s) K_c^2(r) ds dr. \tag{A.9}$$

Second, similar to the proof of (A.9), we show that

$$V_{n,2}^2 = \frac{\sigma_\upsilon^2}{n^4 h} \sum_{t=2}^{n} \sum_{r \neq s}^{t-1} \upsilon_r \upsilon_s x_{r-1} x_{s-1} x_{t-1}^2 K_{ts} K_{tr}$$

$$= \sigma_\upsilon^2 h \times \frac{1}{n^3} \sum_{t=2}^{n} \left(\frac{x_{t-1}}{\sqrt{n}} \right)^2 \sum_{r=2}^{t-1} x_{r-1} \upsilon_r \sum_{s \neq r}^{t-1} x_{s-1} \upsilon_s \frac{K_{ts} K_{tr}}{h^2} = O_p(h). \tag{A.10}$$

Applying (A.9) and (A.10) to (A.8) gives

$$V_n^2 = \sigma^2/2 + o_p(1), \quad \text{which proves (A.6).} \tag{A.11}$$

Now let us check (A.7). A basic calculation similar to (A.10) shows that

$$\sum_{t=2}^{n} E_{t-1}(z_t^4) = \frac{1}{n^8 h^2} \sum_{t=2}^{n} E_{t-1} \left[\left(x_{t-1} \upsilon_t \sum_{r=1}^{t-1} x_{r-1} \upsilon_r K_{tr} \right)^4 \right]$$

$$= \frac{\mu_4^2}{n^8 h^2} \sum_{t=2}^{n} \sum_{r=1}^{t-1} E_{t-1} (x_{t-1} x_{r-1} K_{tr})^4$$

$$+ \frac{\mu_4 \sigma_\upsilon^4}{n^8 h^2} \sum_{t=2}^{n} \sum_{r \neq s}^{t-1} E_{t-1} \left(x_{t-1}^4 x_{r-1}^2 x_{s-1}^2 K_{ts}^2 K_{tr}^2 \right)$$

$$= \frac{\mu_4^2}{n^4 h} \sum_{t=2}^{n} \left(\frac{x_{t-1}}{\sqrt{n}} \right)^4 \sum_{r=1}^{t-1} \left(\frac{x_{r-1}}{\sqrt{n}} \right)^4 \frac{K_{tr}^4}{h}$$

$$+ \frac{\mu_4 \sigma_\upsilon^4}{n^4} \sum_{t=2}^{n} \left(\frac{x_{t-1}}{\sqrt{n}} \right)^4 \sum_{r=1}^{t-1} \left(\frac{x_{r-1}}{\sqrt{n}} \right)^2 \sum_{s \neq r}^{t-1} \left(\frac{x_{s-1}}{\sqrt{n}} \right)^2 \frac{K_{ts}^2 K_{tr}^2}{h^2}$$

$$= O_p\left((\log\log n)^4 \right) [O_p(n^{-2} h^{-1}) + O_p(n^{-1})] \to 0,$$

where the last equation holds since $K_{tr}^4 = O(h)$ and $K_{ts}^2 K_{tr}^2 = O(h^2)$. Theorem 2 in Rio (1995) implies that

$$\sup_{|r| \leq 1} |\xi_n(r)| = O_p(\sqrt{\log\log n}). \tag{A.12}$$

Then, applying Chebyshev's inequality, we have, for any $\varsigma > 0$,

$$\begin{aligned}
\mathrm{P}\left(\sum_{t=2}^{n} E_{t-1}\left[z_t^2 I\left(|z_t| > \varepsilon\right)\right] > \varsigma\right) &= \mathrm{P}\left(\sum_{t=2}^{n} E_{t-1}\left[z_t^2 I\left(\frac{|z_t|}{\varepsilon} > 1\right)\right] > \varsigma\right) \\
&\leq \mathrm{P}\left(\frac{1}{\varepsilon^2}\sum_{t=2}^{n} E_{t-1}\left(z_t^4\right) > \varsigma\right) \\
&\leq \frac{1}{\varepsilon^2 \varsigma}\sum_{t=2}^{n} E\left[E_{t-1}\left(z_t^4\right)\right] = \frac{1}{\varepsilon^2 \varsigma}\sum_{t=2}^{n} E\left(z_t^4\right) \to 0.
\end{aligned}$$

This implies that (A.7) holds. According to Lemma A2, we have

$$n\sqrt{h} I_{n,13} = 2\sum_{t=2}^{n} z_t \xrightarrow{d} \mathrm{MN}(0,\sigma^2). \tag{A.13}$$

