5-2021

Three essays on nonstationary financial econometrics

Yajie ZHANG

Singapore Management University

Three Essays on Nonstationary Financial Econometrics

YAJIE ZHANG

SINGAPORE MANAGEMENT UNIVERSITY

2021

THREE ESSAYS ON NONSTATIONARY FINANCIAL ECONOMETRICS

YAJIE ZHANG

A DISSERTATION

IN

ECONOMICS

Presented to the Singapore Management University in Partial Fulfillment

of the Requirements for the Degree of Doctor of Philosophy in Economics

2021

Supervisor of Dissertation

PhD in Economics, Programme Director

Three Essays on Nonstationary Financial Econometrics

by

Yajie Zhang

Submitted to School of Economics in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Jun Yu (Supervisor/Co-Chair)
Lee Kong Chian Professor of Economics and Finance
Singapore Management University

Peter C.B. Phillips (Co-Chair)
Sterling Professor of Economics and Statistics
Yale University
Distinguished Term Professor of Economics
Singapore Management University

Yichong Zhang
Assistant Professor of Economics
Lee Kong Chian Fellow
Singapore Management University

Shuping Shi
Professor of Economics
Macquarie University

Singapore Management University

2021

Abstract

This dissertation consists of three essays that contribute to the theory of nonstationary time-series analysis.

The first chapter explores the inference procedures for predictive regressions with time-varying characteristics. We extend the self-generated instrumentation, called IVX, to incorporate persistent regressors of functional local-to-unity, functional mildly explosive, and functional mildly stationary roots. The asymptotic distributions of IVX estimators under time-varying parameters are novel and nonpivotal but lead to pivotal distributions of the corresponding Wald statistics that are robust across various roots. The numerical experiments justify the robustness of IVX testing procedures in finite samples. We also verify the existence of time-varying coefficients and the predictability of fundamentals with such unstable parameters using the S&P 500 data.

The second chapter proposes a functional local-to-unity model with autoregressive coefficients that vary smoothly over time. Two sieve estimators, namely a time series and a panel autoregression estimators, are considered to estimate the local-to-unity function. The property of consistency is established. Besides, a consistent specification test to detect parameter instability is proposed. Numerical simulations demonstrate the finite sample performance of the specification test. Finally, we apply the panel estimator and specification test to the price index of China's real estate market and obtain significant empirical results in measuring time-varying growth rates in the data.

The third chapter discusses about time-varying predictive regressions, which are useful in the applications of empirical finance. The relevant theory in this area is mainly restricted to the case in which the model contains the local-to-unity (LUR) or locally stationary regressors only. It is not universal as the prevalent evidence indicates the existence of both time-varying predictability and the mixed-root phenomenon. We investigate a nonparametric predictive regression model with mixed-root regressors and time-varying coefficients, evolving smoothly over time. Further, we present a new variant of the self-generated instrument, called Sieve-IVX, which attains robust inference irrespective of various degrees of persistence. We establish its consistency and provide a Wald test to detect the temporary predictability of economic fundamentals. Numerical simulations show satisfactory finite-sample performances, which support our results.

Contents

Acknowledgement

First of all, I would like to express sincere gratitude to my supervisor, Professor Jun YU, for providing guidance and encouragement in my Ph.D. life. He is not only an outstanding scholar but also a devoted educator. I really appreciate his effort and patience, as he always gives instant feedback and valuable suggestions whenever I need. Professor Jun YU also maintains a effective team, where we can share research ideas and learn from each other.

Meanwhile, my profound gratitude goes to Professor Peter C.B. Phillips. As a leading econometrician in the world, Professor Peter C.B. Phillips spares no effort in helping young students. We all benefit a lot from his teaching and consultations. I have received a lot of constructive guidance during the past four years. His gentle and friendly attitude towards students is like a breeze in summer, which makes me feel refreshing.

Besides, I would like to thank Professor Yichong ZHANG for his kind help starting from my first year in Ph.D. program. With strong academic background, Professor ZHANG is willing to offer his opinion towards our research topics and give instructions based on the latest information. As a young scholar, he is like a friend, who cares about us and supports our choice of life.

I feel very thankful for the chance to join the Ph.D. program in Singapore Management University. In the past five years, I have received great training in research and teaching. I found my role models, Professor Jun YU and Peter C.B. Phillips and experienced different life in Singapore. I would like to express my special thanks to our program director, Professor Anthony TAY for his great contribution in making our Ph.D. journey an enjoyable process.

Last but not least, I want to give thanks to my family and friends, who support me strongly and create happiness in my life. My special thanks go to my friend and co-author, Professor Yanbo LIU for our joint works. My deepest thanks go to my parents, who love me unconditionally wherever I am. I appreciate the long-lasting friendship from my old friends, who always care for me even though there is a long distance between us in the past few years. Your gentle love makes me feel that I am never alone no matter in good times or bad times. And I feel grateful for my new friends in Singapore, who broaden my horizon and express positive vibes along this way. You make me realize that the world is much more than what I have seen in my personal life. In addition, I want to thank my aunt and cousins for their kind understanding and considerate advice. I am very blessed for having all of you in my life.

Chapter 1

Robust Inference with Functional Deviations from Unity in Predictive Regression

1.1 Introduction

The predictability of asset returns is one of the most debatable issues in empirical finance (see Campbell, 2008). The efficient market hypothesis supports the unpredictability of asset returns, whereas empirical evidence shows the predictive power of economic fundamentals for stock returns. Till now, there is little consensus on the topic of predictability, which leads to the stock return predictability puzzle (see Campbell and Thompson, 2008; Welch and Goyal, 2008; Rapach et al., 2010). To address this puzzle, the linear time-invariant predictive regression that relates stock returns to lagged fundamental variables has become a benchmark model to discuss predictability phenomena.

However, size distortions may arise when conventional inference procedures are applied to predictive regressions. These size distortions are due to endogeneity generated by the persistence in economic fundamentals. To eliminate size distortions, Campbell and Yogo (2006) propose a simulation-based likelihood ratio test (CY-Q test, hereafter) and construct its confidence interval by reverting the limit distribution. Nonetheless, the success of this procedure sacrifices the robustness for various roots and multiple regressors. Elliott et al. (2015) establish a nearly optimal test for predictive regressions when nuisance parameters exist in the probability limit. Similarly, this test relies on both numerical algorithms and a local-to-unity (LUR, hereafter) formulation. To address the abovementioned defects, Phillips and Magdalinos (2009, PM hereafter) propose a robust inference procedure called IVX and apply it to the mildly stationary root (MSR, hereafter) and LUR cases. They show the pivotal distribution under the null hypothesis of no predictability. Extensions of IVX are considered in Phillips and Lee (2013, 2016, PL hereafter) and Lee (2016).

Extensive studies on asset returns suggest evidence of time-varying dynamics due to changes in monetary policy, fluctuations in market sentiment, etc. (see Bossaerts and Hillion, 1999; Bekaert et al. 2007). Nevertheless, Pesaran and Timmerman (2002) highlight the need to develop econometric procedures for the model uncertainty issue of predictive regressions. A popular strategy to characterize model uncertainty is to use a predictive regression model with time-varying parameters. In this paper, we model the time-varying roots of economic fundamentals (e.g. persistent regressors) in terms of functional deviations from unity, including functional local-to-unit root (FLUR, hereafter), functional mildly explosive root (FMER, hereafter), and functional mildly stationary root (FMSR, hereafter). Bykhovskaya and Phillips (2018) propose the FLUR process and show that the power envelopes of standard unit root statistics are weakened with functional departure from unity. Bykhovskaya and Phillips (2020) employ FMER and FMSR models to discuss the boundary limit theory on the FLUR process. Parameter instability induced by the above time-varying roots presents a significant challenge to conventional inference procedures of predictive regressions since traditional methods, such as least-squares or simulated methods, are no longer applicable.

This study considers IVX inference procedures on predictive regressions with persistent regressors modeled by FLUR, FMER, and FMSR processes. We prove that the IVX estimators of predictive regressions are of asymptotic normality, and the corresponding Wald statistics follow a pivotal chi-square distribution. Monte Carlo simulations examine the finite sample performance of the IVX-based Wald test. In addition, we apply the IVX procedure to S&P 500 data and justify the predictability of several economic fundamentals.

The rest of this chapter is organized as follows. Section 2 presents the model setup and discusses the size distortions of conventional statistics. Section 3 shows the limit theory of IVX estimates and corresponding tests under FLUR, FMER, and FMSR cases. Section 5 reports simulation findings. Section 6 applies the IVX procedure to financial data sets. Section 6 concludes. The proofs and technical details are given in the appendix.

Throughout the paper, we use the following notation. For some arbitrary matrix M, we use $\|M\|_1$, $\|M\|_2$, $\|M\|_\infty$, and $\|M\|$ to denote the L^1, L^2, L^∞ and spectral norms. The symbol $\mathbb{E}_{t-1}(\cdot) := \mathbb{E}(\cdot|\mathcal{F}_{t-1})$ denotes conditional expectation with respect to the filtration $\mathcal{F}_{t-1}$. The symbol $=_d$ denotes equivalence in distribution. The notation $\Rightarrow$ signifies weak convergence in function space and Euclidean space according to context. The symbol $\xrightarrow{d}$ denotes convergence in distribution.

1.2 Model Setup and Size Distortion

This section defines the model setup and assumptions. We demonstrate the size distortion of t-statistics in predictive regression models with FLUR regressors. The existence of size distortions justifies the necessity for IVX.

1.2.1 Model setup

The standard predictive regression model is given as

$$y_t = \beta_0 + \beta_1' x_{t-1} + u_{0t}, \text{ with } \mathbb{E}(u_{0t}|\mathcal{F}_{t-1}) = 0, \tag{1.1}$$

where β_1 is an $n \times 1$ vector and $\mathcal{F}_{t-1}$ is a natural filtration. A vector of predictors x_{t-1} follows an autoregressive process with roots in a circle of unity, as follows:

$$x_t = R_{Tt} x_{t-1} + u_{xt},\ R_{Tt} = I_n + \frac{C(t/T)}{T^{\alpha}}, \tag{1.2}$$

where T is the sample size and $C(\cdot) := \text{diag}\{c_1(\cdot), c_2(\cdot), ..., c_n(\cdot)\}$. The pair of $(\alpha, C(\cdot))$ represents persistence in the multiple predictors of unknown functional forms. We allow for more general types of persistence: When $C(\cdot)$ is a time-varying function, functional autoregressive roots exist. We eliminate the case of unbounded $C(\cdot)$, since diverging distance parameters generate stationarity or explosiveness (see Phillips, 1987) and persistence disappears. Therefore, we impose regularity conditions on $C(\cdot)$ as the following assumption.

Assumption 1.1. *(i) For each $i = 1, 2, ..., n$, the distance parameter $c_i(r)$ is a deterministic function with bounds as*

$$0 < \inf_{1 \leq i \leq n} |c_i(r)| \leqslant |c_i(r)| \leqslant \sup_{1 \leq i \leq n} |c_i(r)| < +\infty,$$

where $r \in [0, \infty)$.

(ii) For each $i = 1, 2, ..., n$, we have

$$\sup_{1 \leq i \leq n} \int_0^{\infty} |c_i(r)|^2 \, dr < +\infty,$$

with $r \in [0, \infty)$.

In particular, x_t belongs to one of the following persistence categories:

(i) FMSR, if $\alpha \in (0, 1)$ and $C(r) < 0$ for each $r \in [0, \infty)$;

(ii) FLUR, if $\alpha = 1$;

(iii) FMER, if $\alpha \in (0, 1)$ and $C(r) > 0$ for each $r \in [0, \infty)$.

The innovation structure allows for a linear process with intertemporal dependence for u_{xt} and u_{0t}, and incorporates an assumption of conditionally homoskedastic martingale difference sequence (mds) for u_{0t}. Detailed assumptions on innovations are given below.

Assumption 1.2.

$$u_t = \begin{bmatrix} u_{0t} & (1 \times 1) \\ u_{xt} & (n \times 1) \end{bmatrix} = \sum_{j=0}^{\infty} F_j \epsilon_{t-j},$$

where

$$\begin{aligned} \epsilon_t &\sim mds(0, \Sigma),\ \mathbb{E}\left\|\epsilon_1\right\|^q < \infty \textit{ for some } q \geq 4, \\ F_0 &= I_{1+n}, \sum_{j=0}^{\infty} j \left\|F_j\right\| < \infty, F(1) = \sum_{j=0}^{\infty} F_j > 0, \\ F_j &= \begin{bmatrix} F_{0j} \\ F_{xj} \end{bmatrix}, F_{0j} = \begin{cases} [I_1 : 0_{1\times n}] & \textit{for } j = 0 \\ 0_{n\times(1+n)} & \textit{for } j \geq 1 \end{cases}. \end{aligned}$$

Based on Phillips and Solo (1992, PS hereafter), Beveridge–Nelson decomposition accommodates our model as,

$$u_t = F(1)\epsilon_t - \Delta\widetilde{\epsilon}_t,\ \widetilde{\epsilon}_t = \sum_{j=0}^{\infty} \widetilde{F}_j \epsilon_{t-j},\ \widetilde{F}_j = \sum_{s=j+1}^{\infty} F_s \text{ and } F(z) = \sum_{j=0}^{\infty} F_j z^j. \quad (1.3)$$

PS verify that $\sum_{i=0}^{\infty} i \left\|F(i)\right\| < \infty$ is a sufficient assumption for $\sum_{i=0}^{\infty} \left\|\widetilde{F}(i)\right\| < \infty$. Therefore, $\{F(1)\epsilon_t\}_{t=1}^{T}$ is the only leading term of $\{u_t\}_{t=1}^{T}$. We further denote the two-sided long-run covariance as

$$\Omega = \sum_{j=-\infty}^{\infty} \mathbb{E}(u_t u_{t-h}^{'}) = F\left(1\right) \Sigma F\left(1\right)',$$

where

$$F\left(1\right) = \begin{bmatrix} F_0\left(1\right) \\ F_x\left(1\right) \end{bmatrix} = \begin{bmatrix} [I_1 : 0_{1\times n}] \\ F_x\left(1\right) \end{bmatrix},$$

In addition,

$$\Omega = \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix},$$

where $\Omega_{00} = \mathbb{E}(u_{0t}^2)$, $\Omega_{0x} = \sum_{j=-\infty}^{\infty} \mathbb{E}(u_{0t} u_{x,t-h}^{'}) = \Sigma F_x^{'}(1)$, and $\Omega_{xx} = \sum_{j=-\infty}^{\infty} \mathbb{E}(u_{xt} u_{x,t-h}^{'}) = F_x(1)\Sigma F_x^{'}(1)$.

Based on (1.3), the functional central limit theorem applies in the following manner.

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{[Ts]} u_j = \frac{1}{\sqrt{T}}\sum_{j=1}^{[ns]} \begin{bmatrix} u_{0j} \\ u_{xj} \end{bmatrix} \Rightarrow \begin{bmatrix} B_0(s) \\ B_x(s) \end{bmatrix} = \text{BM} \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix},$$

where BM is a Brownian motion. Moreover, Bykhovskaya and Phillips (2018a) provide the following functional laws as

$$\frac{1}{\sqrt{T}} x_{[Tr]} \Rightarrow K_c(r) = \int_0^r \exp\left(\int_s^r C(k)dk\right) \Omega_{xx}^{\frac{1}{2}} dB_x(s),$$

where $K_c(r)$ follows the functional-coefficient Ornstein–Uhlenbeck process as

$$dK_c(r) = C(r)K_c(r)dr + \Omega_{xx}^{\frac{1}{2}} dB_x(r).$$

1.2.2 Size distortion

We demonstrate the size distortions of t-statistics when considering predictive regression models with FLUR regressors. By convention, the centered least squares estimator of $\beta := (\beta_0, \beta_1)$ is

$$\hat{\beta}^{OLS} - \beta = \left(\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1} \left(\sum_{t=1}^{T} X_{t-1}u_{0t}\right), \tag{1.4}$$

where $X_{t-1} := \left(1, x_{t-1}'\right)$ includes both the intercept and FLUR regressors in (1.2). To discuss size distortions, we decompose u_{0t} into two orthogonal components (Phillips, 2007): one is vertical to u_{xt}, while the other is proportional to u_{xt}. The decomposition is as follows,

$$dB_0(r) = dB_{0|x}(r) + \Omega_{0x}\Omega_{xx}^{-1} dB_x(r), \tag{1.5}$$

where $B_{0|x}(\cdot) =_d \text{BM}\left(0, \Omega_{0|x}\right)$. In the single-regressor case ($n = 1$), we have the following limit theory for the least squares estimator,

$$\begin{aligned} T(\widehat{\beta}_1 - \beta_1) &\Rightarrow \frac{\int_0^1 \overline{K}_c(r)dB_{0|x}(r)}{\int_0^1 \overline{K}_c^2(r)dr} \\ &= \frac{\int_0^1 \overline{K}_c(r)dB_{0|x}(r)}{\int_0^1 \overline{K}_c^2(r)dr} + \Omega_{0x}\Omega_{xx}^{-1}\frac{\int_0^1 \overline{K}_c(r)dB_x(r)}{\int_0^1 \overline{K}_c^2(r)dr}, \end{aligned}$$

where $\overline{K}_c(r) := K_c(r) - \int_0^1 K_c(r)dr$. When the estimated standard error is given as s.e.$(\widehat{\beta}_1) = \hat{\Omega}_{00}\left\{\sum_{t=1}^T (x_{t-1}^\mu)^2\right\}^{-\frac{1}{2}}$ with $x_{t-1}^\mu := x_{t-1} - \frac{1}{T}\sum_{t=1}^T x_{t-1}$, and the estimator $\hat{\Omega}_{00}$ represents any consistent estimate for Ω_{00}, the t-statistic under the null hypothesis $H_0 : \beta_1 = 0$ is justified as

$$\begin{aligned} t_{\widehat{\beta}_1 - \beta_1} = \frac{\widehat{\beta}_1 - \beta_1}{s.e.(\widehat{\beta}_1)} &\Rightarrow \left[1 - \frac{\Omega_{0|x}}{\Omega_{xx}\Omega_{00}}\right]\mathcal{N}(0,1) + \frac{\Omega_{0|x}}{\Omega_{xx}\Omega_{00}}\frac{\int_0^1 \overline{K}_c(r)dB_x(r)}{\int_0^1 \overline{K}_c^2(r)dr} \\ &= \left[1 - \lambda^2\right]^{\frac{1}{2}} Z + \lambda \cdot \eta(c), \end{aligned} \tag{1.6}$$

where $Z =_d \mathcal{N}(0,1)$, $\lambda := \Omega_{0|x}/\Omega_{xx}\Omega_{00}$ and $\eta(c) := \frac{\int_0^1 \overline{K}_c(r)dB_x(r)}{\int_0^1 \overline{K}_c^2(r)dr}$. The non-zero factor λ in (1.6) reveals the presence of size distortion in the test for predictability.