By combining (A.2), (A.4), (A.5), and (A.13), we conclude that

$$n\sqrt{h} I_{n,1} \xrightarrow{d} \mathrm{MN}(0,\sigma^2). \tag{A.14}$$

Next, we focus on the second term in (A.1). Let $X_{1t} = (1 \;\; u_t)^T$. We decompose $M_{n,2}$ such that

$$\begin{aligned}
M_{n,2} &= \frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_t^T X_r X_r X_t^T K_{tr} \\
&= \frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}\left(X_{1t}^T X_{1r} + x_{t-1}x_{r-1}\right)\begin{pmatrix} X_{1r}X_{1t}^T & X_{1r}x_{t-1} \\ X_{1t}^T x_{r-1} & x_{t-1}x_{r-1}\end{pmatrix} K_{tr} \\
&= \frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}\begin{pmatrix} X_{1t}^T X_{1r} X_{1r} X_{1t}^T & X_{1t}^T X_{1r} X_{1r} x_{t-1} \\ X_{1t}^T X_{1r} X_{1t}^T x_{r-1} & X_{1t}^T X_{1r} x_{t-1} x_{r-1}\end{pmatrix} K_{tr} \\
&\quad + \frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}\begin{pmatrix} X_{1r} X_{1t}^T x_{t-1} x_{r-1} & X_{1r} x_{t-1}^2 x_{r-1} \\ X_{1t}^T x_{t-1} x_{r-1}^2 & x_{t-1}^2 x_{r-1}^2\end{pmatrix} K_{tr}.
\end{aligned} \tag{A.15}$$

We check each component. First, we have

$$\begin{aligned}
E\left|\frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_{1t}^T X_{1r} X_{1r} X_{1t}^T K_{tr}\right| &\leq \frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} E\left|X_{1t}^T X_{1r} X_{1r} X_{1t}^T K_{tr}\right| \\
&= O(n^{-1}),
\end{aligned} \tag{A.16}$$

which implies that

$$\frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_{1t}^T X_{1r} X_{1r} X_{1t}^T K_{tr} = O_p(n^{-1}).$$

In addition, recall that $\xi_{nt} = x_t/\sqrt{n}$ and $\xi_n(r) = \xi_{n,[nr]} \Rightarrow K_c(r)$. Then, (A.14) implies that

$$\begin{aligned} E\left|\frac{1}{n^3h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_{1t}^T X_{1r} X_{1r} x_{t-1} K_{tr}\right| &= \frac{1}{n^{5/2}h} E\left|\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_{1t}^T X_{1r} X_{1r}\frac{x_{t-1}}{\sqrt{n}} K_{tr}\right| \\ &= O(\sqrt{\log\log n})\frac{1}{n^{5/2}h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} E\left|X_{1t}^T X_{1r} X_{1r} K_{tr}\right| \\ &= O(n^{-1/2}\sqrt{\log\log n}) = o(1). \end{aligned} \quad \textbf{(A.17)}$$

Similarly,

$$\begin{aligned} E\left|\frac{1}{n^3h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_{1t}^T X_{1r} x_{t-1} x_{r-1} K_{tr}\right| &= \frac{1}{n^2h} E\left|\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_{1t}^T X_{1r}\frac{x_{t-1}}{\sqrt{n}}\frac{x_{r-1}}{\sqrt{n}} K_{tr}\right| \\ &= O(\log\log n)\frac{1}{n^2h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} E\left|X_{1t}^T X_{1r} K_{tr}\right| \\ &= O(\log\log n), \end{aligned}$$

$$\begin{aligned} E\left|\frac{1}{n^3h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_{1r} x_{t-1}^2 x_{r-1} K_{tr}\right| &= \frac{1}{n^{3/2}h} E\left|\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_{1r}\frac{x_{t-1}^2}{n}\frac{x_{r-1}}{\sqrt{n}} K_{tr}\right| \\ &= O(n^{1/2}(\log\log n)^{3/2}), \end{aligned}$$

and

$$\begin{aligned} E\left|\frac{1}{n^3h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} x_{t-1}^2 x_{r-1}^2 K_{tr}\right| &= \frac{1}{nh} E\left|\sum_{t=1}^{n}\sum_{r\neq t}^{n}\frac{x_{t-1}^2}{n}\frac{x_{r-1}^2}{n} K_{tr}\right| \\ &= O(n(\log\log n)^2). \end{aligned} \quad \textbf{(A.18)}$$

Together with (A.15), we have

$$M_{n,2} = \begin{pmatrix} 1 & n^{1/2}(\log\log n)^{1/2} \\ n^{1/2}(\log\log n)^{1/2} & n(\log\log n) \end{pmatrix} O_p(\log\log n).$$

Further, under H_0, it can be proved that $\widehat{\beta}_i - \beta_i = O_p(n^{-1/2})$ for $i = 0, 1$ and that $\widehat{\beta}_2 - \beta_2 = O_p(n^{-1})$ based on the central limit theorem and the properties of a nearly integrated process. These, combined with the above result, imply that

$$\begin{aligned} &\left(\widehat{\beta} - \beta\right)^T M_{n,2}\left(\widehat{\beta} - \beta\right) \\ &\quad = \begin{pmatrix} n^{-1/2} & n^{-1} \end{pmatrix} \begin{pmatrix} n^{-1}(\log\log n)^{-1} & n^{-1/2}(\log\log n)^{-1/2} \\ n^{-1/2}(\log\log n)^{-1/2} & 1 \end{pmatrix} \begin{pmatrix} n^{-1/2} \\ n^{-1} \end{pmatrix} \\ &\qquad \times O_p\left(n^{-1}(\log\log n)^2\right) \end{aligned}$$

$$= \left(n^{-1/2}(\log\log n)^{-1} \quad n^{-1}\right)\begin{pmatrix} n^{-1/2} \\ n^{-1} \end{pmatrix} O_p\left(n^{-1}(\log\log n)^2\right)$$
$$= O_p\left(n^{-1}(\log\log n)\right). \tag{A.19}$$

Finally, we analyze the last term in (A.1). Clearly, $E(M_{n,3}) = 0$ since $E(v_t) = 0$. So, we consider the second moment. First, decompose $M_{n,3}$ to

$$M_{n,3} = \frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_t^T X_r v_r X_t^T K_{tr} = \frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}\left(X_t^T X_r X_{1t}^T, X_t^T X_r x_{t-1}\right) v_r K_{tr}$$
$$\equiv (H_1, H_2),$$

where the definitions of H_1 and H_2 are obvious. In addition, recall that $X_t^T X_r = X_{1t}^T X_{1r} + x_{t-1}x_{r-1}$ is dominated by $x_{t-1}x_{r-1}$. Then, applying a similar method to that used for (A.16)–(A.18), we have

$$E\left|\frac{1}{n^6h^2}\sum_{t=1}^{n}\sum_{r\neq t}^{n}(X_t^T X_r)^2 X_{1t}^T X_{1t} K_{tr}^2\right| = O\left(n^{-2}h^{-1}(\log\log n)^2\right),$$

$$E\left|\frac{1}{n^6h^2}\sum_{t=1}^{n}\sum_{r\neq t, r\neq t'}^{n}(X_t^T X_r)^2 x_{t-1}^2 K_{tr}^2\right| = O\left(n^{-1}h^{-1}(\log\log n)^3\right),$$

$$E\left|\frac{1}{n^6h^2}\sum_{t=1}^{n}\sum_{t'\neq t}^{n}\sum_{r\neq t, r\neq t'}^{n} X_t^T X_r X_r^T X_{t'} K_{tr} K_{t'r}\right| = O\left(n^{-1}(\log\log n)^3\right),$$

and

$$E\left|\frac{1}{n^6h^2}\sum_{t=1}^{n}\sum_{t'\neq t}^{n}\sum_{r\neq t, r\neq t'}^{n} X_t^T X_r X_r^T X_{t'} x_{t-1} x_{t'-1} K_{tr} K_{t'r}\right| = O\left((\log\log n)^3\right).$$

Hence,

$$E|H_1 H_1^T| = \frac{\sigma_v^2}{n^6h^2} E\left|\sum_{t=1}^{n}\sum_{t'=1}^{n}\sum_{r\neq t, r\neq t'}^{n} X_t^T X_r X_r^T X_{t'} X_{1t}^T X_{1t} K_{tr} K_{t'r}\right|$$
$$= \frac{\sigma_v^2}{n^6h^2} E\left|\sum_{t=1}^{n}\sum_{r\neq t}^{n}(X_t^T X_r)^2 X_{1t}^T X_{1t} K_{tr}^2\right|$$
$$+ \frac{\sigma_v^2}{n^6h^2} E\left|\sum_{t=1}^{n}\sum_{t'\neq t}^{n}\sum_{r\neq t, r\neq t'}^{n} X_t^T X_r X_r^T X_{t'} X_{1t}^T X_{1t'} K_{tr} K_{t'r}\right| + (s.o.)$$
$$= O\left(n^{-2}h^{-1}(\log\log n)^2\right) + O\left(n^{-1}(\log\log n)^2\right) + (s.o.)$$
$$= O\left(n^{-1}(\log\log n)^2\right) + (s.o.),$$