1.3 IVX Method

The main intuition of IVX is filtering x_t to generate instruments $\tilde{z}_t$ with MSR persistence. In this way, we can eliminate the asymptotic dependence between the numerator and denominator of IVX estimates. We define $\tilde{z}_t$ as

$$\widetilde{z}_t = R_{Tz}\widetilde{z}_{t-1} + \Delta x_t \text{ with } R_{Tz} = I_n + \frac{C_z}{T^\gamma},$$

where $\gamma \in (0,1)$, $C_z = c_z I_n$, $c_z < 0$ and $\widetilde{z}_0 = 0$. We choose the values of parameters γ and c_z. Consequently, the self-generated instrument is an inter-temporal summation of first-differenced regressors Δx_t as $\tilde{z}_t = \sum_{j=1}^{t-1} R_{Tz}^{t-j}\Delta x_j$. By (1.2), the decomposition of $\tilde{z}_t$ is

$$\begin{aligned} \tilde{z}_{t-1} &= \sum_{j=1}^{t-1} R_{Tz}^{t-j-1} u_{xj} + \frac{1}{T^\alpha}\sum_{j=1}^{t-1} R_{Tz}^{t-j-1} C\left(\frac{j+1}{T}\right) x_j \\ &= z_{t-1} + \frac{1}{T^\alpha}\eta_{T,t-1}^{(1)}, \end{aligned} \tag{1.7}$$

where the latent instrument $z_{t-1} := \sum_{j=1}^{t-1} R_{Tz}^{t-j-1} u_{xj}$, and IVX residual $\eta_{T,t-1}^{(1)} := \sum_{j=1}^{t-1} R_{Tz}^{t-j-1} C\left(\frac{j+1}{T}\right) x_j$.

As the instrument is constructed, we have the centered IVX estimator as follows

$$\hat{\beta}^{IVX} - \beta = \left(\sum_{t=1}^{T} \tilde{Z}'_{t-1} X_{t-1} \right)^{-1} \left(\sum_{t=1}^{T} \tilde{Z}'_{t-1} u_{0t} \right),$$

where $\tilde{Z}_{t-1} := \left(1, \tilde{z}'_{t-1}\right)$.

1.3.1 Functional local-to-unity regressors

For FLUR regressors where $\alpha = 1$, the IVX instrument (1.7) has the following decomposition,

$$\tilde{z}_{t-1} = z_{t-1} + \frac{1}{T^{\alpha}} \eta^{(1)}_{T,t-1}.$$

In the literature, PM proved that

$$\sup_{1 \leq t \leq T} \|z_{t-1}\| = O_p\left(T^{\frac{\gamma}{2}}\right), \tag{1.8}$$

while we show

$$\sup_{1 \leq t \leq T} \left\|\eta^{(1)}_{T,t-1}\right\| = O_p\left(T^{\frac{\gamma}{2}+1}\right). \tag{1.9}$$

Based on (1.8) and (1.9), we provide the asymptotic approximations to IVX estimates with FLUR regressors.

Lemma 1.1. *Let Assumptions 1.1 and 1.2 hold. As $T \to \infty$, the approximations to IVX estimates are given by*

$$(i) \quad \frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} \tilde{z}_{t-1} u_{0t} = \frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} z_{t-1} u_{0t} + o_p(1),$$

$$(ii) \quad \frac{1}{T^{1+\gamma}} \sum_{t=1}^{T} \tilde{z}_{t-1} \tilde{z}'_{t-1} = \frac{1}{T^{1+\gamma}} \sum_{t=1}^{T} z_{t-1} z'_{t-1} + o_p(1),$$

$$(iii) \quad \frac{1}{T^{1+\gamma}} \sum_{t=1}^{T} \tilde{z}_{t-1} x'_{t-1} = \frac{1}{T^{1+\gamma}} \sum_{t=1}^{T} z_{t-1} x'_{t-1} + \frac{1}{T^{2+\gamma}} \sum_{t=1}^{T} \eta^{(1)}_{T,t-1} x'_{t-1}.$$

Lemma 1.1 reveals the asymptotic behavior of sample moments in (i) and (iii). In the numerator, the term $\sum_{t=1}^{T} z_{t-1} u_{0t}$ with the latent instrument z_{t-1} dominates $\sum_{t=1}^{T} \eta^{(1)}_{T,t-1} u_{0t}$ of IVX residual $\eta^{(1)}_{T,t-1}$. For the denominator, both $\sum_{t=1}^{T} z_{t-1} x'_{t-1}$ and $\sum_{t=1}^{T} \eta^{(1)}_{T,t-1} x'_{t-1}$ affect the probability limit. Although the abovementioned approximation of FLUR regressors coincides with the LUR case, our IVX estimate of interest still demonstrates novel limit distributions. We collect asymptotic results of the IVX estimate in the following theorem.

Theorem 1.1. *Let Assumptions 1.1 and 1.2 hold. As $T \to \infty$, the limit behaviors of the IVX estimate is given by*

$$(i) \quad \frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} \tilde{z}_{t-1} u_{0t} \xrightarrow{d} \mathcal{N}\left(0, \int_0^{+\infty} e^{rC_z} \Omega_{xx} e^{rC_z} dr\right),$$

$$(ii) \quad \frac{1}{T^{1+\gamma}} \sum_{t=1}^{T} \tilde{z}_{t-1} x'_{t-1}$$

$$\Rightarrow \quad -\left(\int_0^1 dB_x(r) K'_c(r) + \int_0^1 C(r) K_c(r) K'_c(r) dr + \Omega_{xx}\right) C_z^{-1} =: \Phi,$$

$$(iii) \quad T^{\frac{1+\gamma}{2}}\left(\hat{\beta} - \beta\right) \Rightarrow \mathcal{MN}\left(0, \Phi^{-1} \int_0^{+\infty} e^{rC_z} \Omega_{xx} e^{rC_z} dr \Phi'^{-1}\right).$$

Compared with PM, the significant difference is that the functional-coefficient process $K_c(r)$ rather than Ornstein–Uhlenbeck process $J_c(r)$ determines asymptotic behaviors. However, the limiting distribution of the IVX estimate still follows mixed normality and lays down the foundation for pivotal tests. Note that the mixed normality depends on the asymptotic independence between two Gaussian variables, $\sum_{t=1}^{T} \tilde{z}_{t-1} u_{0t}$ and $\sum_{t=1}^{T} \tilde{z}_{t-1} x'_{t-1}$, and the order reductions of latent instrument $\tilde{z}_t$. Since the scale parameter $\gamma \in (0,1)$, the mds sequences $\left\{T^{-\frac{\gamma+1}{2}} z_{t-1}\epsilon_t,\ T^{-\frac{1}{2}} F_x(1)\epsilon_t\right\}_{t=1}^{T}$ are asymptotically independent. Moreover, we can show that the Wald statistic follows a standard χ^2 distribution under the null hypothesis of no predictability.

Theorem 1.2. *Let Assumptions 1.1 and 1.2 hold. Under $H_0 : H\beta = h$, as $T \to \infty$,*

$$W_T := \left(H\hat{\beta} - h\right)' \left\{H\left[\left(X'P_{\tilde{z}}X\right)^{-1} \hat{\Omega}_{00}\right] H'\right\}^{-1} \left(H\hat{\beta} - h\right) \xrightarrow{d} \chi^2(n),$$

where $(X'P_{\tilde{z}}X) := \left[\left(\sum_{t=1}^{T} x_{t-1}\tilde{z}'_{t-1}\right)\left(\sum_{t=1}^{T} \tilde{z}_{t-1}\tilde{z}'_{t-1}\right)^{-1}\left(\sum_{t=1}^{T} x_{t-1}\tilde{z}'_{t-1}\right)'\right]$.

As u_{0t} is a mds sequence, a simple consistent estimator for Ω_{00} is $\hat{\Omega}_{00}$ $\left(:= \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_{0t}^2\right)$ where $\widehat{u}_{0t} = y_t - \hat{\beta}x_{t-1}$ and $\hat{\beta}$ is an IVX estimate. For a cointegrating system with serial correlations, a heteroskedasticity and autocorrelation consistent (HAC) estimator (Andrews, 1991b) can consistently estimate Ω_{00}.

According to the asymptotic results, IVX procedures have obvious advantages. Unlike Campbell and Yogo (2006), our IVX statistic is robust across various model formulations as the W_T test for FLUR regressors is identical to the case of LUR regressors (PM 2009). Moreover, we do not need to simulate critical values as the pivotal distribution χ^2 is free of nuisance parameters. These features demonstrate the flexibility of IVX statistics.

1.3.2 Functional mildly explosive regressors

Different from the predictive regression model with FLUR regressors, the asymptotic theory for FMER regressors is more complicated. Although the asymptotic approximations of the FMER case are consistent with PL, the limit theory is novel as provided for the first time. In the FMER case, the remainder term $\frac{1}{T^\alpha}\eta_{T,t-1}^{(1)}$ contains explosive roots of exponential rates and dominates the latent instrument z_{t-1} in the IVX estimator. To consistently approximate sample moments, the pointwise convergence is insufficient, as our results should accommodate a wide range of α. Instead, the Skorokhod embedding theorem is employed to establish the uniform convergence.

To simplify our discussion, we define an intermediate argument, $\tilde{x}_t$, as,

$$\tilde{x}_t := \sum_{j=1}^{t}\left[\exp\left(-\frac{1}{T^\alpha}\sum_{l=t-j+1}^{t} C\left(\frac{t-l+1}{T}\right)\right)\right] F_x(1)\epsilon_j.$$

With this auxiliary term, we obtain the asymptotic behavior of FMER regressors, x_t.

Lemma 1.2 (Pointwise Approximations)**.** *Let Assumptions 1.1 and 1.2 hold. Define $\bar{x}_{k_T} = \frac{1}{T^{\alpha/2}}\tilde{x}_{k_T}$ with $k_T = [Tr]$. When $\frac{T^\alpha}{k_T} + \frac{k_T}{T} \to 0$, the limit process of x_t is given by*

$$\bar{x}_{k_T} \xrightarrow{d} X_{C(0)} =_d \mathcal{N}\left(0, \int_0^{+\infty} e^{-C(1)p}\Omega_{xx}e^{-C(1)p}dp\right).$$

The pointwise limiting theory for any $t \in \{1, 2, ..., T\}$ is insufficient for approximating the sample moments over a wide range of α. Therefore, a uniform approximation to the stabilized FMER process is indispensable. We illustrate the formal statement of the Skorokhod embedding theorem in the following lemma.

Lemma 1.3 (Uniform Approximations)**.** *Let rate restriction $\alpha q > 2$ hold. With the same k_T defined in Lemma 1.2, there exists a suitably expanded probability space, such that*

$$\sup_{k_T \leq j-1 \leq T} \left\| \frac{\hat{\tilde{X}}_{j-1}}{T^{\alpha/2}} - X_{C(0)} \right\| = o_{a.s.}(1),$$

where $\hat{\tilde{X}}_{j-1} := \exp\left[-\frac{1}{T^\alpha}\sum_{i=1}^{j-1} C\left(\frac{j}{T}\right)\right] x_{j-1}$, *and* $X_{C(0)} =_d \mathcal{N}\left(0, \int_0^{+\infty} e^{-C(1)p}\Omega_{xx}e^{-C(1)p}dp\right)$.

Condition, $\alpha q > 2$, ensures the applicability of the Skorokhod embedding theorem for FMER regressors. Besides, the IVX residual $\frac{1}{T^\alpha}\eta_{T,t}^{(1)}$ also contains explosive roots and dominates the latent instrument z_{t-1}. Therefore, a similar uniform approximation applies to $\frac{1}{T^\alpha}\eta_{T,t}^{(1)}$.

Lemma 1.4. *Let Assumptions 1.1 and 1.2 hold. Assume that k_T satisfies the following conditions: (i) $\frac{T^\alpha}{k_T} + \frac{T^\gamma}{k_T} \to 0$; (ii) $\frac{T-k_T}{T^\gamma} + \frac{T-k_T}{T^\alpha} \to 0$. For all $t \in [k_T, T]$, we have*

$$\frac{1}{T^{\frac{\alpha}{2}+(\alpha\wedge\gamma)}} \exp\left(-\frac{1}{T^\alpha}\sum_{i=1}^{t} C\left(\frac{i}{T}\right)\right)\eta_{T,t}^{(1)} = C_{\alpha\gamma}X_{C(0)} + o_p(1),$$

where

$$C_{\alpha\gamma} := \begin{cases} -C_z^{-1} & \text{if } \alpha > \gamma, \\ C(0)^{-1} & \text{if } \alpha < \gamma, \\ [C(1) - C_z]^{-1} & \text{if } \alpha = \gamma. \end{cases}$$

The adjustment rate of IVX residuals $\eta_{T,t}^{(1)}$ is exponential and involves parameters α and γ from MSR and FMER autoregressive coefficients. This exponential adjustment rate ensures the asymptotic dominance of IVX residuals over the latent instrument. Combining the asymptotics of $\frac{1}{T^\alpha}\eta_{T,t}^{(1)}$, and x_t, we show the asymptotic theory of the IVX estimate.

Lemma 1.5. *Let Assumptions 1.1 and 1.2 hold.*

(i) For the IVX numerator, as $T \to \infty$*,*

$$\frac{1}{T^{(\alpha\wedge\gamma)}}\sum_{j=1}^{T} u_{0t}\tilde{z}_{t-1}' \exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T} C\left(\frac{j}{T}\right)\right] \Rightarrow \mathcal{MN}\left(0, V\right),$$

where $V := \int_0^{+\infty} e^{-pC(1)} C_{\alpha\gamma} X_{C(0)} X_{C(0)}' C_{\alpha\gamma} e^{-pC(1)} dp \cdot \Omega_{00}$.

(ii) For the IVX denominator, as $T \to \infty$*,*

$$\frac{1}{T^{\alpha+(\alpha\wedge\gamma)}}\sum_{t=1}^{T} \exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T} C\left(\frac{j}{T}\right)\right] \tilde{z}_{t-1}\tilde{x}_{t-1}' \exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T} C\left(\frac{j}{T}\right)\right] \Rightarrow \Phi,$$

where $\Phi := \int_0^{+\infty} e^{-pC(1)} X_{C(0)} X_{C(0)}' e^{-pC(1)} dp \cdot C_{\alpha\gamma}$.

(iii) For the IVX estimator, as $T \to \infty$*,*

$$T^{\alpha} \exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T} C\left(\frac{j}{T}\right)\right]\left(\hat{\beta} - \beta\right) \Rightarrow \mathcal{MN}\left(0, \Phi^{-1} V \Phi'^{-1}\right).$$

Lemma 1.5 (i) verifies that $\sum_{t=1}^{T} z_{t-1}u_{0t}$ vanishes in the probability limit, whereas $\frac{1}{T^{\alpha}}\sum_{t=1}^{T} \eta_{T,t-1}^{(1)} u_{0t}$ contributes to the mixed normality due to the exponential rate of IVX residuals. In Lemma 1.5 (ii), $\frac{1}{T^{\alpha}}\sum_{t=1}^{T} \eta_{T,t-1}^{(1)} x_{t-1}'$ dominates $\sum_{t=1}^{T} z_{t-1}x_{t-1}'$ in the denominator. The limiting distribution for the IVX estimator with FMER regressors follows mixed normality, since the self-generated instrument eliminates the asymptotic dependence between numerator $\sum_{t=1}^{T} \tilde{z}_{t-1}u_{0t}$ and denominator $\sum_{t=1}^{T} \tilde{z}_{t-1}x_{t-1}'$. Based on the asymptotic normal distribution of the IVX estimator, the IVX-based Wald test is χ^2-distributed.

Theorem 1.3. *Let Assumptions 1.1 and 1.2 hold. As* $T \to \infty$*, under* $H_0 : H\beta = h$*,*

$$W_T = \left(H\hat{\beta} - h\right)' \left\{H\left[\left(X'P_{\tilde{z}}X\right)^{-1}\hat{\Omega}_{00}\right]H'\right\}^{-1}\left(H\hat{\beta} - h\right) \xrightarrow{d} \chi^2\left(n\right),$$

where

$$\left(X'P_{\tilde{z}}X\right)^{-1} := \left[\left(\sum_{t=1}^{T} x_{t-1}\tilde{z}_{t-1}'\right)\left(\sum_{t=1}^{T} \tilde{z}_{t-1}\tilde{z}_{t-1}'\right)^{-1}\left(\sum_{t=1}^{T} x_{t-1}\tilde{z}_{t-1}'\right)'\right]^{-1},$$

$\hat{\Omega}_{00} := \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{0t}^{2}$ *with* $\widehat{u}_{0t} := y_t - \hat{\beta}x_{t-1}$ *and* $\hat{\beta}$ *is an IVX estimator.*

An essential feature of the IVX instrumentation is that the complexity in the limiting distributions does not affect the pivotal distribution of the Wald test. It relies on the self-normalizing property in the test statistics. Combining our results with those of PL, we can claim that regardless of whether the parameters of the mildly explosive regressors are time-invariant or not, the unified framework of IVX always generates pivotal test statistics. The robustness of the IVX limit theory benefits empirical work in which there is inevitable uncertainty about the degree of persistence and parameter instability.

1.3.3 Functional mildly stationary regressors

When $c_i(\cdot) < 0$ for each $i \in \{1, 2, 3, \ldots, n\}$, the persistent regressor x_{t-1} belongs to the category of the FMSR process. The discussion for FMSR regressors comprises two scenarios: (i) $\gamma < \min\{\alpha, 1\}$; (ii) $\gamma \geq \alpha$. If $\gamma < \alpha$, we can apply the same derivations as in the FLUR case. The main results are summarized in the following lemma without proof.

Lemma 1.6. *Let Assumptions 1.1 and 1.2 hold. As* $T \to \infty$,

$$\sup_{1\leq t\leq T} \mathbb{E}\left\|\eta_{T,t-1}^{(1)}\right\|^2 = O_p\left(T^{(\alpha\vee\gamma)+2(\alpha\wedge\gamma)}\right).$$

If $\gamma < \alpha$, the term related to latent instrument z_{t-1} dominates in the IVX numerator. However, both the latent instrument and IVX residual affect the limit behavior of the denominator. The abovementioned results establish foundations for the asymptotic analysis when $\gamma < \alpha$.

Theorem 1.4. *Let Assumptions 1.1 and 1.2 hold. When* $\gamma < \alpha$, *as* $T \to \infty$,

(i) $\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\tilde{z}_{t-1}u_{0t} \xrightarrow{d} \mathcal{N}\left(0, \int_0^{+\infty} e^{rC_z}\Omega_{xx}e^{rC_z}dr\right)$,

(ii) $\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\tilde{z}_{t-1}x_{t-1}' \Rightarrow -C_z^{-1}\Omega_{xx} =: \Psi$,

(iii) $T^{\frac{1+\gamma}{2}}\left(\hat{\beta}-\beta\right) \Rightarrow \mathcal{N}\left(0, \Psi^{-1}\int_0^{+\infty} e^{rC_z}\Omega_{xx}e^{rC_z}dr\left(\Psi^{-1}\right)'\right)$.

When $\gamma \geq \alpha$, by summations by parts,

$$\begin{aligned} \tilde{z}_t &= \sum_{j=1}^{t} R_{Tz}^{t-j} \Delta x_j = x_t - R_{Tz}^t x_0 - \sum_{j=1}^{t} \left(\Delta R_{Tz}\right)^{t-j} x_{j-1} \\ &= x_t - R_{Tz}^t x_0 + \frac{C_z}{T^\gamma} \eta_{Tt}^{(1)}. \end{aligned}$$

The intuition for this step is that when $\gamma \geq \alpha$, x_{t-1} is more persistent than z_{t-1}. The persistent regressor x_{t-1} replaces z_{t-1} and becomes the new latent instrument. Under this case, the endogeneity still vanishes asymptotically as the persistence disappears. The following lemma summarizes the main results of asymptotic approximations when $\alpha \in (0, \gamma)$.