where $(s.o.)$ denotes some smaller order term. Similarly,

$$
\begin{aligned}
E|H_2^2| &= \frac{\sigma_v^2}{n^6h^2} E\left|\sum_{t=1}^{n}\sum_{r\neq t}^{n}(X_t^T X_r)^2 x_{t-1}^2 K_{tr}^2\right| \\
&\quad + \frac{\sigma_v^2}{n^6h^2} E\left|\sum_{t=1}^{n}\sum_{t'\neq t}^{n}\sum_{r\neq t, r\neq t'}^{n} X_t^T X_r X_r^T X_{t'} x_{t-1} x_{t'-1} K_{tr} K_{t'r}\right| + (s.o.) \\
&= O\left(n^{-1}h^{-1}(\log\log n)^3\right) + O\left((\log\log n)^3\right) + (s.o.) \\
&= O\left((\log\log n)^3\right) + (s.o.).
\end{aligned}
$$

These imply that $M_{n,3} = \left(n^{-1/2}(\log\log n)^{-1/2}, 1\right) O_p\left((\log\log n)^{3/2}\right)$, and that

$$
M_{n,3}(\widehat{\beta}-\beta) = O_p\left(n^{-1}(\log\log n)\right). \tag{A.20}
$$

By applying (A.14), (A.19), and (A.20) to (A.1) and applying Assumption C.4, we obtain the asymptotical distribution of J_n in (10).

Finally, we need to prove the consistency of the estimate of σ^2. Note that $\tilde{v}_t = (y_t - X_t^T \beta_t) + (X_t^T \beta_t - X_t^T \widehat{\beta}_t) = v_t - X_t^T[\widehat{\beta}_t - \beta_t] = v_t + o_p(1)$ by (8). Hence, applying the same method used to obtain (A.9), we have

$$
\begin{aligned}
\tilde{\sigma}^2 &= \frac{1}{n^4h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}\tilde{v}_t^2\tilde{v}_r^2\left(X_t^T X_r\right)^2 K_{tr}^2 \\
&= \frac{1}{n^4h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}[v_t+o_p(1)]^2[v_r+o_p(1)]^2\left(X_t^T X_r\right)^2 K_{tr}^2 \\
&= \frac{1}{n^4h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}[v_t+o_p(1)]^2[v_r+o_p(1)]^2\left(X_{1t}^T X_{1r} + x_{t-1}x_{r-1}\right)^2 K_{tr}^2 \\
&= \frac{1}{n^4h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}[v_t+o_p(1)]^2[v_r+o_p(1)]^2 \\
&\quad \times\left[(X_{1t}^T X_{1r})^2 + x_{t-1}^2 x_{r-1}^2 + 2X_{1t}^T X_{1r} x_{t-1} x_{r-1}\right] K_{tr}^2 \\
&= \frac{1}{n^4h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} v_t^2 v_r^2 x_{t-1}^2 x_{r-1}^2 K_{tr}^2 + o_p(1),
\end{aligned}
$$

and

$$
E_{t-1}(\tilde{\sigma}^2) = \frac{\sigma_v^2}{n^4h}\sum_{t=2}^{n}\sum_{r=1}^{t-1} x_{t-1}^2 x_{r-1}^2 E_{t-1}(K_{tr}^2 v_r^2) + o_p(1) \xrightarrow{p} \sigma^2.
$$

Therefore, we finish the proof of Theorem 1. ∎

Proof of Theorem 2. Under H_a, we have $y_t = X_t^T \beta_t + v_t$. So, we decompose the OLS estimate of β to

$$\widehat{\beta} = \left(\sum_t u_t X_t^T\right)^{-1} \sum_t X_t \left(u_t^T \beta_t + v_t\right)$$
$$= \left(\sum_t X_t X_t^T\right)^{-1} \sum_t X_t X_t^T \beta_t + \left(\sum_t X_t X_t^T\right)^{-1} \sum_t X_t v_t. \tag{A.21}$$

Based on the well-known properties of a stationary and nearly unit root process, we obtain the order of each element in the second term of (A.21) as

$$\left(\sum_t X_t X_t^T\right)^{-1} \sum_t X_t v_t = \begin{pmatrix} n & \sum_t u_t & \sum_t x_{t-1} \\ \sum_t u_t & \sum_t u_t^2 & \sum_t u_t x_{t-1} \\ \sum_t x_{t-1} & \sum_t u_t x_{t-1} & \sum_t x_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_t v_t \\ \sum_t u_t v_t \\ \sum_t x_{t-1} v_t \end{pmatrix}$$
$$= \begin{pmatrix} n^{-1} & n^{-1} & n^{-3/2} \\ n^{-1} & n^{-1/2} & n^{-3/2} \\ n^{-3/2} & n^{-3/2} & n^{-2} \end{pmatrix} \begin{pmatrix} n^{1/2} \\ n^{1/2} \\ n \end{pmatrix} O_p(1) = \begin{pmatrix} O_p(n^{-1/2}) \\ O_p(1) \\ O_p(n^{-1}) \end{pmatrix}.$$