Lemma 1.7. *Let Assumptions 1.1 and 1.2 hold. When $\alpha \in (0, \gamma)$, as $T \to \infty$,*

$$\begin{aligned} &(i) \quad \frac{1}{T^{\frac{1+\alpha}{2}}} \sum_{t=1}^{T} \tilde{z}_{t-1} u_{0t} = \frac{1}{T^{\frac{1+\alpha}{2}}} \sum_{t=1}^{T} x_{t-1} u_{0t} + o_p(1), \\ &(ii) \quad \frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} \tilde{z}_{t-1} x_{t-1}' = \frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} x_{t-1} x_{t-1}' + o_p(1), \\ &(iii) \quad \frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} \tilde{z}_{t-1} \tilde{z}_{t-1}' = \frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} x_{t-1} x_{t-1}' + o_p(1), \end{aligned}$$

where both $\frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} \tilde{z}_{t-1} x_{t-1}'$ and $\frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} \tilde{z}_{t-1} \tilde{z}_{t-1}'$ converge to $V_{xx} := \int_0^1 \left[\int_0^\infty e^{zC(r)} \Omega_{xx} e^{zC(r)} dz\right] dr$.

When $\alpha < \gamma$, the term associated with the latent instrument x_t dominates the one of IVX residual $\frac{1}{T^\alpha} \eta_{T,t}^{(1)}$. Nonetheless, the latent instrument x_{t-1} in the denominator dominates the IVX residual when $\alpha < \gamma$, different from the discussions when $\alpha \geq \gamma$. Based on these approximations, we provide the asymptotic theory for the IVX estimate.

Theorem 1.5. *Let Assumptions 1.1 and 1.2 hold. When $\alpha \leq \gamma$, as $T \to \infty$,*

$$\begin{aligned} &(i) \quad \frac{1}{T^{\frac{1+\alpha}{2}}} \sum_{t=1}^{T} \tilde{z}_{t-1} u_{0t} \xrightarrow{d} \mathcal{N}(0, V_{xx}), \\ &(ii) \quad \frac{1}{T^{1+\alpha}} \sum_{t=1}^{T} \tilde{z}_{t-1} x_{t-1}' \Rightarrow V_{xx}, \\ &(iii) \quad T^{\frac{1+\alpha}{2}} \left(\hat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, V_{xx}^{-1}\right). \end{aligned}$$

We check that when $\alpha < \gamma$, the limiting distribution of the IVX estimate is a Gaussian process and contributes to the pivotal Wald statistic without any modifications. Combining the results of $\alpha \leq \gamma$ and $\alpha > \gamma$, we have the following theorem.

Theorem 1.6. *Let Assumptions 1.1 and 1.2 hold. As $T \to \infty$, under $H_0 : H\beta = h$,*

$$W_T = \left(H\hat{\beta} - h\right)' \left\{H\left[\left(X'P_{\tilde{z}}X\right)^{-1}\hat{\Omega}_{00}\right]H'\right\}^{-1}\left(H\hat{\beta} - h\right) \xrightarrow{d} \chi^2(n),$$

where

$$\left(X'P_{\tilde{z}}X\right)^{-1} := \left[\left(\sum_{t=1}^{T} x_{t-1}\tilde{z}'_{t-1}\right)\left(\sum_{t=1}^{T} \tilde{z}_{t-1}\tilde{z}'_{t-1}\right)^{-1}\left(\sum_{t=1}^{T} x_{t-1}\tilde{z}'_{t-1}\right)'\right]^{-1},$$

$\hat{\Omega}_{00} := \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{0t}^2$ *with* $\widehat{u}_{0t} := y_t - \hat{\beta}x_{t-1}$ *and* $\hat{\beta}$ *is an IVX estimator.*

Kostakis et al. (2014) discuss the predictability of the economic fundamentals of time-invariant coefficients. They also prove the pivotal distribution of the IVX test under the null hypothesis of no predictive phenomenon. Even if parameter instability exists, our testing statistics remain robust with a χ^2 distribution. This property of robustness confirms the broad applicability of IVX procedure. Moreover, the IVX procedure provides a unified approach for verifying predictability: The W_T statistic can identify the predictability of persistent regressors of various persistence and unstable parameters, including FLUR, FMER, FMSR processes.

1.4 Monte Carlo Simulation

In this section, we conduct numerical simulations to evaluate the finite sample performance of IVX test with FLUR regressors. The data generating process is as follows:

$$\begin{aligned} y_t &= \beta_1' x_{t-1} + u_{0t}, \\ x_t &= R_{Tt} x_{t-1} + u_{xt}, \end{aligned} \tag{1.10}$$

where $R_{Tt} = I_n + \frac{C(t/T)}{T}$ and

$$u_t = \begin{pmatrix} u_{0t} \\ u_{xt} \end{pmatrix} \sim iid\ \left(0_{(1+n)}, \Sigma_{(1+n)\times(1+n)}\right). \tag{1.11}$$

We consider both normal and heavy-tailed innovations. In the single-regressor ($n = 1$) case, the covariance matrix of innovations is parameterized as

$$\Sigma = \begin{pmatrix} 1 & -0.95 \\ -0.95 & 1 \end{pmatrix}. \tag{1.12}$$

The negative value of covariance between u_{0t} and u_{xt} is consistent with the empirical findings in the financial market.

The instruments are constructed as in (1.7). We set $C_z = -5$ and $\gamma = 0.9$ to observe the size and local power of the Wald statistics in various forms of functional persistence. We set the sample size T as 250 and the number of simulation paths as 2,500.

To investigate the performance of IVX method in terms of size and local power, we adopt a sequence of Pitman's local alternatives as $H_{\beta_{1n}} : \beta_{1n} = \frac{b}{T}$ with integer values of $b \in [0, 20]$. We evaluate the empirical size by the frequency of rejections under $H_0 : \beta_1 = 0$ ($b = 0$ equivalently). The empirical power is the frequency of rejections under $H_1 : \beta_1 \neq 0$ ($b \neq 0$ equivalently). The nominal size is 0.05. Moreover, we employ the Bonferroni Q-test (CY-Q test, hereafter) and power-enhanced Q-test (modified CY-Q test, hereafter) of Campbell and Yogo (2006) as benchmarks for efficiency comparisons.

We provide simulated results in Figures 1.1-1.4. Figures 1.1 and 1.2 give the size and power performance of IVX inferences, when $C(t/T)$ is piecewise constant. We consider cases in which regressors switch between different nonstationary regimes. Figure 1.1 shows the simulation results when $\alpha = 1$ (LUR). The IVX method outperforms simulation-based tests when regressors change from a stationary regime to an explosive regime and from a unit-root region to an explosive region. When structural breaks occur between stationary and unit root regimes, IVX slightly underperforms in terms of power. Figure 1.2 shows the superiority of IVX inferences when $\alpha = 0.5$ (mildly integrated or mildly explosive cases). The test based on IVX method performs better than the CY-Q and modified CY-Q test statistics when there are structural breaks in the distance parameters of LUR regressors.

To verify the robustness of the IVX procedures, Figures 1.3 and 1.4 present the performance of the Wald statistics when $C(t/T)$ are trigonometric functions. In Figure 1.3 ($\alpha = 1$), the IVX method performs better in both size and power when $C(t/T)$ takes the form of $5\sqrt{t/T}$, $15\sin(t/T)$, $15\cos(t/T)$, and $-15\cos(t/T)$. The performance of IVX is nearly identical to CY-Q when $C(t/T)$ is $-15\sin(t/T)$. When $C(t/T)$ belongs to the stationary domain as $-5\sqrt{t/T}$, IVX slightly underperforms in terms of power. In Figure 1.4 ($\alpha = 0.5$), IVX performs better in terms of size and power throughout all functional forms as $5\sqrt{t/T}$, $-5\sqrt{t/T}$, $15\sin(t/T)$, $-15\sin(t/T)$, $15\cos(t/T)$, and $-15\cos(t/T)$.

Since the IVX inference procedure outperforms in most of the cases, numerical simulations verify the applicability of IVX procedures to the case of parameter instability. Besides, the simulated results show the robustness of the IVX procedure. We can conclude that the self-generated instrument is friendly to empirical studies when there is uncertainty about the functional forms or degrees of persistence in the regressors.

1.5 Empirical Illustrations

The predictability puzzle on stock return has been a controversial topic for years. In conventional prediction models, the coefficients of persistent regressors are time-invariant, as

$$x_t = R_T x_{t-1} + u_{xt} \text{ with } R_T = I_n + \frac{C}{T},$$

where C is a constant matrix. In this study, we generalize C to be a time-varying function. To detect the predictability of economic fundamentals on stock returns, we employ the updated dataset of Welch and Goyal (2008). In the empirical practice, the full sample period is from January 1927 to December 2017.

The persistent regressors belong to the following two categories:

1) Characteristics of stocks: dividend–price ratio (*d/p*), earnings–price ratio (*e/p*), book-to-market ratio (*b/m*), stock variance (*svar*), dividend–payout ratio (*d/e*), and net equity expansion(*ntis*);

2) Interest-related variables: 3-month treasury bill rate (*tbl*), term spread(*tms*), long-term government bond return (*ltr*), default yield spread (*dfy*), default return spread (*dfr*), and inflation (*infl*).

The empirical analysis has two parts. First, we estimate the slope in (1.2) by kernel estimation and demonstrate the existence of parameter instability (see Appendix A). Second, we use the IVX procedure to show the predictability of the selected economic fundamentals.

1.5.1 Kernel estimation

To accommodate time-varying coefficients, we conduct kernel estimations on potential regressors and plot the estimated paths of slopes. The asymptotic theory of kernel estimation is given in Appendix A; see also recent works by Phillips et al. (2017), Li et al. (2016), and Li et al. (2020) on cointegrating regression with functional coefficients. Under the framework of FLUR, the distance parameter is not consistently estimated, and the standard testing procedures fail for both the distance parameter and the slope.

Figure 1.5 plots the estimated paths of slopes and displays the unstable slopes in regressors, such as *svar, ntis, tms, dfr,* and *infl.* We find that variables such as *d/e* and *tbl* show time-varying features in specific sub-periods. There are also slight fluctuations in *b/m* and *dfy*, but they are not as significant as those of variables mentioned before. The kernel estimation results verify the existence of parameter instability in the potential predictors.

1.5.2 IVX inference

We employ IVX inferences to test the predictability of economic variables on the S&P 500 index return. Based on the univariate predictive regression model, we provide the estimated values of β_1 and corresponding IVX-based t-statistics in Table 1.1. For the sample period from January 1927 to December 2017, the variable *b/m,* with time-varying slope, shows significant predictive power on the asset return. Since Campbell and Yogo (2006) and Kostakis et al. (2012) claim that the predictability for stock return appears to be weaker for post-1952 data, we test a subsample period from January 1952 to December 2017. Within the subsample period, there are more predictors with statistical significance, which are *svar*, *tbl,* and *tms*. From kernel estimates, we know that the slopes of *svar* and *tms* are time-varying. If we apply simulation-based methods on these variables with time-varying roots, we are likely to obtain misleading results.

Furthermore, we discuss the predictability of multivariate regressors based on the results of the univariate case. We combine the predictors that have predictive power under the criteria of IVX statistics. To avoid strong correlations between the predictors, we choose regressors from different categories, characteristics of stock, or interest-related variables. In Table 1.2, we provide the empirical results for both the full sample and subsample periods. The results show that the combination of *b/m* and *tbl* provides predictive powers for both the full sample and subsample. Other combinations, such as (*svar*, *tbl*), and (*svar, tms*), also can predict S&P 500 returns significantly in the post-1952 periods.

The model setup in this study helps to deal with predictors with time-varying parameters. We find that some of the widely used fundamentals, such as *b/m* and *tms*, have predictive power on stock returns. The empirical study also shows that the IVX-based inference procedure is a robust tool to verify the predictability of regressors with potential parameter instability.

1.6 Conclusion

This study shows that the IVX method developed in PM is robust under functional-deviated regressors characterized as FLUR, FMER, and FMSR. An essential feature of IVX is that it can apply to regressors with different persistence, including the case of deterministic departure from unity (PM and PL), and the case of time-varying roots discussed in this paper. Unlike the methods based on numerical simulations, we do not need to identify the type of persistence and the existence of time-varying roots. These advantages offer substantial convenience to empirical studies in macroeconomics and finance.

According to the simulation results, the method from Campbell and Yogo (2006) tends to yield a slight size distortion in the local-to-unity case. The existence of size distortions means that we still reject the true model with a high frequency. By comparison, the IVX method appears to control size distortions better and accommodates more general types of regressors. The empirical results based on IVX estimation verify the predictability of some economic fundamentals on stock returns. We also specify the time-varying patterns in these potential predictors by consistent kernel estimations. This application demonstrates the usefulness of our study and confirms the robustness of IVX instruments.

A future research direction is to increase the local power of IVX estimation in this work. According to the Neyman–Pearson lemma, the likelihood-ratio test is the most powerful. The simulation-based test proposed by Campbell and Yogo (2006) is a type of likelihood ratio test which demonstrates desirable power behavior. Comparing the power function of the IVX instrument with the likelihood ratio test of stationary regressors, we observe slight underperformance of IVX in the finite sample. It is possibly unrealistic for IVX to be the best method in every perspective since there seems to be a trade-off between robustness and efficiency. However, the phenomenon illustrates a possible direction for enhancing the IVX inference procedure.

Tables and Figures

Figure 1.1: Size and power performance of IVX, CY-Q, and Modified CY-Q tests with discrete c(t/n) when alpha=1

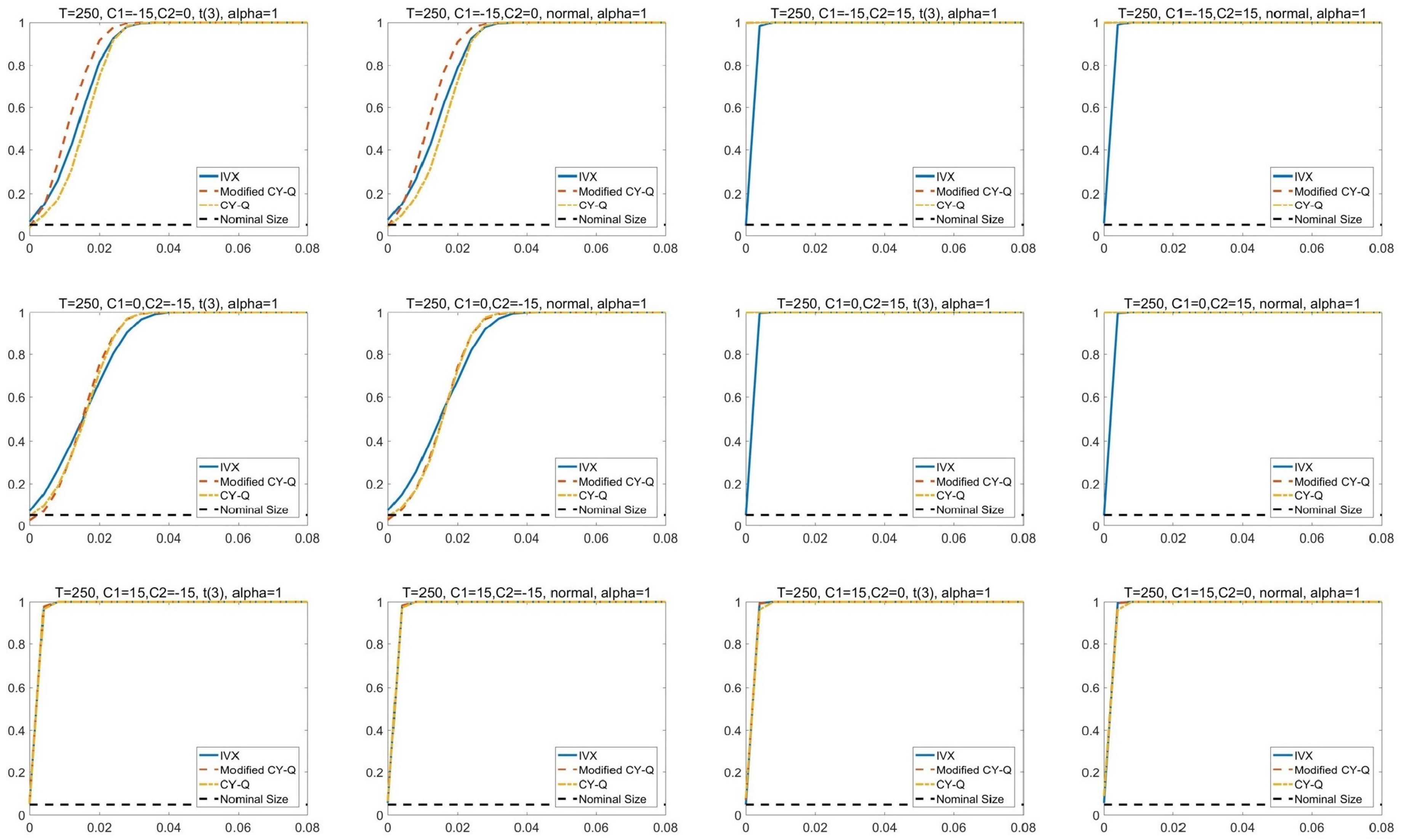

Figure 1.2: Size and power performance of IVX, CY-Q, and Modified CY-Q tests with discrete c(t/n) when alpha=0.5

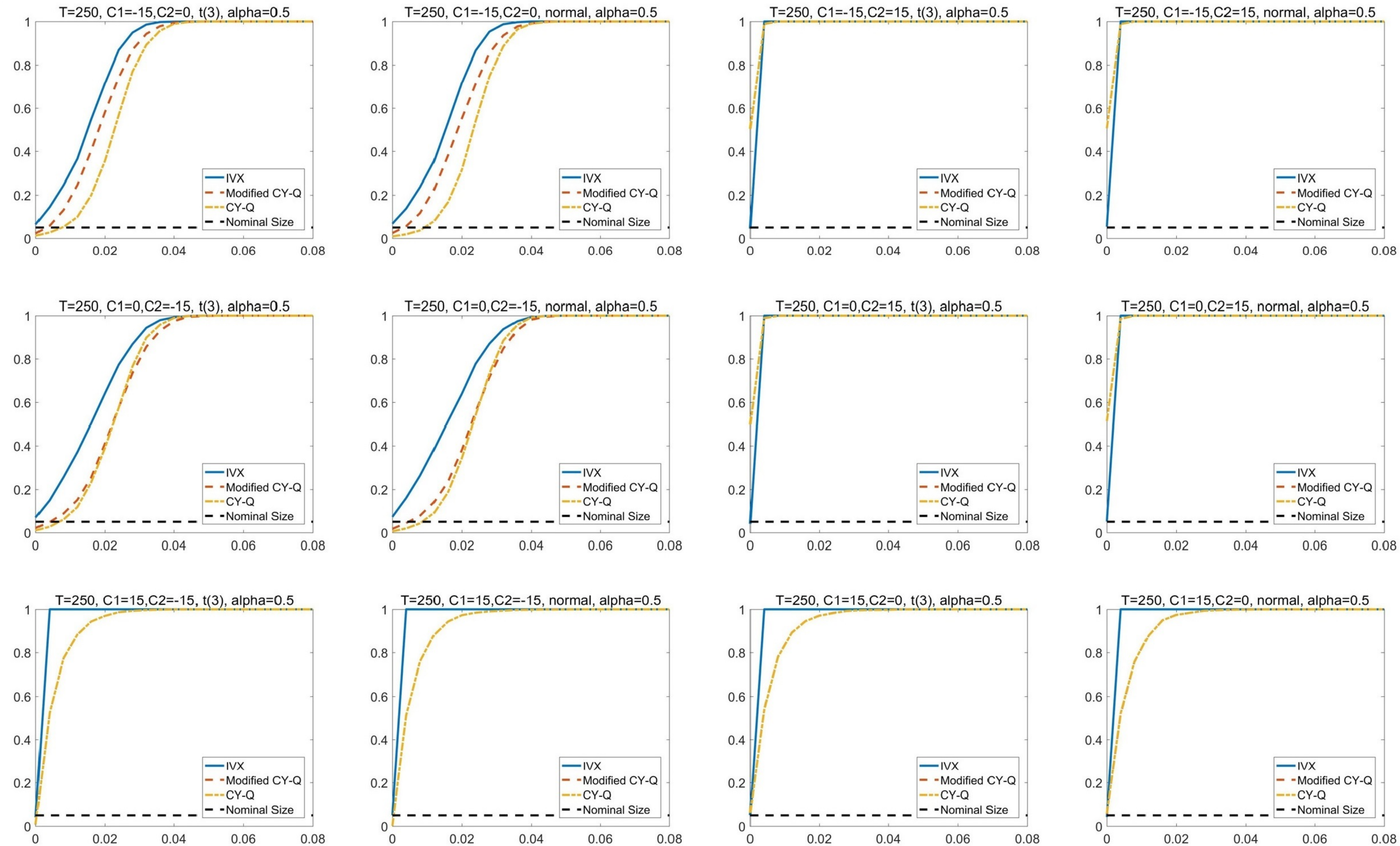

Figure 1.3: Size and power performance of IVX, CY-Q, and Modified CY-Q tests with continuous c(t/n) when alpha=1

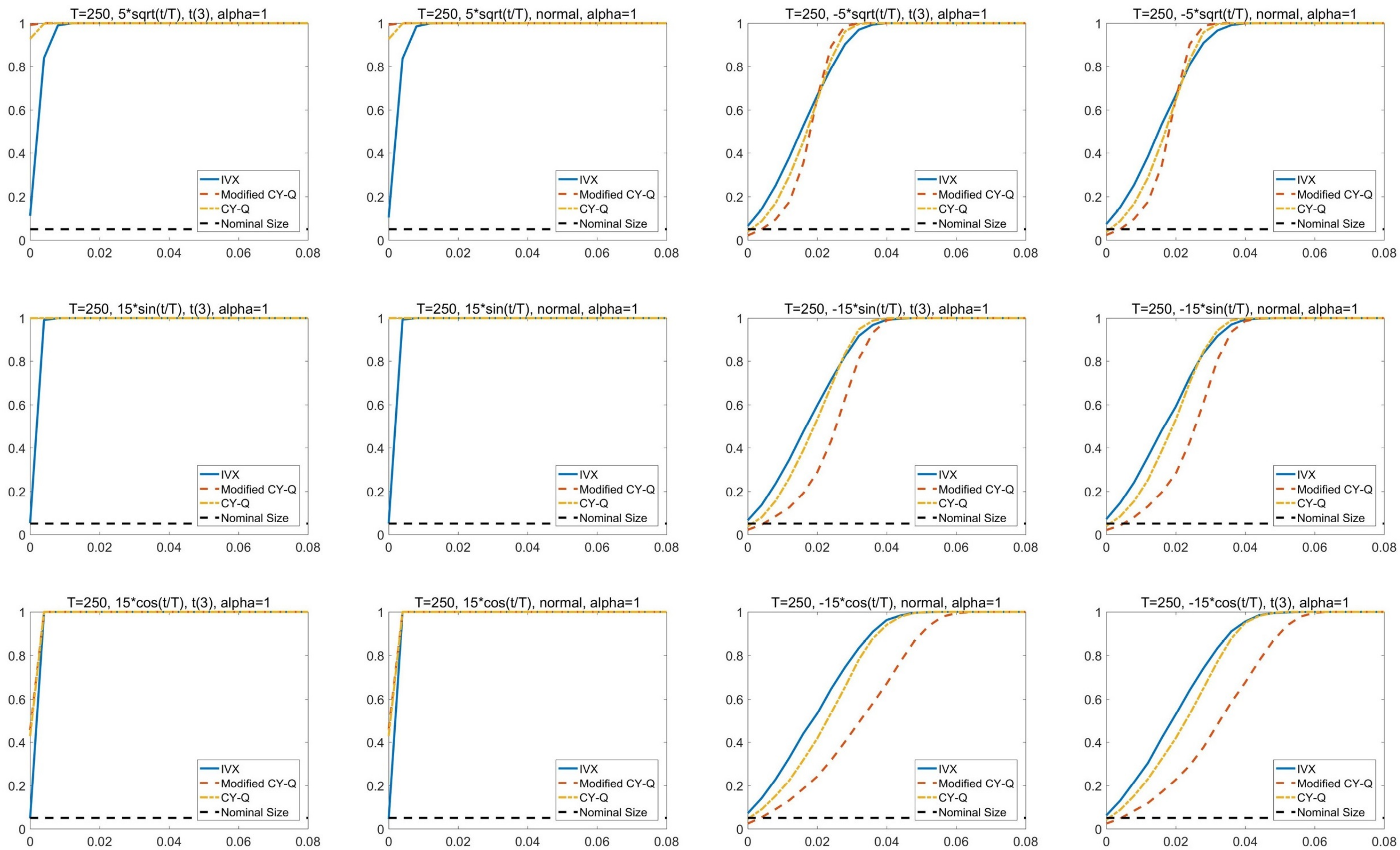

Figure 1.4: Size and power performance of IVX, CY-Q, and Modified CY-Q tests with continuous c(t/n) when alpha=0.5

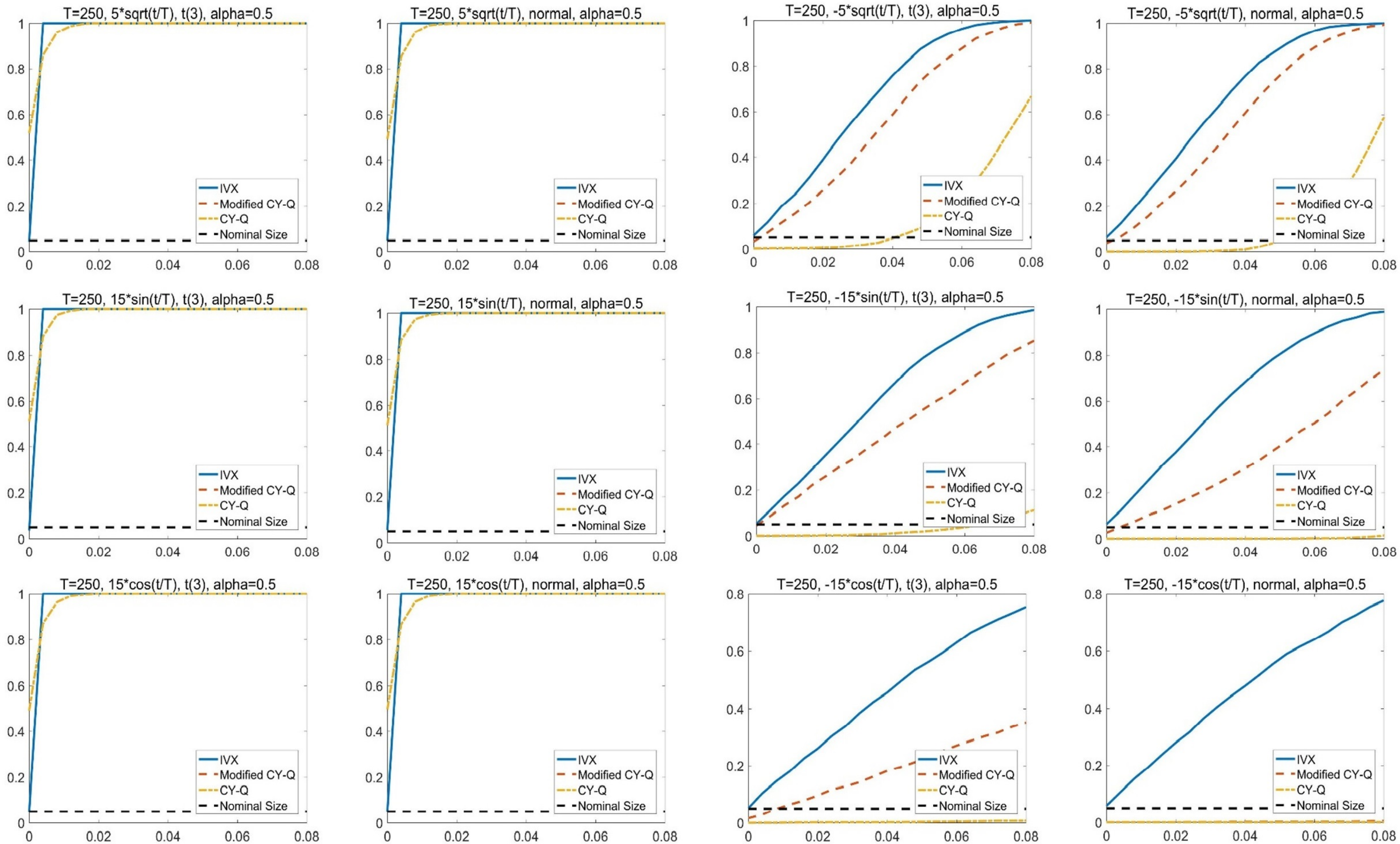

Figure 1.5: Empirical results of nonparametric estimation

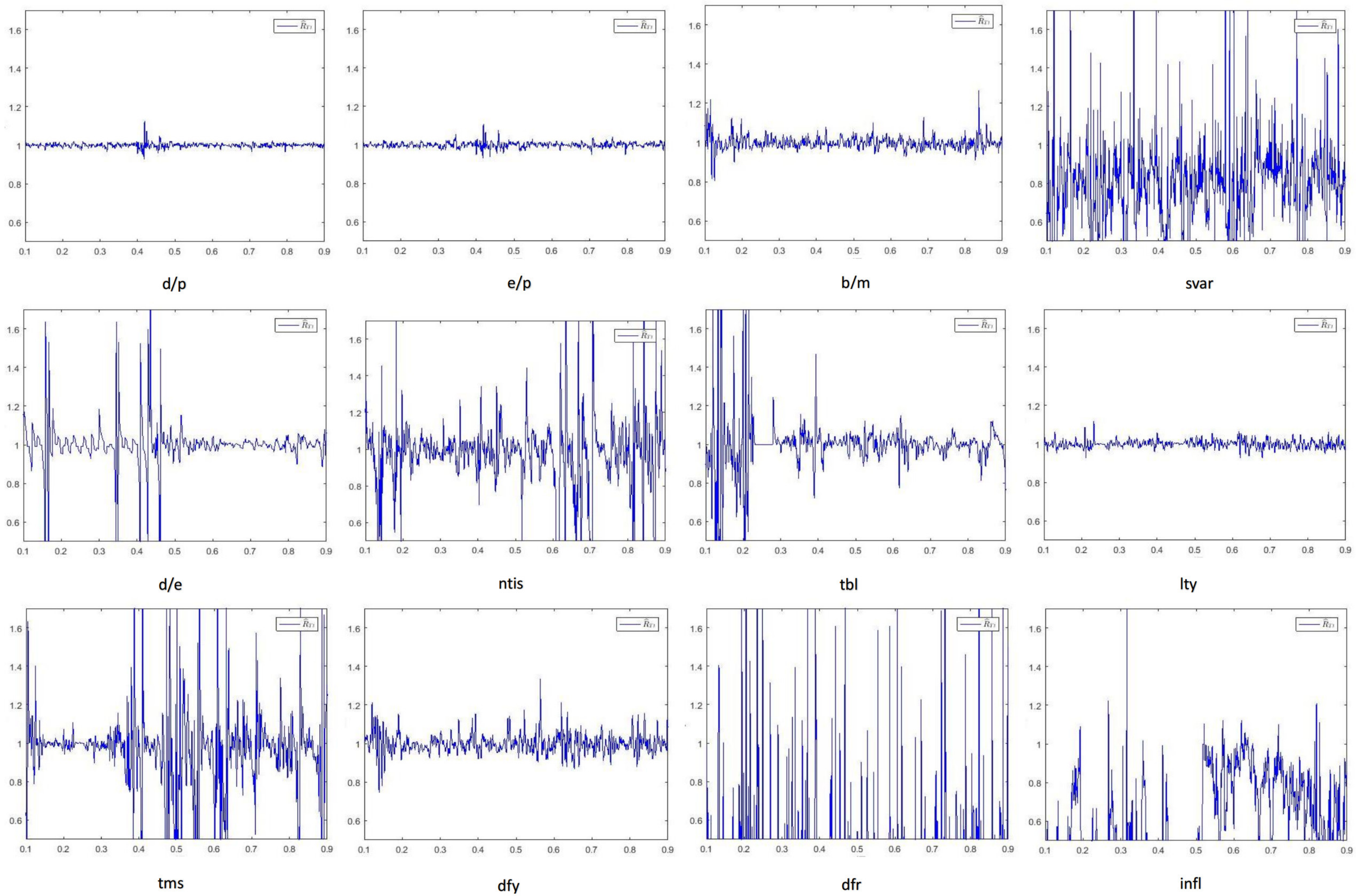

Table 1.1: Empirical results for univariate regressions

	d/p	*e/p*	*b/m*	*svar*	*d/e*	*ntis*	*tbl*	*lty*	*tms*	*dfy*	*dfr*	*infl*
1927:01–2017:12												
$\hat{\beta}_1$	-0.07	-0.04	1.26	-22.29	-0.45	-11.52	-12.92	-9.49	12.60	22.43	13.77	-50.50
t-stat	-0.74	-0.33	**2.04**	-0.77	-1.10	-1.69	-1.75	-1.32	1.12	1.06	1.14	-1.57
1952:01–2017:12												
$\hat{\beta}_1$	-0.14	-0.17	0.54	-113.86	-0.22	-1.13	-17.35	-11.40	20.24	24.64	11.52	-86.46
t-stat	-1.70	-1.64	0.85	**-3.15**	-0.63	-0.13	**-2.63**	-1.73	**2.05**	0.90	1.09	-1.94

Table 1.2: Empirical results for multivariate regressions

	b/m, tbl	*b/m, tms*	*svar, tbl*	*svar, tms*
1927:01–2017:12				
$\hat{\beta}_{1,1}$	1.19	1.26	-17.33	-26.30
$\hat{\beta}_{1,2}$	-21.77	-0.35	-13.29	16.34
Wald-stat	**9.87**	4.58	3.67	2.27
1952:01–2017:12				
$\hat{\beta}_{1,1}$	0.95	0.64	-96.65	-120.67
$\hat{\beta}_{1,2}$	-25.45	14.75	-19.77	26.09
Wald-stat	**13.20**	4.63	**17.06**	**15.27**

Chapter 2

Nonparametric Estimation and Inference for Functional Local-to-unity Processes

2.1 Introduction

Extensive studies on asset prices suggest evidence of time-varying dynamic properties. A popular strategy to model time-varying dynamic properties is to use a model with time-varying parameters. For example, Plazzi et al. (2010) demonstrate the existence of time-varying parameters in the real estate market. In the literature, economic reasons have been provided to explain why time-varying parameters are needed. For example, Bossaerts and Hillion (1999) and Bekaert et al. (2007) characterize the model instability in the stock market caused by changes in market sentiment and monetary policy.

To estimate models with time-varying parameters, traditional methods, such as least-squares, may no longer be applicable. More sophisticated estimation methods are upon request. Moreover, the validity of standard inference procedures, which are developed for models with constant parameters, may fail to work for models with time-varying parameters.

In this paper, we investigate the dynamics in economic variables with intertemporal self-dependence. If the self-relationship is linear and time-invariant, the most widely used autoregressive model is the AR(1) model. Given that many economic time series have wandering around behavior, AR models with the exact unit root and a local-to-unit root (LUR) have been extensively employed in practice; see Chan and Wei (1987), Phillips (1987, 1988). To allow for a time-varying property in the context of the LUR model, Bykhovskaya and Phillips (2020, BP hereafter) propose a functional local-to-unit root (FLUR) model, whose slope is a function of time. BP show how to construct the point optimal FLUR tests by generalizing the LUR asymptotics to cases where the localized departure from unity is a time-varying function rather than a constant. In this paper, we test for the functional departure from unity by developing nonparametric inferences.

In the nonparametric literature, two classes of nonparametric estimation methods co-exist. The first is the class of local approximation methods, while the second type is a class of global approximation methods. Local approximation methods, such as the Nadaraya-Watson or local polynomial estimates, have been used for functional co-integration by Juhl and Xiao (2005), Wang and Phillips (2009a, 2009b), Li et al. (2020), and Phillips et al. (2017). Global approximation methods, based on the trigonometric basis, have been used by Phillips (1998), Park and Hahn (1999), Cai et al. (2009), Bierens and Martins (2010, BM hereafter), and Martins (2018). The sieve method that approximates time-varying parameters by basis functions with diverging dimensions is easier to compute than local approximation methods. For computational convenience, we use the sieve method to estimate unknown functional departures from unity. We show that the proposed estimator for slope is consistent as the dimension of the orthogonal basis increases. We also develop the asymptotic distribution of the estimator. It is established that the asymptotic distribution, being the ratio of stochastic integrals, is nonstandard. Moreover, we show that the time-unstable distance parameters cannot be estimated consistently.

In model specifications, we establish the consistent estimate of the distance parameter by grouping homogeneous cross-sectional units. Under conditions of joint convergence, the panel estimator of the distance parameter is consistent and normally distributed. We also provide a Wald-type statistic to detect the time-varying patterns. We establish the joint asymptotics and the power envelope of the Wald statistic. To illustrate our proposed test, we conduct an empirical study using the price index of China's real estate market. For the group of large- and medium-sized cities, we find substantial evidence of time-varying rates of growth in the price index.

The remainder of this chapter is organized as follows. Section 2 introduces the model setup and assumptions. Section 3 presents the preliminary results of the orthogonal basis. Section 4 demonstrates an asymptotic theory of the time-series estimator. Section 5 discusses panel FLUR autoregressions and shows the consistency of the specification statistic. Section 6 shows numerical simulations. Section 7 applies the specification test to China's real estate markets. Section 8 concludes. The appendix collects the technical proofs of the theorems.

Throughout the paper, we use the following notation. For an arbitrary matrix M, we use $\|M\|_\infty$, $\|M\|_2$ and $\|M\|$ to denote the L^∞, L^2 and the spectral norms of matrix M. We use $\Rightarrow$, $\xrightarrow{p}$, $\xrightarrow{d}$ to denote a weak convergence in a functional space, convergence in probability, and convergence in distribution, respectively. Finally, equality in distribution is represented by $=_d$.