This implies that the second term in (A.21) has a smaller order than does the first term. So we only focus on the first term in the remaining proof. First, let

$$D_n = n^{-4}\Big[n\Sigma u_t^2 \Sigma x_{t-1}^2 + 2\Sigma u_t \Sigma x_{t-1} \Sigma u_t x_{t-1} - \Sigma u_t^2 (\Sigma x_{t-1})^2$$
$$- n(\Sigma u_t x_{t-1})^2 - (\Sigma u_t)^2 \Sigma x_{t-1}^2\Big],$$

$$S_n = \begin{pmatrix} S_{n,11} & S_{n,12} & S_{n,13} \\ S_{n,12} & S_{n,22} & S_{n,23} \\ S_{n,13} & S_{n,23} & S_{n,33} \end{pmatrix}, \quad \text{and} \quad R_n = \begin{pmatrix} R_{n,1} \\ R_{n,2} \\ R_{n,3} \end{pmatrix},$$

where

$$S_{n,11} = n^{-3}\left[\Sigma u_t^2 \Sigma x_{t-1}^2 - (\Sigma u_t x_{t-1})^2\right], S_{n,12} = n^{-3}\left[\Sigma u_t x_{t-1} \Sigma x_{t-1} - \Sigma u_t \Sigma x_{t-1}^2\right],$$
$$S_{n,13} = n^{-3}\left[\Sigma u_t \Sigma u_t x_{t-1} - \Sigma u_t^2 \Sigma x_{t-1}\right], S_{n,22} = n^{-3}\left[n\Sigma x_{t-1}^2 - (\Sigma x_{t-1})^2\right],$$
$$S_{n,23} = n^{-3}\left[\Sigma x_t \Sigma x_{t-1} - n\Sigma x_t x_{t-1}\right], S_{n,33} = n^{-3}\left[n\Sigma u_t^2 - (\Sigma u_t)^2\right],$$
$$R_{n,1} = n^{-1}\left[\Sigma \beta_{0t} + \Sigma u_t \beta_{1t} + \Sigma x_{t-1} \beta_{2t}\right],$$
$$R_{n,2} = n^{-1}\left[\Sigma u_t \beta_{0t} + \Sigma u_t^2 \beta_{1t} + \Sigma u_t x_{t-1} \beta_{2t}\right],$$

and

$$R_{n,3} = n^{-1}\left[\Sigma x_{t-1} \beta_{0t} + \Sigma u_t x_{t-1} \beta_{1t} + \Sigma x_{t-1}^2 \beta_{2t}\right].$$

Then, let $\widehat{\beta} = \left(\sum_t X_t X_t^T\right)^{-1} \sum_t X_t X_t^T \beta_t + (s.o.) = D_n^{-1} S_n R_n + (s.o.)$. Define $B_1 = \int_0^1 K_c(r)dr$, $B_2 = \int_0^1 K_c^2(r)dr$, $B_3 = \int_0^1 K_c(r)dW_u(r) + \Omega_1$, $\sigma_a^2 = E(u_t^2)$, and

$\omega_j = E(\beta_{jt})$, where $K_c(r)$ is defined in (9), and $\Omega_1 = \sum_{k=2}^{\infty} E(u_1 u_k)$. Then, it can be easily proved that $D_n = \sigma_a^2(B_2 - B_1) + o_p(1) \equiv D + o_p(1)$. Also, we have that

$$\begin{aligned}
&S_{n,11} = \sigma_a^2 B_2 - n^{-1} B_3^2 + (s.o.), \quad S_{n,12} = n^{-1/2} B_3 B_1 + (s.o.),\\
&\quad S_{n,13} = n^{-1/2}\sigma_a^2 B_1 + (s.o.),\\
&S_{n,22} = (B_2 - B_1^2) + (s.o.), \quad S_{n,23} = n^{-1} B_3 + (s.o.), \quad S_{n,33} = n^{-1}\sigma_a^2 + (s.o.),\\
&R_{n,1} = \omega_0 + n^{1/2} B_1 \omega_2, \quad R_{n,2} = \sigma_a^2 \omega_2 + B_3 \omega_2, \text{ and}\\
&\quad R_{n,3} = n^{1/2}\omega_0 B_1 + B_3\omega_1 + n B_2 \omega_2.
\end{aligned}$$

Now we consider the order of $\widehat{\beta}$ under the different scenarios given in Theorem 2.

Case (I1): $\boldsymbol{\beta}_{1t} \neq \boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_{2t} \neq \boldsymbol{\beta}_2$. We split this case into two subcases: (I1a) neither β_{1t} nor β_{2t} is constant, and (I1b) only one is constant.

(I1a): Neither $\boldsymbol{\beta}_{1t}$ nor $\boldsymbol{\beta}_{2t}$ is constant. By extracting the leading term of $n^{-4} S_n R_n$, we have

$$\begin{aligned}
\widehat{\beta} &= D^{-1}\begin{pmatrix} S_{n,11}R_{n,1} + S_{n,12}R_{n,2} + S_{n,13}R_{n,3} \\ S_{n,12}R_{n,1} + S_{n,22}R_{n,2} + S_{n,23}R_{n,3} \\ S_{n,13}R_{n,1} + S_{n,23}R_{n,2} + S_{n,33}R_{n,3} \end{pmatrix} + (s.o.)\\
&= D^{-1}\begin{pmatrix} n^{1/2}C_1 + o_p(n^{1/2}) \\ C_2 + o_p(1) \\ C_3 + o_p(1) \end{pmatrix},
\end{aligned}$$

where $C_1 = 2\sigma_a^2 B_1 B_2 \omega_2$, $C_2 = B_1^2 B_2 \omega_2 + B_2 B_3 \omega_2 + (B_2 - B_1^2)(\sigma_a^2 + B_3)\omega_2$, and $C_3 = \sigma_a^2(B_1^2 + B_2)\omega_2$. Therefore, recalling the last two terms in (A.1), we have

$$\begin{aligned}
I_{n,2} &= \left(\widehat{\beta} - \beta_t\right)^T M_{n,2}\left(\widehat{\beta} - \beta_r\right) = \frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_t^T X_r X_t^T\left(\widehat{\beta} - \beta_t\right)\left(\widehat{\beta} - \beta_r\right)^T X_r K_{tr}\\
&= \frac{1}{n^3 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}(1 + u_t u_r + x_{t-1}x_{r-1})\\
&\quad \times\left[(n^{1/2}D^{-1}C_1 - \beta_{0t}) + u_t(D^{-1}C_2 - \beta_{1t}) + x_{t-1}(D^{-1}C_3 - \beta_{3t})\right]\\
&\quad \times\left[(n^{1/2}D^{-1}C_1 - \beta_{0r}) + u_r(D^{-1}C_2 - \beta_{1r}) + x_{r-1}(D^{-1}C_3 - \beta_{2r})\right] K_{tr} + (s.o.)\\
&= \frac{n}{n^2 h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}\left(\frac{1}{n} + \frac{u_t u_r}{n} + \frac{x_{t-1}x_{r-1}}{n}\right)\\
&\quad \times\left[D^{-1}C_1 + \frac{u_t}{\sqrt{n}}(D^{-1}C_2 - \beta_{1t}) + \frac{x_{t-1}}{\sqrt{n}}(D^{-1}C_3 - \beta_{3t})\right]\\
&\quad \times\left[D^{-1}C_1 + \frac{u_r}{\sqrt{n}}(D^{-1}C_2 - \beta_{1r}) + \frac{x_{r-1}}{\sqrt{n}}(D^{-1}C_3 - \beta_{2r})\right] K_{tr} + (s.o.)\\
&= O_p(n).
\end{aligned}$$

Similarly,

$$
\begin{aligned}
I_{n,3} &= 2M_{n,3}(\widehat{\beta}-\beta_t) = \frac{2}{n^3h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_t^T X_r v_r X_t^T(\widehat{\beta}-\beta_t)K_{tr} \\
&= \frac{2}{n^3h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}\Big[(n^{1/2}D^{-1}C_1-\beta_{0t})+u_t(D^{-1}C_2-\beta_{1t})+x_{t-1}(D^{-1}C_3-\beta_{3t})\Big] \\
&\quad \times (1+u_tu_r+x_{t-1}x_{r-1})v_rK_{tr}+(s.o.) \\
&= \frac{2\sqrt{n}}{n^2h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}\left[D^{-1}C_1+\frac{u_t}{\sqrt{n}}\left(D^{-1}C_2-\beta_{1t}\right)+\frac{x_{t-1}}{\sqrt{n}}\left(D^{-1}C_3-\beta_{3t}\right)\right] \\
&\quad \times \left(\frac{u_tu_r}{n}+\frac{x_{t-1}x_{r-1}}{n}\right)v_rK_{tr}+(s.o.) \\
&= O_p(\sqrt{n}).
\end{aligned}
$$

Recall that $I_{n,1}=O_p(n^{-1}h^{-\frac{1}{2}})$. Hence, $I_{n,2}$ is the leading term of I_n and it follows that $J_n = n\sqrt{h}I_n = O_p(n^2\sqrt{h})$, and it diverges to $+\infty$ at the rate of $n^2\sqrt{h}$.