2.2 Model and Assumptions

Following BP, the model setup is given as

$$\begin{aligned} y_t &= R_{Tt} y_{t-1} + u_{xt}, \\ \text{with } R_{Tt} &= \exp\left(\frac{C(t/T)}{T}\right) \text{ and } t = 1, 2, ..., T, \end{aligned} \tag{2.1}$$

where $\{y_t\}_{t=1}^{T}$ is a sequence of $n\times 1$ random vectors, $R_{Tt}(:= R_T\,(t/T))$ is an $n\times n$ matrix of time-varying parameters, and $C(t/T) = diag\,\{c_1(t/T), c_2(t/T), \ldots,\ c_n(t/T)\}$ $(:= C_{Tt})$, with $c_j\,(\cdot)$ being a smooth function defined on $[0, 1]$ for all j. By satisfying the rational expectation hypothesis, the persistence of economic variables, like real-estate prices, is similar to the LUR process. It is too restrictive to assume the autoregressive root is stable across time. Therefore, we employ the model (2.1) with time-varying slopes to approximate economic fundamentals; for instance, the prices of housing markets. To conduct further investigations, we impose the following assumptions.

Assumption 2.1. *(i)* $\{u_{xt}\}_{t=0}^{\infty}$ *is a martingale difference sequence with the second moment* $\Sigma (:= \mathbb{E}\left(u_{xt}u_{xt}^{'}\right))$ *and the finite* p^{th} *moment for some* $p \geq 4$*;*

(ii) There exist $\bar{c}$ *and* $\underline{c}$*, such that* $\underline{c} < c_i(\cdot) < \bar{c}$ *for any* $i = 1, 2, ..., n$*, where* $-\infty < \underline{c} < \bar{c} < \infty$.

Assumption 1 (i) imposes a martingale property for the innovations. These innovations can be generalized as a stationary linear process

$$u_{xt} = \sum_{i=0}^{\infty} \Phi_x(i)\epsilon_{t-i},$$

where $\{\epsilon_t\}_{t=0}^{\infty}$ is a sequence of martingale difference processes with a second moment $\Sigma_\epsilon := \mathbb{E}\left(\epsilon_t\epsilon_t^{'}\right)$ and a finite p^{th} moment for some $p \geq 4$. The Beveridge-Nelson-Phillips decomposition implies that

$$u_{xt} = u_{xt}^{*} + \left(\widetilde{u}_{x,t-1} - \widetilde{u}_{xt}\right),$$

where $u_{xt}^{*} = \Phi_x(1)\epsilon_t$, and $\widetilde{u}_{xt} = \sum_{i=0}^{\infty} \widetilde{\Phi}_x(i)\epsilon_{t-i}$ with $\widetilde{\Phi}_x(i) = \sum_{j=i+1}^{\infty} \Phi_x(j)$. Phillips and Solo (1992, PS hereafter) show that $\sum_{i=0}^{\infty} i \left\|\Phi_x(i)\right\| < \infty$ is sufficient to justify the condition that $\sum_{i=0}^{\infty} \left\|\widetilde{\Phi}_x(i)\right\| < \infty$. Then, the dominating term u_{xt}^{*} is also a martingale difference sequence. This generalization will reserve for the future research. Assumption 1 (ii) assumes bounded ranges for each entry of $C(t/T)$.

Assumption 2.2. *(i)* $C(t/T)$ *is* q^{th}*-order differentiable with bounded derivatives on* $[0, 1]$ *for some* $q \geq 1$*;*

(ii) $\frac{2}{2q-1} < \frac{p-2}{3p}$ *and* $\frac{1}{k} + \frac{T^{\frac{2}{2q-1}}}{k} + \frac{k}{T^{\frac{p-2}{3p}}} \to 0$.

Assumption 2.2 (i) assumes the differentiability of the FLUR coefficients. If $p \geq 4$, then Assumption 2.2 (ii) requires the smallest integer value for q to be 4.

Consequently, $C(\cdot)$ needs to be at least 4^{th}-order differentiable. According to BM, linear combinations of the orthogonal basis can approximate an unknown slope R_{Tt} at the rate of $O(k^{-q})$. As a result, when the dimensions of the orthogonal basis diverge at an appropriate speed, the approximation bias is negligible.

2.3 Orthogonal Basis and Sieve Estimator

We consider a pointwise trajectory, $\prod R_T := (R_T(r_1), R_T(r_2), ..., R_T(r_d))'$, where $\{r_i\}_{i=1}^d$ are selected grid points on $[0, 1]$. To approximate $\prod R_T$, we use a sequence of orthogonal basis $\{\phi_i(\cdot)\}_{i=1}^k$. We assume that the dimension of the orthogonal basis, k, diverges under the restrictions of Assumption 2.2 (ii). For $R_T(\cdot)$, the k-dimensional approximation $R_T^{(k)}(\cdot)$ has the form of

$$R_T^{(k)}(\cdot) := \sum_{i=1}^{k} \beta_{ki}\phi_i(\cdot), \tag{2.2}$$

where β_{ki} is an $n \times n$ diagonal matrix. To simplify notations, we combine $\{\beta_{ki}\}_{i=1}^k$ into an $nk \times n$ coefficient matrix β_k so that $\beta_k := (\beta_{k1}, \beta_{k2}, ..., \beta_{kk})'$. We merge a k-dimensional orthogonal basis $\{\phi_i(\cdot)\}_{i=1}^k$ into a vector so that $f_k(\cdot) := (\phi_1(\cdot), \phi_2(\cdot), ..., \phi_k(\cdot))'$. Hence, we can rewrite (2.2) as

$$R_T^{(k)}(\cdot) := (f_k'(\cdot) \otimes I_{n\times n})\, \beta_k.$$

Similarly, the pointwise trajectory $\prod R_T^{(k)}$ has the representation of

$$\prod R_T^{(k)} = \left[(f_k(r_1), f_k(r_2), ..., f_k(r_d))' \otimes I_{n\times n}\right] \beta_k = S_k\beta_k,$$

where $S_k := (f_k(r_1), f_k(r_2), ..., f_k(r_d))' \otimes I_{n\times n}$. Therefore, the problem of estimating the infinite-dimensional argument $R_T(\cdot)$ degenerates into the estimation of a finite-dimensional coefficient matrix β_k.

Define a consistent estimator for β_k to be $\widehat{\beta}_k$. Based on $\widehat{\beta}_k$, the estimators of $R_T^{(k)}(\cdot)$ and $\prod R_T^{(k)}$ have the following expressions

$$\widehat{R}_T^{(k)}(\cdot) := \sum_{i=1}^{k} \widehat{\beta}_{ki}\phi_i(\cdot),$$

and

$$\prod \widehat{R}_T^{(k)} := \left[(f_k(r_1), f_k(r_2), ..., f_k(r_d))' \otimes I_{n\times n}\right] \widehat{\beta}_k.$$

For our model, it is natural to consider the least-squares estimator $\widehat{\beta}_{Tk}$ as,

$$\widehat{\beta}_{Tk} := \left(Y'_{Tk,-1}Y_{Tk,-1}\right)^{-1} Y'_{Tk,-1}Y_T = \beta_k + \left(Y'_{Tk,-1}Y_{Tk,-1}\right)^{-1} Y'_{Tk,-1}U_{xk}, \quad (2.3)$$

where $Y_{Tk,-1} := (y_{k0}, y_{k1}, ..., y_{k,T-1})'$, $U_{xk} := (u_{xk,1}, u_{xk,2}, ..., u_{xk,T})'$, $y_{k,t-1} := f_k\left(\frac{t}{T}\right) \otimes y_{t-1}$, and $u_{xk,t} := u_{xt} + \left(R_T - R_T^{(k)}\right)\left(\frac{t}{T}\right) y_{t-1}$. Equivalently, the expression (2.3) comes from the fitted equation as follows:

$$y_t = \beta_k y_{k,t-1} + u_{xk,t}. \quad (2.4)$$

Following Elliot (1964), we adopt the Chebyshev basis as an orthogonal basis. The orthogonal basis is defined as,

$$\phi_0\left(\frac{[Tr]}{T}\right) = 1 \text{ and } \phi_j\left(\frac{[Tr]}{T}\right) = \sqrt{2}\cos\left(j\pi\left(\frac{[Tr]}{T}\right)\right),$$

where $r \in [0,1]$ and $j = 1, 2, ..., k$. Orthogonality comes from the cosine functions as

$$2\int_0^1 \phi_i(r)\phi_j(r)dr = \begin{cases} 0 & i \neq j \\ \pi & i = j \neq 0 \\ 2\pi & i = j = 0 \end{cases}.$$

Based on the orthogonal property, Elliot (1964) demonstrates that $\{\phi_i(\cdot)\}_{i=1}^{\infty}$ form a set of basis functions in the Hilbert space $L^2[0,1]$. With any square-integrable function $R_T(r)$ on $[0,1]$, we have its orthogonal decomposition as

$$R_T\left(\frac{[Tr]}{T}\right) = \sum_{i=1}^{\infty} \beta_{T,i}\phi_i\left(\frac{[Tr]}{T}\right), \quad (2.5)$$

where $\beta_{T,i} := \int_0^1 R_T\left(\frac{[Tr]}{T}\right)\phi_i\left(\frac{[Tr]}{T}\right)dr$. In this paper, we define the approximation for $R_T\left(\frac{[Tr]}{T}\right)$ as $R_T^{(k)}\left(\frac{[Tr]}{T}\right)$. The k-dimensional approximation $R_T^{(k)}\left(\frac{[Tr]}{T}\right)$

has the expression of

$$R_T^{(k)}\left(\frac{[Tr]}{T}\right) := \sum_{i=1}^{k} \beta_{Tk,i}\phi_i\left(\frac{[Tr]}{T}\right), \tag{2.6}$$

where $\beta_{Tk,i} := \beta_{T,i}$ as in (2.5). To asymptotically eliminate the distance between (2.5) and (2.6), we need to show

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\left\|R_T\left(\frac{[Tr]}{T}\right) - R_T^{(k)}\left(\frac{[Tr]}{T}\right)\right\|^2 \to 0,$$

as illustrated in the following lemma.

Lemma 2.1. *Let Assumptions 2.1 and 2.2 hold. For any $k \geq 1$, we have*

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\left\|R_T\left(\frac{[Tr]}{T}\right) - R_T^{(k)}\left(\frac{[Tr]}{T}\right)\right\|^2 = O_p(k^{-2q}).$$

BM bound the approximation error by the following inequality:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\left\|R_T\left(\frac{[Tr]}{T}\right) - R_T^{(k)}\left(\frac{[Tr]}{T}\right)\right\|^2 \leq \frac{\int_0^1 \left\|R_T^{(q)}(r)\,dr\right\|}{\pi^{2q}(k+1)^{2q}},$$

where $R_T^{(q)}(r)$ is the q^{th}-order derivative of $R_T(r)$. When the dimension of basis functions diverges, the approximation error diminishes asymptotically.

Denote

$$K_{T,c}^*(r) := \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\exp\left(\frac{1}{T}\sum_{j=t}^{[Tr]} C\left(\frac{j}{T}\right)\right)u_{xt}.$$

The stochastic process $K_c(r)$ follows $dK_c(r) = C(r)K_c(r)dr + \Sigma^{\frac{1}{2}}dW_x(r)$. For asymptotic derivations, we need a uniform approximation of $K_{T,c}^*(r)$ to $K_c(r)$ as the functional central limit theorem developed by BM is insufficient. The modified Skorokhod embedding theorem ensures the following uniform approximation.

Lemma 2.2. *Let Assumptions 2.1 and 2.2 hold. Define $\alpha := \frac{p-2}{2p}$. As $T \to \infty$, we have*

$$\sup_{0\leq r\leq 1}\left\|K_{T,c}^*(r) - K_c(r)\right\| = O_p(T^{-\alpha}).$$

2.4 Asymptotic Theory for Sieve Estimator

Following (2.3), the least-squares estimator for β_k has the decomposition

$$\begin{aligned}
\widehat{\beta}_{Tk} &= \beta_k + \left(Y'_{Tk,-1}Y_{Tk,-1}\right)^{-1} Y'_{Tk,-1}U_{xk} \\
&= \beta_k + \left(Y'_{Tk,-1}Y_{Tk,-1}\right)^{-1} Y'_{Tk,-1}\left(U_{xT} + Y_{Tk,-1}\left(R_T - R_T^{(k)}\right)\left(\frac{t}{T}\right)\right) \\
&= \beta_k + \left(Y'_{Tk,-1}Y_{Tk,-1}\right)^{-1} Y'_{Tk,-1}U_{xT} + \\
&\quad \left(Y'_{Tk,-1}Y_{Tk,-1}\right)^{-1} Y'_{Tk,-1}Y_{Tk,-1}\left(R_T - R_T^{(k)}\right)\left(\frac{t}{T}\right),
\end{aligned}$$

where the term $\left(Y'_{Tk,-1}Y_{Tk,-1}\right)^{-1} Y'_{Tk,-1}U_{xT}$ contributes to the nonstandard limiting distribution, and $\left(Y'_{Tk,-1}Y_{Tk,-1}\right)^{-1} Y'_{Tk,-1}Y_{Tk,-1}\left(R_T - R_T^{(k)}\right)\left(\frac{t}{T}\right)$ represents the dominated approximation error. According to Assumption 2.2 and Lemma 2.1, both the approximation error and the asymptotic error are diminishing as $T \to \infty$. The negligibility of these errors justifies the consistency of our sieve estimates.

Theorem 2.1. *Let Assumptions 2.1 and 2.2 hold. As $T \to \infty$, we have*

$$\prod \widehat{R}_T^{(k)} = \prod R_T + O_p\left(kT^{-1}\right) + O_p\left(k^{-q}\right).$$

The asymptotic error term is of order $O_p\left(kT^{-1}\right)$, and the approximation bias decreases at the rate of $O_p\left(k^{-q}\right)$. Further, we define the sample moment as

$$A_{Tk} := S_k\left(Y'_{Tk,-1}Y_{Tk,-1}\right)^{-1} S_k.$$

Based on A_{Tk}, we have the following limiting distribution of the time-series sieve estimator.

Theorem 2.2. *Let Assumptions 2.1 and 2.2 hold. As $T \to \infty$, we have*

$$A_{Tk}^{-\frac{1}{2}}\left(\prod \widehat{R}_T^{(k)} - \prod R_T\right) \Rightarrow \mathbf{Z},$$

where

$$\mathbf{Z} = \lim_{k\to\infty}\left[S_k\left(\int_0^1\left(f_k(r)f_k^{'}(r)\right)\otimes\left(K_c(r)K_c^{'}(r)\right)dr\right)^{-1}S_k\right]^{-\frac{1}{2}}\times$$
$$S_k\left(\int_0^1\left(f_k(r)f_k^{'}(r)\right)\otimes\left(K_c(r)K_c^{'}(r)\right)dr\right)^{-1}\left(\int_0^1 f_k(r)\otimes K_c(r)dW_x^{'}(r)\right),$$

and $W_x(\cdot)$ *is a standard Brownian motion.*

Remark 2.1. *In the appendix A, we show that the kernel estimate of FLUR is consistent and asymptotically normal. According to Theorem 2.2, the sieve estimator of FLUR is also consistent but not of normality anymore. The reason for the asymptotic non-normality is that* $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}u_{xt}$ *and* $K_{T,c}^{*}(r)$ *are asymptotically correlated.*

Remark 2.2. *As the sample moment is diverging of the rate* $O_p(T/k)$*, the convergence rate for the sieve estimate is* $O_p(k/T)$*. Similarly, the convergence rate of the kernel method is* $O_p\left(\frac{1}{T\sqrt{h}}\right)$ *in appendix A. For the time-invariant case, Phillips (1987) shows that the least-squares estimate converges at the rate of* $O_p\left(\frac{1}{T}\right)$*, faster than that of the sieve estimate.*

2.5 Panel Specification Test

Based on (2.1), the failure of a specification test happens due to the inconsistent estimations on $C(\cdot)$. To resolve this problem, we consider the panel autoregressions as in Moon and Phillips (2004). The primary benefit of the panel approach is that slope homogeneity can massively enhance the powers of the tests. Assumptions of homogeneity across individuals can either be deduced by economic theory or be verified by the machine learning algorithm. Therefore, without introducing additional complexity, we obtain the theoretical results by imposing a homogeneity assumption.

We consider the scalar case, where $n = 1$. We collect M homogeneous time series in (2.1). For each individual, there are T observations. With m being the index for cross-sectional identity, we have

$$y_{mt} = R_{Tt} y_{m,t-1} + u_{mx,t}, \tag{2.7}$$

where $R_{Tt} = \exp\left(\frac{C(t/T)}{T}\right)$ for each $m = 1, 2, ..., M$, and $t = 1, 2, ..., T$.

We assume that $\{y_{mt}\}_{t=1}^{T}$ and $\{u_{mx,t}\}_{t=1}^{T}$ follow Assumptions 2.1 and 2.2. In addition, we impose Assumption 2.3.

Assumption 2.3. *(i) Innovations $\{u_{mx,t}\}_{t=1}^{T}$ are independent across $m \in \{1, 2, ..., M\}$;*

(ii) $\frac{k}{\sqrt{M}} \to 0$.

When $n = 1$, we define the panel sieve estimator as

$$\widehat{\beta}_{Tk} = \left(\sum_{m=1}^{M} Y'_{mTk,-1} Y_{mTk,-1}\right)^{-1} \left(\sum_{m=1}^{M} Y'_{mTk,-1} Y_{mTk}\right), \tag{2.8}$$

where $Y_{mTk,-1} = (y_{mk,0}, y_{mk,1}, ..., y_{mk,T-1})'$, $U_{mxk} = (u_{mxk,1}, u_{mxk,2}, ..., u_{mxk,T})'$, $y_{mk,t-1} = f_k\left(\frac{t}{T}\right) \otimes y_{m,t-1}$, and $u_{mxk,t} = u_{mxt} + \left(R_T - R_T^{(k)}\right)\left(\frac{t}{T}\right) y_{m,t-1}$. The estimated pointwise trajectory on the chosen grid $\{r_1, r_2, ..., r_d\} \subseteq [0, 1]$ is

$$\prod \widehat{R}_T^{(k)} = \left[(f_k(r_1), f_k(r_2), ..., f_k(r_d))' \otimes I_{n\times n}\right] \widehat{\beta}_{Tk}.$$

We justify the asymptotic normality of the panel sieve estimator for the scalar FLUR process.

Theorem 2.3. *Let Assumptions 2.1-2.3 hold. As $(M, T) \to \infty$,*

$$A_{MTk}^{-\frac{1}{2}} \left(\prod \widehat{R}_T^{(k)} - \prod R_T\right) \xrightarrow{d} N(0, I_d),$$

where I_d is d-dimensional identity matrix and $A_{MTk} := S_k \left(\sum_{m=1}^{M} Y'_{mTk,-1} Y_{mTk,-1}\right)^{-1} S'_k$.

Remark 2.3. *The convergence rate of the panel estimator $\widehat{C}(\cdot)$ is $O_p\left(\frac{k}{\sqrt{M}}\right)$. When parameters are time-invariant, Moon and Phillips (2004) show that the convergence rate is $O_p\left(\frac{1}{\sqrt{M}}\right)$, which is is faster.*

Remark 2.4. *The cross-sectional least-squares estimator for each $t \in \{1, 2, ..., T\}$ is an alternative estimator. However, the cross-sectional estimator may suffer* *from inconsistency when M diverges at a slower speed than T. Moreover, the cross-sectional least-squares estimator is less efficient than our panel estimate under the joint limit framework.*

With asymptotic normality, we propose a consistent Wald statistic to test parameter instability.