(I1b): One of $\boldsymbol{\beta}_{1t}$ or $\boldsymbol{\beta}_{2t}$ is constant, say, $\boldsymbol{\beta}_{1t}=\boldsymbol{\beta}_1$, but $\boldsymbol{\beta}_{2t}\neq\boldsymbol{\beta}_2$. Define

$$
R_{n,1}^* = \sum_t \beta_{0t} + \sum_t x_{t-1}\beta_{2t}; \quad R_{n,2}^* = \sum_t u_t\beta_{0t} + \sum_t u_t x_{t-1}\beta_{2t};
$$

and $\quad R_{n,3}^* = \sum_t x_{t-1}\beta_{0t} + \sum_t x_{t-1}^2\beta_{2t}.$

Then, by a simple calculation and a similar method to that used above, we have

$$
\begin{aligned}
\widehat{\beta} &= D^{-1}n^{-1}\begin{pmatrix} S_{n,11}R_{n,1}^*+S_{n,12}R_{n,2}^*+S_{n,13}R_{n,3}^* \\ S_{n,12}R_{n,1}^*+S_{n,22}R_{n,2}^*+S_{n,23}R_{n,3}^*+nD_n\beta_1 \\ S_{n,13}R_{n,1}^*+S_{n,23}R_{n,2}^*+S_{n,33}R_{n,3}^* \end{pmatrix} \\
&= D^{-1}n^{-1}\begin{pmatrix} S_{n,11}R_{n,1}^*+S_{n,12}R_{n,2}^*+S_{n,13}R_{n,3}^* \\ S_{n,12}R_{n,1}^*+S_{n,22}R_{n,2}^*+S_{n,23}R_{n,3}^* \\ S_{n,13}R_{n,1}^*+S_{n,23}R_{n,2}^*+S_{n,33}R_{n,3}^* \end{pmatrix} + \begin{pmatrix} 0 \\ \beta_1 \\ 0 \end{pmatrix} + (s.o.) \\
&= D^{-1}\begin{pmatrix} n^{1/2}C_1^*+o_p(n^{1/2}) \\ C_2^*+o_p(1) \\ C_3^*+o_p(1) \end{pmatrix} + \begin{pmatrix} 0 \\ \beta_1 \\ 0 \end{pmatrix} + (s.o.),
\end{aligned}
$$

where $C_1^* = B_1\sigma_a^2\omega_2(B_2+1)$, $C_2^* = B_3\omega_2(B_2+1)$, and $C_3^* = B_3\omega_2(B_1^2+1)$. Then, the analyses of the orders of $I_{n,2}$, $I_{n,3}$, and J_n are similar.

Case (I2): $\boldsymbol{\beta}_{1t}=\boldsymbol{\beta}_1, \boldsymbol{\beta}_{2t}=\boldsymbol{\beta}_2$ but $\boldsymbol{\beta}_{0t}\neq\boldsymbol{\beta}_0$. When β_{1t} and β_{2t} are constants, a simple calculation shows that

$$
\widehat{\beta} = D^{-1}n^{-1}\begin{pmatrix} S_{n,11}\sum_t\beta_{0t}+S_{n,12}\sum_t x_t\beta_{0t}+S_{n,13}\sum_t x_{t-1}\beta_{0t} \\ S_{n,12}\sum_t\beta_{0t}+S_{n,22}\sum_t x_t\beta_{0t}+S_{n,23}\sum_t x_{t-1}\beta_{0t}+nD_n\beta_{1t} \\ S_{n,13}\sum_t\beta_{0t}+S_{n,23}\sum_t x_t\beta_{0t}+S_{n,33}\sum_t x_{t-1}\beta_{0t}+nD_n\beta_{2t} \end{pmatrix} + (s.o.)
$$