Theorem 2.4. *Let Assumptions 2.1-2.3 hold. Under the null hypothesis, H_0 : R_{Tt} is time-invariant, as $(M, T) \to \infty$,*

$$\frac{\left(\prod \widehat{R}_T^{(k)} - \prod \widetilde{R}_T\right)' A_{MTk}^{-1} \left(\prod \widehat{R}_T^{(k)} - \prod \widetilde{R}_T\right)}{\widehat{\Sigma}} \xrightarrow{d} \chi^2(d), \tag{2.9}$$

where $\widehat{\Sigma} := \frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T}\widehat{u}_{mx,t}^2$ with $\widehat{u}_{mx,t} = y_{mt} - \widehat{R}_{Tt}^{(k)} y_{m,t-1}$ for $m = 1, 2, ..., M$ and $t = 1, 2, ..., T$. The pooled least-squares estimator is denoted by $\widetilde{R}_T$. Under the alternative hypothesis, H_1 : R_{Tt} is time-varying, as $(M, T) \to \infty$,

$$\frac{\left(\prod \widehat{R}_T^{(k)} - \prod \widetilde{R}_T\right)' A_{MTk}^{-1} \left(\prod \widehat{R}_T^{(k)} - \prod \widetilde{R}_T\right)}{\widehat{\Sigma}} = O_p\left(\frac{M}{k}\right) \to \infty.$$

2.6 Monte Carlo Simulations

This section investigates the finite sample performance of the sieve statistic in the panel FLUR. We examine the following data generating process as

$$y_{mt} = \exp\left(\frac{C(t/T)}{T}\right) y_{m,t-1} + u_{mx,t}.$$

The functional time-varying slopes are defined as

Model 1: $C(t/T) = 20\cos(2\pi t/T)$;

Model 2: $C(t/T) = 12\sqrt{t/T}$;

Model 3: $C(t/T) = 5t/T + 5|\sin(4\pi t/T)|$;

Model 4: $C(t/T) = 15\sin(t/T)$.

The innovations follow either normal distributions or heavy-tail t-distributions with 3 degrees of freedom. Noise is set to be independent across individuals. The homoscedastic variance is set as 2. The numbers of time-series observations are 100 and 200. The dimensions for the cross-sectional units are chosen as 20, 50 and 80. The simulation is conducted with 2,000 replications. Sieve basis functions are selected as either orthogonal trigonometric basis (Chebyshev basis) or as B-splines. The B-splines are employed to justify the robustness of our approach. To satisfy Assumptions 2.1-2.3, we select k as $5T^{\frac{1}{4}}$ and d as 1 with $r_d = 0.5$.

The plots of $C\left(t/T\right)$ and $\widehat{C}\left(t/T\right)$ are given in Figure 2.1. From theoretical derivations, we know that the convergence rate for $C\left(t/T\right)$ estimates is $O\left(\frac{k}{\sqrt{M}}\right)$. This feature is essential, as the validity of the specification tests relies heavily on the consistency of estimates. When the cardinality of cross-sectional units is as small as 20, the estimates are close to the true values. These results demonstrate the excellent finite sample performance of our estimates. Besides, the robustness of this approach is shown by various functional slopes.

The powers of specification tests are shown in Tables 2.1-2.4. Tables 2.1 and 2.2 represent cases with normally distributed errors, and Tables 2.3 and 2.4 refer to performances with t(3) distributions. We evaluate the powers by applying two sieve functions, the orthogonal basis and the B-spline. The simulated results show that cross-sectional asymptotics improve powers. With the larger cardinality of individuals, the precision of the estimates and the powers of specification tests are upgraded. Besides, with an identical model setup, the powers of specification tests with B-splines are nearly the same as those with an orthogonal basis. Excellent performances with B-splines demonstrate the robustness of our asymptotic theory. Tables 2.5-2.8 display the empirical sizes when the nominal level is set at 5%. Regardless of whether the residuals are normal or heavy-tailed, the empirical sizes are well controlled around 5% with both sieve approximations.

2.7 Empirical Illustrations

We apply sieve inferences on the data concerning China's real estate market. Su and Ju (2018) use classification algorithms on the housing prices in China's large- and medium-sized cities. They identify three latent groups. With identical slopes in each group, they illustrate that the growth rates of real estate prices are more persistent in large cities. In this paper, following the membership in Su and Ju (2018), we detect the time-varying patterns in annualized growth rates using the price index (PI hereafter) on the 33 Tier-1 and Tier-2 cities. Table 2.10 shows the list of these cities. Fang et al. (2016) construct real estate PIs for 120 major Chinese cities from 2003 m1 to 2013 m3, based on sequential sales of new homes.

First, we follow the membership structures in Su and Ju (2018), where they apply C-Lasso method to group the 69 cities in China according to their housing price. The results show that the housing price in the group of big cities show similar characteristics in terms of persistence. In this paper, we merge 33 Tier-1 and Tier-2 cities into a panel model. The plots of the annual growth rates are provided in Figure 2.2. We can observe the common behaviors in the cities of interest, which support the construction of homogeneous panel autoregression.

Second, we apply panel sieve inferences on R_{Tt}. To verify our conjecture on time-varying slopes, we conduct the panel specification test. We select the grid points as 0.2, 0.4, 0.6, and 0.8, and compute the statistics on all grid points. The resulting statistics are reported in Table 2.9. From these results, we can observe that all statistics are higher than critical values. These findings provide strong evidence for time-varying patterns. The results also verify the usefulness of the proposed specification statistics.

2.8 Conclusion

In this paper, we propose a sieve inference procedure built on an orthogonal trigonometric basis. Time-series sieve estimators on slopes are consistent, and they converge to the ratios of stochastic integrals. We provide a panel specification test with excellent finite-sample performance, which is supported by numerical simulations. In addition, we apply panel specification tests to assess China's real estate market and obtain significant empirical results.

The sieve basis functions that we use are the orthogonal trigonometric basis. Other basis functions, such as splines or father and mother wavelets, are also widely used in the literature, especially the B-splines. We prefer orthogonal trigonometric basis to B-splines, due to their smoothness. In stationary time series analysis, the sample moment converges to its non-random population moment. However, the probability limit of the sample moment in the FLUR model involves a stochastic integral. In the proof, we implement an integration-by-part formula, which requires the differentiability of basis functions. If smoothness is not necessary for theoretical justifications, then the B-spline becomes applicable. Indeed, the extensive numerical simulations verify our conjectures and demonstrate the usefulness of B-splines in FLUR models.

Tables and Figures

Figures and tables on numerical experiments and empirical illustrations are given.

Monte Carlo simulations

Table 2.1: Powers of Specification Test (Orthogonal Trigonometric Basis) with Normality

Model	M=20, T=100	M=20, T=200	M=50,T=100	M=50,T=200	M=80,T=100	M=80,T=200
Model 1	0.8855	0.8725	0.9115	0.887	0.973	0.9665
Model 2	0.844	0.811	0.91	0.8835	0.983	0.977
Model 3	0.8335	0.8225	0.9335	0.8905	0.9785	0.98
Model 4	0.8415	0.8459	0.9455	0.9185	0.9885	0.9865

Table 2.2: Powers of Specification Test (B-spline) with Normality

Model	M=20, T=100	M=20, T=200	M=50,T=100	M=50,T=200	M=80,T=100	M=80,T=200
Model 1	0.879	0.8545	0.9485	0.9005	0.995	0.9205
Model 2	0.8565	0.8705	0.9555	0.9095	0.997	0.95
Model 3	0.874	0.8455	0.9535	0.915	0.9955	0.9555
Model 4	0.8889	0.8715	0.9685	0.927	0.998	0.9725

Table 2.3: Powers of Specification Test (Orthogonal Trigonometric Basis) with t(3)

Model	M=20, T=100	M=20, T=200	M=50,T=100	M=50,T=200	M=80,T=100	M=80,T=200
Model 1	0.8515	0.849	0.867	0.8648	0.9515	0.9385
Model 2	0.8295	0.8315	0.921	0.8785	0.984	0.9815
Model 3	0.837	0.836	0.9375	0.9	0.987	0.985
Model 4	0.862	0.8455	0.9315	0.9175	0.983	0.9815

Table 2.4: Powers of Specification Test (B-spline) with t(3)

Model	M=20, T=100	M=20, T=200	M=50,T=100	M=50,T=200	M=80,T=100	M=80,T=200
Model 1	0.8475	0.8395	0.8715	0.868	0.966	0.962
Model 2	0.861	0.849	0.9525	0.9115	0.9965	0.933
Model 3	0.882	0.855	0.96	0.925	0.9965	0.95
Model 4	0.844	0.846	0.9585	0.937	0.992	0.9645

Table 2.5: Sizes of Specification Test (Orthogonal Trigonometric Basis) with Normality

Model	M=20, T=100	M=20, T=200	M=50,T=100	M=50,T=200	M=80,T=100	M=80,T=200
C=-5	0.064	0.06	0.0663	0.0607	0.0671	0.0625
C=5	0.0422	0.0425	0.0512	0.0395	0.0601	0.0471
C=-10	0.0666	0.0625	0.0608	0.0605	0.0631	0.0559
C=10	0.0422	0.044	0.0538	0.0401	0.0606	0.0476

Table 2.6: Sizes of Specification Test (Orthogonal Trigonometric Basis) with t(3)

Model	M=20, T=100	M=20, T=200	M=50,T=100	M=50,T=200	M=80,T=100	M=80,T=200
C=-5	0.0567	0.0527	0.0649	0.0573	0.0565	0.0604
C=5	0.0482	0.0461	0.0519	0.0396	0.0526	0.0544
C=-10	0.0548	0.0506	0.0578	0.052	0.0545	0.0549
C=10	0.0466	0.0473	0.0545	0.0393	0.0485	0.0482

Table 2.7: Sizes of Specification Test (B-spline) with Normality

Model	M=20, T=100	M=20, T=200	M=50,T=100	M=50,T=200	M=80,T=100	M=80,T=200
C=-5	0.064	0.06	0.0663	0.0607	0.0671	0.0625
C=5	0.0422	0.0425	0.0512	0.0395	0.0601	0.0471
C=-10	0.0666	0.0625	0.0608	0.0605	0.0631	0.0559
C=10	0.0422	0.044	0.0538	0.0401	0.0606	0.0476

Table 2.8: Sizes of Specification Test (B-spline) with t(3)

Model	M=20, T=100	M=20, T=200	M=50,T=100	M=50,T=200	M=80,T=100	M=80,T=200
C=-5	0.064	0.06	0.0663	0.0607	0.0671	0.0625
C=5	0.0422	0.0425	0.0512	0.0395	0.0601	0.0471
C=-10	0.0666	0.0625	0.0608	0.0605	0.0631	0.0559
C=10	0.0481	0.0405	0.0637	0.041	0.0536	0.0516

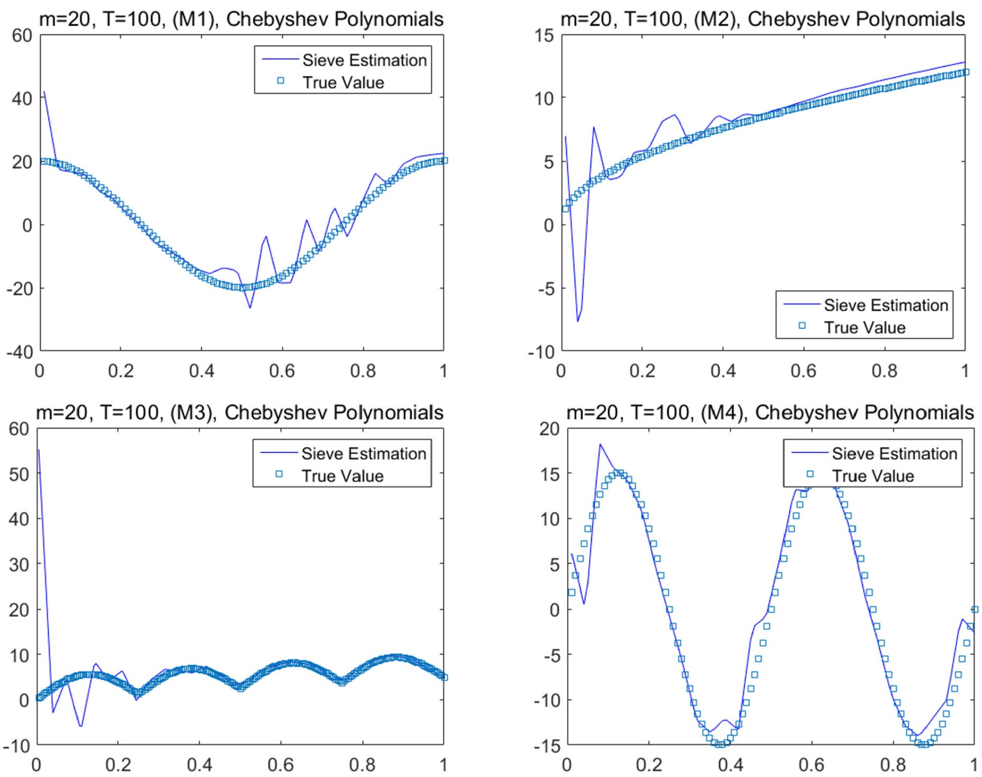

Figure 2.1: Plots of sieve estimates and true values of C(t/T) using the orthogonal trigonometric basis (Chebyshev polynomial)

Empirical illustrations

Time series plots on annual growth rates of a real-estate price index (PI thereafter) for 33 China's Tier-1 and Tier-2 cities are given in the Figure 2.2. The period is from January 2003 to March 2013.

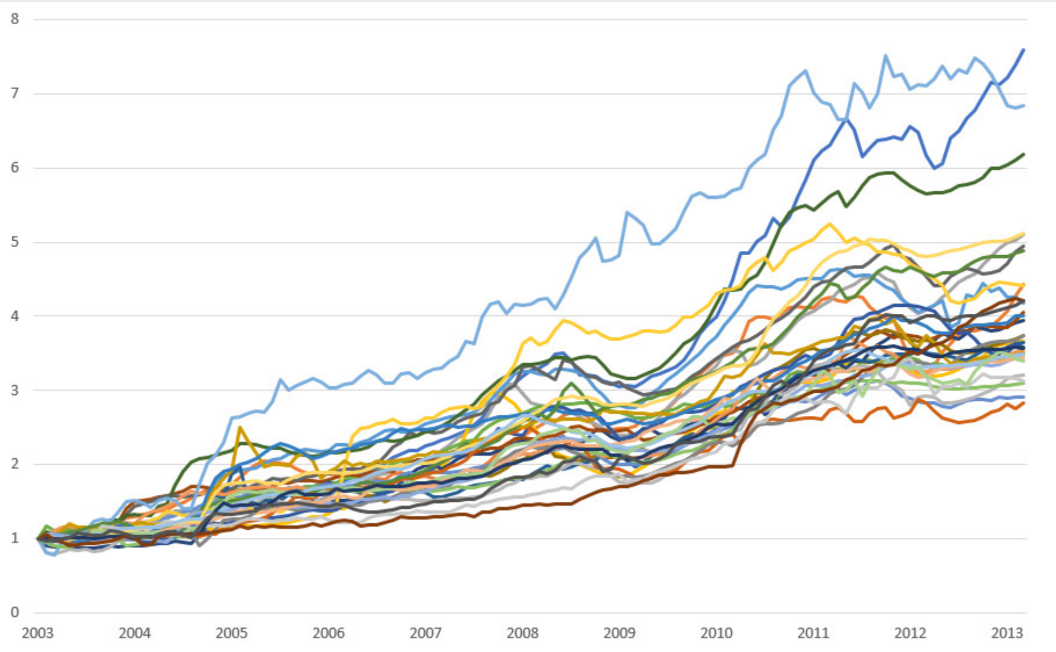

Figure 2.2: Real estate PIs for 33 Tier1 and Tier 2 cities in China (2003m1 - 2013 m3)

Model specification tests on China's real estate markets are implemented for FLUR panel autoregressions as shown in the following table.

Table 2.9: Constancy Specification Test based on Chinese Real-estate Market

Grid	0.2	0.4	0.6	0.7	0.8
Test	-8.2648	8.501	2.9358	7.6516	2.2192

Table 2.10: List of the cities

Tier 1	Beijing, Shanghai, Guangzhou, Shenzhen
Tier 2	Tianjin, Shijiazhuang, Hohhot, Shenyang, Dalian, Changchun, Harbin, Nanjing, Wuxi, Suzhou, Hangzhou, Ningbo, Wenzhou, Hefei, Fuzhou, Xiamen, Nanchang, Jinan, Qingdao, Zhengzhou,Changsha, Nanning, Haikou, Chongqing, Chengdu, Kunming, Xi'an, Xining,Urumq

Chapter 3

Predictive Regression with Time-varying Parameters: A Sieve-IVX Approach

3.1 Introduction

The empirical finance literature has a long history of discussing the predictive regression models in which the stock return is regressed on the lagged value of the fundamental variables. A wide range of macroeconomic and financial variables are considered as potential predictors (e.g., the dividend-price ratio, the default-yield spread, etc.). The earlier empirical studies, including Fama (1981), Campbell & Shiller (1988a,b), Fama & French (1988, 1989), and Fama (1990), adopt the framework of parametric predictive regression models in which the slope coefficients are time-invariant.

However, due to significant changes in market sentiments, the burst of speculative bubbles, rare disasters, and regime switches in monetary and debt management policies, the assumption of constant coefficients is highly suspicious. A series of recent empirical studies have recognized the limitations of time-invariant models and start to accommodate the unstable parameters in their discussions. For instance, Timmermann (2008) suggests that the local predictability is detected while the asset returns are not predictable for most periods. Dangl et al. (2012) evaluate the time-varying predictive regression models in a comprehensive Bayesian framework and show that the models with time-varying coefficients dominate those with constant coefficients. Pyun (2019) introduces an out-of-sample forecasting strategy for monthly market returns using the time-varying predictive regression model. Besides, the predictive regression model with time-varying coefficients employs current information to forecast future stock returns. It has been widely applied to study the lack of predictability in the field of empirical asset pricing (Ang & Behaert, 2007; Avramov, 2002; Lettau & Van Nieuwerburgh, 2008; Dangel et al., 2012; Henkel et al., 2011). These time-varying approaches have led to some novel stylized facts on the predictability of stock returns. However, the above attempts initiated by the empirical researchers illustrate the necessity of considering the parameter instability in predictive regression model. But they still demonstrate several shortcomings to be resolved. The adopted procedures are more interested in estimation and forecast, while lack solid theoretical foundations to facilitate rigorous inference procedures. The statistical theory of the abovementioned procedure is also underdeveloped, which requires more asymptotic treatments.

The above empirical studies motivate us to consider the following preliminary simulated experiment in the presence of time-varying predictability. We simulate $1,000$ sample paths from the data generating process (DGP) $y_t = B\left(\frac{t}{T}\right) x_{t-1} + u_{0t}$, $x_t = x_{t-1} + u_{xt}$, in which a variety of functional forms in $B(\cdot)$ are explored. Assume

$$\begin{pmatrix} u_{0t} \\ u_{xt} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.95 \\ -0.95 & 1 \end{pmatrix} \right).$$

We further consider the OLS-Wald test under the null hypothesis $\mathcal{H}_0^a : B(0.5) = 0$. The empirical rejection rate under the null hypothesis $\mathcal{H}_0^a$ with nominal size 5% is given in Table 3.1. It is evident that the parametric inference approach produces severe size distortions when the true model contains time-varying coefficients. When we test $\mathcal{H}_0^a$ using the OLS-Wald test, the empirical size is

Table 3.1: Empirical size for OLS estimator with univariate regressor

$B\left(t/T\right)$	$T=25$	$T=50$	$T=75$	$T=100$	$T=150$	$T=200$
0	0.096	0.066	0.066	0.069	0.071	0.066
$\frac{t}{T}-\frac{1}{2}$	0.553	0.684	0.736	0.782	0.835	0.862
$\cos\left(\frac{\pi t}{T}\right)$	0.627	0.724	0.792	0.818	0.856	0.882
$\sin\left(\frac{\pi t}{T}\right)-1$	0.625	0.828	0.936	0.973	0.995	0.999

distant from its nominal level of 5%. It will lead to spurious predictability for most cases. The observation is not surprising as the OLS estimation fails to consider the instability in slopes. The experiment also illustrates the risk of imposing the piecewise-constant coefficients. Even though the time-invariant parameters are assumed for a short horizon (e.g., $T \leq 50$), the perturbation will produce the over-rejection phenomenon and ruin the validity of the parametric inference procedures in the model with structural breaks.

Therefore, a time-varying coefficient regression model is a more suitable candidate for detecting predictability, which calls upon the non-/semi-parametric estimation methods. However, the econometric development of the time-varying predictive phenomenon is only considered by assuming the piecewise constancy in parameters accompanied with multiple structural breaks or the piecewise Lipschitz-continuous functions over small open balls (Gonzalo & Pitarakis, 2012; Demetrescu et al., 2020; Georgiev et al., 2020; Georgiev et al., 2021). In contrast, this paper proposes a semi-parametric predictive regression model with a time-varying slope evolving smoothly over time and we allow for regressors with different degrees of persistence. Without imposing constant assumption in parameters across arbitrary time horizons, our model has the flexibility to test for time-varying predictability. The modeling strategy intends to reduce the size distortion caused by parameter instability. This paper employs the nonparametric sieve method to estimate the proposed predictive regression model. The sieve method can be conducted via various basis functions (e.g., polynomials, B-splines, etc.). It can also extend to the multivariate model without introducing the degenerate signal matrix.

The literature of nonlinear and nonparametric cointegration models is closely related to our discussions of time-varying predictive regression model. Specifically, various nonlinear and nonparametric cointegration models have been estimated via kernel methods (Park & Phillips, 2001; Cai et al., 2009; Wang & Phillips, 2009ab; Xiao, 2009; Gao & Phillips, 2013; Li et al., 2016; Phillips et al., 2017). Unfortunately, the usual asymptotic theory of kernel estimation completely breaks down for non-/semi-parametric cointegration model (Phillips et al., 2017). The failure is due to the singular signal matrix induced by kernel smoothing and nonstationary regressors. Instead, the sieve method via global approximation is employed as an alternative in the cointegration models. For instance, the trigonometric basis functions (Park & Hahn, 1999) and Hermite polynomials (Dong & Gao, 2019; Dong, Gao & Peng, 2019, 2020; Dong & Linton, 2018; Dong, Linton & Peng, 2020) are considered for non-/semi-parametric estimations in the multivariate models with nonstationary regressors. However, the available results of cointegration models cannot be applied directly in this paper, since they only consider the unit root behaviors.

Most of the research works containing time-varying coefficients adopt highly restrictive model setups, as only local-to-unity regressors (Cai et al., 2014; Georgiev et al., 2018) or locally stationary regressors (Yousuf et al., 2020) are considered. Comparatively, it is widely recognized that the mixed-root phenomenon exists in the predictive regression model (Phillips, 2014, 2015; Kostakis et al., 2015; Phillips & Lee, 2013, 2016; Yang et al., 2020; Lin & Tu, 2020; Tu & Wang, 2021). Correspondingly, this paper discusses the semi-parametric predictive regression model with mixed-root regressors. We propose a new variant of the self-generated instrument, called the Sieve-IVX instrument, to estimate the time-varying slope. Considering the square integrability of the slope coefficient, a set of basis functions (e.g., polynomials, splines, wavelet) can achieve consistent approximation when the dimension of basis functions increases. Therefore, the weights of basis functions can be estimated using the standard IVX estimation (Phillips & Magdalinos, 2009), which help establish the nonparametric estimation. This paper shows the consistency of the Sieve-IVX estimator in various norms. Based on the pointwise convergence asymptotics, this paper also provides the Sieve-IVX-Wald test for predictability and model specification at a finite set of time points. Under the null hypothesis of no predictability, the Sieve-IVX-Wald tests converge to the chi-square distributions as T goes to infinity.

The simulation results show the finite-sample performance of the Sieve-IVX-Wald test in terms of empirical size and power. In both univariate and multivariate predictive regression models, the Sieve-IVX-Wald test excellently controls the empirical size around the nominal level. Although the power performance is sacrificed in some sense, its power function still approaches unity when the local alternative deviates from the null hypothesis. Comparatively, severe size distortions are observed for the OLS-Wald and the IVX-Wald tests, corresponding to the primary motivation to capture the time-varying predictability in this paper.

The remainder of this chapter is structured as follows. Section 2 presents the model setup and assumptions. Section 3 introduces the Sieve-IVX estimator based on the arbitrary basis functions. Section 4 provides the asymptotic theory of the proposed estimator and the corresponding Sieve-IVX-Wald tests. Section 5 shows simulated results. Section 6 concludes.

Throughout the paper, we employ the following notations. The $n \times n$ dimensional identity matrix is defined as I_n. For some arbitrary matrix M, we use $\|M\|_\infty$, $\|M\|_2$ and $\|M\|$ to denote the L^∞, L^2 and spectral norms of matrix M. We employ $\to_p$ to denote convergence in probability. We employ $\rightsquigarrow$ to denote the weak convergence in Euclidean and functional spaces. Equivalence in distribution is denoted by $=_d$. If $A \leq c \cdot B$ for some real value c, then $A \lesssim B$. The minimum value between the two arguments is denoted by $\wedge$, e.g., $\alpha \wedge \gamma = \min\{\alpha, \gamma\}$.

3.2 Model Setup

The preditive regression model with time-varying coefficient is given as

$$y_t = B'\left(\frac{t}{T}\right)x_{t-1} + u_{0t},\ t = 1, 2, ..., T, \tag{3.1}$$

where $B\left(\frac{t}{T}\right)$ is an n-dimentional vector $\left(B_1\left(\frac{t}{T}\right), ..., B_n\left(\frac{t}{T}\right)\right)'$ with each entry $B_j\left(\frac{t}{T}\right)$ being a time-varying smooth function defined on $[0, 1]$ for all j. The predictor x_t is an n-dimentional persistent regressor, and u_{0t} is a stationary error term. T denotes the sample size.

The data generating process (DGP) of (3.1) extends the predictive regression model of the time-invariant parameters by allowing for the time-varying coefficients for predictors. In the existing literature, standard predictive regression model accommodates a constant slope coefficient such as $B\left(\frac{t}{T}\right) = B$ for all $t = 1, 2, ..., T$. The usual interest is to test the null hypothesis that y_t is unpredictable by x_{t-1} with a time-invariant parameter across the whole time horizon, that is $\mathcal{H}_0 : B = 0$ against the alternative hypothesis that $\mathcal{H}_1 : B \neq 0$. In contrast, the proposed model contains the time-varying parameter that changes smoothly over time. Comparatively, this paper discusses the temporary predictive phenomenon on the chosen grid points. We are intended to test the null hypothesis

$$\mathcal{H}_{0,t^\star} : B\left(\frac{t^\star}{T}\right) = 0,$$

on the grid time points $t^\star$. Correspondingly, the alternative hypothesis considered in this paper follows

$$\mathcal{H}_{1,t^\star} : B\left(\frac{t^\star}{T}\right) \neq 0,$$

on at least one point of the chosen grids $t^\star$.

Several restrictions need to be imposed on $B(\cdot)$ to ensure that the identification and inference procedures that will be formalized. The assumptions for the coefficients $B(\cdot)$ are provided in the following Assumption 3.1.

Assumption 3.1.

(i) *The slope coefficient* $B_j(\cdot)$ *is bounded on the support* $[0,1]$*: there exist constant values* $\overline{B}$ *and* $\underline{B}$*, such that* $\underline{B} \leq B_j(\cdot) \leq \overline{B}$ *for any* j*;*

(ii) $B_j(\cdot)$ *is* q^{th}*-order differentiable with bounded derivatives on the support* $[0,1]$*:* $\underline{B} \leq B_j^s(\cdot) \leq \overline{B}$ *for any* $1 \leq j \leq n$ *and* $1 \leq s \leq q$*, where* $B_j^s(\cdot)$ *is the* s^{th} *order derivative.*

Assumption 3.1 provides the boundedness and smoothness conditions for the time-varying coefficient and facilitates the sieve expansion which will be discussed later.

To characterize the mixed-root phenomenon in regressors, the n-dimentional vector of predictors x_t follows the autoregressive process with roots in the vicinity of unity, as follows:

$$x_t = R_T x_{t-1} + u_{xt}, \; R_T = I_n + \frac{C}{T^{\alpha}}, \tag{3.2}$$

where $C \, (:= diag\{c_1, c_2, ..., c_n\})$ is an $n \times n$ diagonal matrix, $\alpha \, (:= diag\{\alpha_1, \alpha_2, ..., \alpha_n\})$ is an $n \times n$ diagonal matrix with each diagonal element $\alpha_j \in (0,1]$, and u_{xt} is a n-dimensional stationary linear processe. The DGP of (3.2) considers the mixed-root phenomenon that incorporates four types of persistent regressors as defined in the following assumption.

Assumption 3.2.

(i) *Unit Root (UR):* $\alpha_j > 1$ *or* $c_j = 0$*, for any* $1 \leq j \leq n$*;*

(ii) *Local-to-Unity Root (LUR):* $\alpha_j = 1$ *and* $c_j \in (-\infty, \infty)$*, for any* $1 \leq j \leq n$*;*

(iii) *Mildly Integrated Root (MIR):* $\alpha_j \in (0,1)$ *and* $c_j \in (-\infty, 0)$*, for any* $1 \leq j \leq n$*;*

(iv) *Stationary Root (I(0)):* $\alpha_j = 0$ *and* $c_j \in (-\infty, 0)$*, for any* $1 \leq j \leq n$*.*

The exact degrees of persistence of the fundamental variables cannot be trivially specified. To the best knowledge of us, there is no theoretical establishment in distinguishing among UR, LUR, MIR, and stationary regressors. Moreover, Phillips & Lee (2013, 2016) show that the spurious relationship generated by the persistent regressors induces the endogeneity problem and further leads to a severe size distortion. The size distortions caused by the persistent regressors and the time-varying slope coefficients can dramatically undermine the validity of the inference procedure based on OLS estimation. Therefore, we need a robust treatment irrespective of the exact type of underlying persistence.It perfectly suits the discussions of possible misspecifications.

Moreover, this paper imposes a flexible assumption for regression errors. We assume that the innovation of the predictive regression model of (3.1), u_{0t}, follows a conditional heteroskedastic GARCH process, and the errors of the predictors, u_{xt}, admit the stationary linear processes. The detailed discussions of the innovations are formally presented in Assumption 3.3 below.

Assumption 3.3.

(i) Let $\epsilon_t = (\eta_t, e_t')'$, with η_t and e_t as in (3.4) and (3.5), denote an $\mathbb{R}^{n+1}$-valued martingale difference sequence with respect to the natural filtration $\mathcal{F}_t = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots)$ satisfying

$$\mathbb{E}_{\mathcal{F}_{t-1}}[\epsilon_t \epsilon_t'] = \Sigma_\epsilon \text{ a.s. and } \sup_{t \in \mathbb{Z}} \mathbb{E}\|\epsilon_t\|^{2s} < \infty, \tag{3.3}$$

for some $s > 1$, where Σ_ϵ is a positive definite matrix.

(ii) The process $\{u_{0t}\}_{t\in\mathbb{Z}}$ admits the following GARCH(p_1,p_2) representation,

$$u_{0t} = H_t^{1/2}\eta_t, \; H_t = \varphi_0 + \sum_{i=1}^{p_1} \varphi_{1i} u_{0,t-i}^2 + \sum_{k=1}^{p_2} \varphi_{2k} H_{t-k}, \tag{3.4}$$

where $\{\eta_t\}_{t\in\mathbb{Z}}$ is a matigale difference sequence with respect to $\mathcal{F}_t$, φ_0 is a constant vector, φ_{1i} and φ_{2k} are nonnegative for all i, k, and $\sum_{i=1}^{p_1}|\varphi_{1i}| + \sum_{k=1}^{p_2}|\varphi_{2k}| < 1$.

(iii) The process $\{u_{xt}\}$ follows a stationary linear process,

$$u_{xt} = \sum_{j=0}^{\infty} F_j e_{t-j}, \tag{3.5}$$

where $\{F_j\}_{j\geq 0}$ is a sequence of constant matrices such that $F_0 = I_n$, $\sum_{j=0}^{\infty} F_j$ has full rank, and $\sum_{j=0}^{\infty} \|F_j\| < \infty$.

Assumption 3.3(i) imposes conditional homoskcedasticity on the martingale difference sequence ϵ_t. Assumption 3.3(ii) allows conditional heteroskedasticity for the prediction error u_{0t}. The conditional heteroskedasticity characterized by the stationary GARCH process suits the setup of most empirical applications. Therefore, Assumption 3.3(ii) facilitates the asymptotic development under the framework of a quite general and realistic setup. Assumption 3.3(iii) accommodates a stationary linear process for u_{xt}.

Based on our model setup, this paper intends to provide a robust inference procedure for testing predictability to solve the over-rejection problems in predictive detection. In the following sections, we construct a Sieve-IVX estimator for the time-varying coefficient in the predictive regression model (3.1) with mixed-root predictors. Based on the proposed estimator, we also derive the corresponding asymptotic properties, including the uniform convergence rate and asymptotic normality.

3.3 Sieve-IVX Estimator

To construct a sieve estimator for $B(t/T)$, we consider the approximation using a sequence of basis functions $\{\phi_i(\cdot)\}_{i=1}^{k}$. The k-dimensional approximation denoted by $B^{(k)}(\cdot)$ can be constructed by $\sum_{i=1}^{k} \beta_{k,i}\phi_i(\cdot)$, where $\beta_{k,i}$ is an $n \times 1$ vector. To simplify the notations, we further combine $\{\beta_{k,i}\}_{i=1}^{k}$ into an $nk \times 1$ coefficient matrix β_k expressed as $\left(\beta'_{k,1}, \beta'_{k,2}, ..., \beta'_{k,k}\right)'$. We also rewrite the k-dimensional basis functions $\{\phi_i(\cdot)\}_{i=1}^{k}$ as a vector $f_k(\cdot)$, which takes the form of $(\phi_1(\cdot), \phi_2(\cdot), ..., \phi_k(\cdot))'$. Therefore, it is equivalent to write the k-dimensional approximation as $B^{(k)}(\cdot) := (f_k'(\cdot) \otimes I_n)\,\beta_k$.

Define $x_{k,t-1} = f_k\left(\frac{t}{T}\right) \otimes x_{t-1}$. The time-varying predictive regression model (3.1) can be rewritten as

$$y_t = \beta_k' x_{k,t-1} + u_{0k,t}, \tag{3.6}$$

where $u_{0k,t} := u_{0t} + \left(B - B^{(k)}\right)'\left(\frac{t}{T}\right) x_{t-1}$. Therefore, the sieve estimator is equivalent to the OLS estimator of model (3.6), and it can eliminate the size distortion problem in the corresponding inference procedure.

Next, we intend to remove the other source of size distortions generated by the persistent regressors via the IVX instrument. The main intuition of the IVX instrument is filtering x_t to generate instruments $\widetilde{z}_t$ with MIR persistence (Phillips & Magdalinos, 2009). The idea of the standard filtering method is to filter the persistent regressor x_t to generate $\widetilde{z}_t$ as

$$\widetilde{z}_t = F\widetilde{z}_{t-1} + \Delta x_t,$$

where F is a filtering coefficient and the Δx_t is first-differenced regressor. When $F = 0_{n\times 1}$, then $\widetilde{z}_t = \Delta x_t$ and the first-differenced operator is applied to remove the persistence in x_t. The above approach can eliminate the endogeneity problem induced by the spurious relationship but it sacrifices the efficiency of estimation by reducing the signal-to-noise ratio. Comparatively, when $F = I_n$, the level data is kept without filtering. In this way, the estimation efficiency is preserved while the endogeneity problem jeopardizes the estimating accuracy.

The IVX instrument derives the MIR persistence between the first-differenced and level data. It ultimately attains the balance between the estimation efficiency and bias control.The instrument, $\widetilde{z}_t$, is defined as

$$\widetilde{z}_t = R_{Tz}\widetilde{z}_{t-1} + \Delta x_t \text{ with } R_{Tz} = I_n + \frac{C_z}{T^\gamma}, \tag{3.7}$$

where $\gamma \in (0,1)$, $C_z = c_z I_n$, $c_z < 0$, and $\widetilde{z}_0 = 0$. To implement the IVX instrument, we have the flexibility of choosing the parameters γ and c_z. Usually, we choose $\gamma = 0.7$, 0.8, or 0.9, and $c_z = -1$, -10. By the definition (3.7), the self-generated instrument is an intertemporal summation of first-differenced regressors Δx_t as $\widetilde{z}_t = \sum_{j=1}^{t-1} R_{Tz}^{t-j} \Delta x_j$. By applying $\widetilde{z}_{t-1}$ as an instrument (IV) for x_{t-1}, we can remove the persistence of the regressors and further eliminate the asymptotic dependence between the numerator and denominator of the estimates (Phillips & Magdalinos, 2009).

However, the original IVX instrument cannot be applied directly to the semi-parametric model setting in (3.1). Instead, considering the model (3.6) in which the regressors are persistent and large-dimensional (e.g., $k \to \infty$), a modified IVX procedure is applicable to eliminate the nonstandard distribution generated by the latent persistency in x_t, which ensures that the pivotal distribution of the testing statistics still works in the new setting (3.6).

Since the regressor has the format $x_{k,t-1} = f_k\left(\frac{t}{T}\right) \otimes x_{t-1}$ in (3.6), we design a Sieve-IVX approach using the following variant of the IVX instrument $\widetilde{z}_{k,t-1} = f_k\left(\frac{t}{T}\right) \otimes \widetilde{z}_{t-1}$. In this way, $\widetilde{z}_{k,t-1}$ can be used as the IV for $x_{k,t-1}$ in model (3.6). It is natural to consider the Sieve-IVX estimator for β_k, which is defined as follows,

$$\widehat{\beta}_k = \left(\widetilde{Z}'_{k,-1} X_{k,-1}\right)^{-1} \widetilde{Z}'_{k,-1} Y = \beta_k + \left(\widetilde{Z}'_{k,-1} X_{k,-1}\right)^{-1} \widetilde{Z}'_{k,-1} U_{0k}, \tag{3.8}$$

where $X_{k,-1} = (x_{k,0}, x_{k,1}, ..., x_{k,T-1})'$, $\widetilde{Z}_{k,-1} = (\widetilde{z}_{k,0}, \widetilde{z}_{k,1}, ..., \widetilde{z}_{k,T-1})'$, and $Y = (y_1, y_2, ..., y_T)'$.

Correspondingly, the Sieve-IVX estimator of the time-varying coefficient $B(t/T)$ in the predictive regression model (3.1) can be defined as

$$\widehat{B}^{(k)}\left(\frac{t}{T}\right) = \left(f'_k\left(\frac{t}{T}\right) \otimes I_n\right) \widehat{\beta}_k. \tag{3.9}$$

Further, in order to establish the asymptotic properties of the Sieve-IVX estimator $\widehat{B}(\cdot)$, we impose Assumptions 3.4 and 3.5 in the following discussion.