$$= D^{-1}n^{-1}\begin{pmatrix} S_{n,11}\sum_t \beta_{0t} + S_{n,12}\sum_t x_t\beta_{0t} + S_{n,13}\sum_t x_{t-1}\beta_{0t} \\ S_{n,12}\sum_t \beta_{0t} + S_{n,22}\sum_t x_t\beta_{0t} + S_{n,23}\sum_t x_{t-1}\beta_{0t} \\ S_{n,13}\sum_t \beta_{0t} + S_{n,23}\sum_t x_t\beta_{0t} + S_{n,33}\sum_t x_{t-1}\beta_{0t} \end{pmatrix} + \begin{pmatrix} 0 \\ \beta_{1t} \\ \beta_{2t} \end{pmatrix} + (s.o.)$$

$$= D^{-1}\begin{pmatrix} C_1^{**} + o_p(1) \\ C_2^{**} + o_p(1) \\ C_3^{**} + o_p(1) \end{pmatrix} + \begin{pmatrix} 0 \\ \beta_{1t} \\ \beta_{2t} \end{pmatrix},$$

where $C_1^{**} = n^{1/2}B_1B_2\sigma_a^2\omega_0$, $C_2^{**} = n^{1/2}B_1\omega_0(B_2 - B_1^2)$, and $C_3^{**} = \sigma_a^2B_1^2\omega_0$. Applying these on the last two terms in (A.1), we obtain

$$\begin{aligned} I_{n,2} &= \frac{1}{n^3h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_t^T X_r X_t^T \left(\widehat{\beta} - \beta_t\right)\left(\widehat{\beta} - \beta_r\right)^T X_r K_{tr} \\ &= \frac{1}{n^3h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}(1 + u_tu_r + x_{t-1}x_{r-1})K_{tr} \\ &\quad \times\Big[\left(D^{-1}C_1^{**} - \beta_{0t}\right) + u_t\left(n^{-1/2}D^{-1}C_2^{**} - \beta_{1t}\right) \\ &\qquad + x_{t-1}\left(n^{-1/2}D^{-1}C_3^{**} - \beta_{3t}\right)\Big] \\ &\quad \times\Big[(D^{-1}C_1^{**} - \beta_{0r}) + u_r(n^{-1/2}D^{-1}C_2^{**} - \beta_{1r}) + x_{r-1}(n^{-1/2}D^{-1}C_3^{**} - \beta_{3r})\Big] \\ &\quad + (s.o.) \\ &= \frac{1}{n^2h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}\left[(D^{-1}C_1^{**} - \beta_{0t}) + u_tD^{-1}C_2^{**} + \frac{x_{t-1}}{\sqrt{n}}D^{-1}C_3^{**}\right] \\ &\quad \times\left[(D^{-1}C_1^{**} - \beta_{0r}) + u_tD^{-1}C_2^{**} + \frac{x_{t-1}}{\sqrt{n}}D^{-1}C_3^{**}\right] \\ &\quad \times\left(\frac{1}{n} + \frac{u_tu_r}{n} + \frac{x_{t-1}x_{r-1}}{n}\right)K_{tr} + (s.o.) \\ &= O_p(n), \end{aligned}$$

and

$$\begin{aligned} I_{n,3} &= \frac{2}{n^3h}\sum_{t=1}^{n}\sum_{r\neq t}^{n} X_t^T X_r v_r X_t^T \left(\widehat{\beta} - \beta_t\right) K_{tr} \\ &= \frac{2}{n^3h}\sum_{t=1}^{n}\sum_{r\neq t}^{n}(1 + u_tu_r + x_{t-1}x_{r-1})v_rK_{tr} \\ &\quad \times\Big[n^{1/2}\left(D^{-1}C_1^{**} - \beta_{0t}\right) + u_t\left(n^{1/2}D^{-1}C_2^{**} - \beta_{1t}\right) \\ &\qquad + x_{t-1}\left(n^{-1/2}D^{-1}C_3^{**} - \beta_{3t}\right)\Big] + (s.o.) \end{aligned}$$

$$
\begin{aligned}
&= \frac{2\sqrt{n}}{n^2 h} \sum_{t=1}^{n} \sum_{r \neq t}^{n} \left[\left(D^{-1} C_1^{**} - \beta_{0t} \right) + u_t D^{-1} C_2^{**} + \frac{x_{t-1}}{\sqrt{n}} D^{-1} C_3^{**} \right] \\
&\quad \times \left[\frac{1}{n} + \frac{u_t u_r}{n} + \frac{x_{t-1} x_{r-1}}{n} \right] v_r K_{tr} + (s.o.) \\
&= O_p(\sqrt{n}).
\end{aligned}
$$

Therefore, for this case, $I_{n,2}$ is the leading terms of I_n, $J_n = n\sqrt{h} I_n = O_p(n^2\sqrt{h})$, and it diverges to $+\infty$ at the rate of $n^2\sqrt{h}$. Theorem 2 is proved. ■