Assumption 3.4.

(i) Assume the smallest eigenvalues $\lambda_{\min}\left(f_k(r) f'_k(r)\right)$ *and* $\lambda_{\max}\left(f_k(r) f'_k(r)\right)$

are bounded away from zero uniformly in k *and* $r \in [0,1]$,

$$0 < \underline{c}_{\Phi} \leq \lambda_{\min}\left(f_k\left(r\right) f_k'\left(r\right)\right) \leq \lambda_{\max}\left(f_k\left(r\right) f_k'\left(r\right)\right) \leq \bar{c}_{\Phi} < \infty.$$

(ii) Assume the approximation error to the nonparametric function $B\left(t/T\right)$ *satisfies the uniform convergence rate as follows,*

$$\sup_{0 \leq r \leq 1} \left\| B\left(r\right) - B^{(k)}\left(r\right) \right\| = O\left(k^{-q}\right),$$

where q *is the order of smoothness for the slope coefficient* $B\left(\cdot\right)$.

(iii) There exists a sequence of constants $\zeta\left(k\right)$ *such that* $\sup_{0 \leq r \leq 1} \left\| f_k\left(r\right) \right\| \leq \zeta\left(k\right)$.

Assumption 3.5. *As* $T \to +\infty$, *we assume that* $k \to +\infty$, $\zeta^2\left(k\right) \cdot \frac{k}{T^{1+(\alpha \wedge \gamma)}} \to 0$, *and* $\frac{T^{1+(\alpha \wedge \gamma)}}{k^{2q}} \to 0$.

Assumption 3.4 is typically assumed in the literature of sieve method (Newey, 1997; Chen & Chirstensen, 2015, 2018; Belloni et al., 2015). In particular, the parameter q is related to the smoothness of function $B\left(\cdot\right)$ and determines the accuracy of the corresponding sieves approximation. Besides, as Newey (1997) shows, $\zeta\left(k\right) = O\left(k^{1/2}\right)$ for B-splines and $\zeta\left(k\right) = O\left(k\right)$ for power series. Assumption 3.5 imposes the rate restrictions on the parameters k and $\zeta\left(k\right)$, which depend on the parameters α, γ, and q.

3.4 Asymptotic Theory for Sieve-IVX Estimator

The asymptotic theory for the Sieve-IVX estimator $\widehat{B}^{(k)}\left(t/T\right)$ relies on the properties of $\widehat{\beta}_k$ defined in (3.8). We can rewrite $\widehat{\beta}_k$ as the following decomposition:

$$\begin{aligned}
\widehat{\beta}_k &= \beta_k + \left(\widetilde{Z}'_{k,-1}X_{k,-1}\right)^{-1}\widetilde{Z}'_{k,-1}U_{0k} \\
&= \beta_k + \left(\widetilde{Z}'_{k,-1}X_{k,-1}\right)^{-1}\widetilde{Z}'_{k,-1}U_0 + \left(\widetilde{Z}'_{k,-1}X_{k,-1}\right)^{-1}\sum_{t=1}^{T}\widetilde{z}_{k,t-1}x'_{t-1}\left(B - B^{(k)}\right)\left(\frac{t}{T}\right) \\
&= \beta_k + A_1 + A_2,
\end{aligned}$$

where $U_{0k} = (u_{0k,1}, u_{0k,2}, ..., u_{0k,T})'$ and $U_0 = (U_{01}, U_{02}, ..., U_{0T})'$. The two bias terms are defined as:

$$\begin{aligned}
A_1 : &= \left(\widetilde{Z}'_{k,-1}X_{k,-1}\right)^{-1}\widetilde{Z}'_{k,-1}U_0, \\
A_2 : &= \left(\widetilde{Z}'_{k,-1}X_{k,-1}\right)^{-1}\sum_{t=1}^{T}\widetilde{z}_{k,t-1}x'_{t-1}\left(B - B^{(k)}\right)\left(\frac{t}{T}\right),
\end{aligned}$$

where the term A_1 contributes to the estimation error, while the term A_2 contributes to the approximation error. The convergence rate for A_1 and A_2 are summarized in the following proposition.

Proposition 3.1. *Let Assumptions 3.1∼3.5 hold. Then,*

(i) the estimation error: $\|A_1\| = O_p\left(\sqrt{\frac{k}{T^{1+\alpha\wedge\gamma}}}\right)$;

(ii) the approximation error: $\|A_2\| = O_p\left(k^{-q}\right)$;

(iii) the bias order for $\widehat{\beta}_k$*:* $\|\widehat{\beta}_k - \beta_k\| = O_p\left(\sqrt{\frac{k}{T^{1+\alpha\wedge\gamma}}} + k^{-q}\right)$.

As expected, the estimation error increases as the number of the sieve basis functions, k, increases. In contrast, the approximation error decreases exponentially as k increases. This comparison illustrates the trade-off between controlling the estimation error and reducing the approximation error. In addition, the estimation error also depends on the scaling parameters α and γ. When $\gamma < \alpha$, the Sieve-IVX instrument reduces the persistence in the regressor x_t and generates the convergence rate $\sqrt{k/T^{1+\gamma}}$ for the estimator $\widehat{\beta}_k$; When $\gamma \geq \alpha$, the regressor x_t has a lower order of persistence and generates the convergence rate $\sqrt{k/T^{1+\alpha}}$ in estimating the pseudo-true value β_k. Comparatively, the fact that the convergence rate of the Sieve-IVX estimator is slower than the rate of the parametric IVX estimator, as $\sqrt{1/T^{1+\alpha\wedge\gamma}}$, shows that the estimating efficiency is sacrificed for the robust model specifications. Finally, the convergence rate of the Sieve-IVX estimator for β_k follows immediately from Proposition 3.1 (i) and (ii).

The asymptotic negligibility of both terms is sufficient to prove the consistency of the Sieve-IVX estimators. Following Newey (1997), we provide the results of the mean-square and uniform convergence rates of the Sieve-IVX estimator in the following theorem.

Theorem 3.1. *Let Assumptions 3.1∼3.5 hold. As $T \to \infty$,*

(i) $\frac{1}{T}\sum_{t=1}^{T}\left\|\widehat{B}^{(k)}\left(\frac{t}{T}\right) - B\left(\frac{t}{T}\right)\right\|^2 = O_p\left(\frac{k}{T^{1+\alpha\wedge\gamma}} + k^{-2q}\right)$,

(ii) $\sup_{1\le t\le T}\left\|\widehat{B}^{(k)}\left(\frac{t}{T}\right) - B\left(\frac{t}{T}\right)\right\| = O_p\left(\zeta(k)\left(\sqrt{\frac{k}{T^{1+\alpha\wedge\gamma}}} + k^{-q}\right)\right)$.

The mean-square rate obtained here is different than that of Newey (1997). First, since the variable that influences the value of the slope coefficient is the time index, t, we justify the L^2-rate for the sample mean square error (MSE) of the slope coefficient rather than the population MSE. Also, the derived convergence rate of the Sieve-IVX estimator depends on the nuisance parameters α and γ, a unique property for the Sieve-IVX estimators.

The uniform convergence rate is derived based on Newey (1997) and cannot attain the optimal rate of the sieve estimations. Comparatively, some more recent results, such as Belloni et al. (2015) and Chen & Christensen (2015, 2018), establish the optimal uniform rate for the sieve estimators and discuss the types of basis functions that can attain such optimal rate. As far as our concern is to ensure that the convergence result still holds under the supreme norm, the technical tools of Newey (1997) are sufficient for our discussions.

Moreover, considering the functional coefficients that vary over time, the predictability needs to be tested on the grid points of a time-series path. To test for the predictive ability of the regressors, we consider a pointwise trajectory,

$$\prod B := \left(B'(r_1), B'(r_2), ..., B'(r_d)\right)',$$

where $\{r_j\}_{j=1}^d$ are the selected grid points on $[0,1]$. The pointwise trajectory, $\prod B^{(k)}$ is represented as

$$\begin{aligned}\prod B^{(k)} &= \left(B^{(k)\,\prime}(r_1), B^{(k)\,\prime}(r_2), ..., B^{(k)\,\prime}(r_d)\right)' \\ &= \left[\left(f_k(r_1), f_k(r_2), ..., f_k(r_d)\right)' \otimes I_n\right]\beta_k \\ &= P_k\beta_k,\end{aligned}$$

where $P_k := \left(f_k(r_1), f_k(r_2), ..., f_k(r_d)\right)' \otimes I_n$. Therefore, the pointwise trajectory $\prod B$ is the argument of interest in establishing the inference procedures for testing predictability.

The convergence rate for $\prod \widehat{B}^{(k)}$ follows directly from Theorem 3.1 and is shown in the following corollary.

Corollary 3.1. *Let Assumptions 3.1∼3.5 hold. As $T \to \infty$,*

$$\prod \widehat{B}^{(k)} - \prod B = O_p\left(\sqrt{\frac{k}{T^{1+\alpha\wedge\gamma}}} + k^{-q}\right),$$

for a finite number of grid points $\{r_j\}_{j=1}^d$.

To state the aymptotic normality results for $\prod \widehat{B}^{(k)}$, it is useful to define the components of the asymptotic variance formula. Let

$$M_k := P_k\left(\widetilde{Z}'_{k,-1}X_{k,-1}\right)^{-1}\left(\widetilde{Z}'_{k,-1}\widetilde{Z}_{k,-1}\right)\left(X'_{k,-1}\widetilde{Z}_{k,-1}\right)^{-1}P'_k,$$

and $\Sigma_{00} := \mathbb{E}\left[u_{0t}^2\right]$. Following Newey (1997), this paper provides the pointwise asymptotic normality of the Sieve-IVX estimator in the following theorem.

Theorem 3.2. *Let Assumptions 3.1∼3.5 hold. As $T \to \infty$, it is shown that*

$$M_k^{-\frac{1}{2}}\left(\prod \widehat{B}^{(k)} - \prod B\right) \rightsquigarrow \mathcal{N}\left(0, \Sigma_{00}\cdot I_{nd}\right),$$

for a finite number of grid points $\{r_j\}_{j=1}^d$.

Theorem 3.2 includes the pointwise test as a specific case, in which we just need to choose one point $r \in [0,1]$. When the null hypothesis holds as $\mathcal{H}_0 : B(r) = 0$, the corresponding Sieve-IVX-Wald test induces a pivotal distribution in testing predictability at the given time point $r \in [0,1]$. In addition, we can also establish the Hausman-type model specification test with a pivotal distribution under the corresponding null hypothesis. The abovementioned Sieve-IVX-Wald statistics are discussed in Theorems 3.3 and 3.4.

Theorem 3.3 (Sieve-IVX-Wald Test for Testing Predictability)**.** *Let Assumptions 3.1∼3.5 hold. Under null hypothesis* $\mathcal{H}_0 : \prod B = 0$*, the test statistic*

$$\left(\prod \widehat{B}^{(k)}\right)' \left(\widehat{\Sigma}_{00} M_k\right)^{-1} \left(\prod \widehat{B}^{(k)}\right) \rightsquigarrow \chi^2\left(nd\right),$$

where $T \to \infty$*, and* $\widehat{\Sigma}_{00}$ *is any consistent estimator for* Σ_{00}*.*

Theorem 3.4 (Sieve-IVX-Wald Test for Model Specification)**.** *Let Assumptions 3.1∼3.5 hold. Under null hypothesis* $\mathcal{H}_0 : \prod B = B \cdot I_d$*, the test statistic*

$$\left(\prod \widehat{B}^{(k)} - \widehat{B}^{IVX} \cdot I_d\right)' \left(\widehat{\Sigma}_{00} M_k\right)^{-1} \left(\prod \widehat{B}^{(k)} - \widehat{B}^{IVX} \cdot I_d\right) \rightsquigarrow \chi^2\left(nd\right),$$

where $T \to \infty$*,* $\widehat{B}^{IVX}$ *is the parametric IVX estimation for the slope, and* $\widehat{\Sigma}_{00}$ *is any consistent estimator for* Σ_{00}*.*

Theorems 3.3 and 3.4 have the potentials to detect the temporary local predictability and conduct model specifications on a finite set of grid points. Unfortunately, the pointwise convergence is sometimes insufficient for the empirical analysis as the finite number of grid points have zero measure on the whole time horizon. The ideal method is to develop a uniform inference procedure to detect the predictability based on the whole time interval. Some recent papers provide results for uniform inference procedures based on the nonparametric series estimators under the independent errors or the mixingales of the time-series settings (see Belloni et al., 2015; Li & Liao, 2020). However, the existing approaches do not allow trivial extensions to the Sieve-IVX estimator. The concerns are given as follows. In the cases of the independent and identically distributed errors, or the stationary mixingales, the nonparametric estimators have the signal matrices with nonrandom limits. The strong approximation is only needed for the Gaussian process of the nonparametric numerator. However, in the nonparametric predictive regression with persistent regressors, both the numerator and denominator converge weakly to nonstandard distributions. A new methodology that employs the empirical process theory is called upon for incorporating this new scenario. Further developments on the uniform inference procedures for the time-varying predictive regression model remain still an open question and will be left for future research.

The number of basis functions k, namely the tuning parameter, has to be appropriately chosen. If not selected properly, the Sieve-IVX estimator becomes inconsistent or derives a slower convergence rate. The requirements for choosing the tuning parameter, k, are given in the following way. First, the number k needs to be large enough; Otherwise, the approximation error will dominate the Gaussian-distributed estimation error. Meanwhile, the number k needs to diverge at a slower speed than the parametric convergence rate. Thus, the rate restriction for k should satisfy the following conditions:

$$\frac{k}{T^{1+\alpha\wedge\gamma}} + \frac{T^{\frac{1+\alpha\wedge\gamma}{2q+1}}}{k} \to 0,$$

where the parameter γ is selected by researchers. Theoretically, the optimal choice of the tunning parameter k is given by

$$k \asymp T^{\frac{1+\alpha\wedge\gamma}{2q+1}}, \tag{3.10}$$

where the notation $A \asymp B$ denotes the case in which $A = O_p(B)$.

Unfortunately, the optimal choice for k is unavailable since the parameter α is usually latent and unobserved. To facilitate the empirical discussion, we follow the convention and employ the leave-one-out cross-validation (CV), which minimizes the asymptotic mean-squared-error (AMSE) criterion function to choose the optimal k. A similar procedure has been used in the kernel estimations of the semi-parametric time-varying predictive regression model (Li et al., 2016). In the next section, extensive simulation results also verify the validity of the leave-one-out CV approach.

3.5 Simulation Results

We conduct simulations to evaluate the finite-sample performances of Sieve-IVX-Wald test statistics on predictability detections.

3.5.1 Univariate regression

For the univariate case, the data generating process is defined as follows

$$\begin{aligned} y_t &= B(t/T)\, x_{t-1} + u_{0t}, \\ x_t &= R_T x_{t-1} + u_{xt}, \quad R_T = 1 + \frac{C}{T^{\alpha}}, \end{aligned} \tag{3.11}$$

where y_t and x_t are scalars. For $t = 1, ..., T$, the innovation sequence u_{0t} admits a GARCH (1,1) representation:

$$u_{0t} = H_t^{1/2} \eta_t, \quad H_t = \varphi_0 + \varphi_1 u_{0,t-1}^2 + \varphi_2 H_{t-1} \tag{3.12}$$

where $\eta_t \overset{i.i.d.}{\sim} N(0,1)$. We set $\varphi_0 = 1, \varphi_1 = 0.2$ and $\varphi_2 = 0.3$, respectively. The innovation sequence for x_t is also assumed to follow a standard normal distribution, $u_{xt} \overset{i.i.d.}{\sim} N(0,1)$. Denote the contemporaneous correlation coefficient between η_t and u_{xt} to be $\rho\, (:= \mathbb{E}[\eta_t u_{xt}])$. In the simulation study, we consider the case in which the correlation coefficient $\rho = -0.95$, whose value is consistent with the empirical findings in the financial market.

To compare with the parametric estimators, we employ a class of functional forms $B(t/T)$. Both the time-invariant coefficient (Model 1) and time-varying coefficients (Models 2 and 3) are considered in our simulation study:

$$\begin{aligned} &\text{Model 1: } B(t/T) = 0, \\ &\text{Model 2: } B(t/T) = \frac{t}{T} - \frac{1}{2}, \\ &\text{Model 3: } B(t/T) = \sin\left(\prod \frac{t}{T}\right) - 1. \end{aligned}$$

This section conducts the Monte Carlo simulation using $1,000$ repetitions with values $C \in \{0, -10\}$, $\alpha = 0.7$, $\rho = -0.95$, different functional form of coefficient $B(t/T)$ (e.g., Models 1$\sim$3), and the sample size $T = 200$.

For each of the simulated paths, we obtain the Sieve-IVX estimator by applying the k-dimensional series to approximate the time-varying function $B(t/T)$. The IVX instrument is constructed as in equation (3.7) by normalizing the parameter C_z to be -10. The IVX rate parameter γ is set as $\{0.5, 0.6, 0.7, 0.8, 0.9\}$ respectively. The optimal bandwidth k is chosen by minimizing the leave-one-out Cross-Validation (CV). Since this section intends to demonstrate the superiority of the proposed Sieve-IVX estimator over the competing counterparts, this paper applies the Sieve-IVX, Sieve-OLS, IVX, and OLS estimators for model comparisons in Models 1$\sim$3.

To evaluate the performance of the Sieve-IVX-Wald test, we focus on testing the predictability at the midpoint of time (i.e., $B(0.5)$). To be more specific, we employ the null hypothesis H_0: $B(0.5) = b$, where $b = 0$ in Models 1$\sim$3. The simulation results regarding the empirical size with a univariate regressor are shown in Table 3.2. In all the simulation studies, we choose the nominal size to be 0.05.

First, we observe for the unit root case ($C = 0$). The Sieve-IVX-Wald test in which $\gamma \leq 0.8$ has excellent size controls across all the functional forms of $B(\cdot)$. In the MIR case ($C = -10$, $\alpha = 0.7$), the size distortions are relatively lower and can be entirely eliminated by choosing a smaller value of γ, say $\gamma = 0.5$.

Next, we consider the comparison of size performance between Wald tests based on Sieve-IVX and IVX estimators. Evidently, the Sieve-IVX procedure demonstrates huge advantages of size control over its counterpart under the null hypothesis across different functional forms. For all choices of $B(t/T)$ and γ, the empirical sizes of the Sieve-IVX -Wald tests are close to 0.05. However, the IVX-Wald tests only have size control for the time-invariant coefficient case ($B(t/T) = 0$). They show severe size distortions for the time-varying models, which range from 0.1 to 0.9. The findings are not surprising as the parameter instability will inevitably lead to high rejection rates under the null. The above comparison corresponds to the motivation of this paper, considering the time-varying features of predictability. When the parameter instability cannot be omitted, the Sieve-IVX procedure will reduce the chance of making Type I error and obtain more reliable decisions of predictability. Besides, the Sieve-OLS-Wald test also exhibits size distortions, especially when the regressor has high persistency ($C = 0$).

In all, this paper finds that the Sieve-IVX-Wald test shows evident improvements over the size controls under the null hypothesis of both parameter instability and persistent property. The above numerical experiment again demonstrates the necessity of proposing the doubly robust method with both the self-generated instrument and the nonparametric estimation.

This paper also examines the following sequence of local alternatives to show the power performance of the Sieve-IVX-Wald tests:

$$B\,(0.5) = \frac{\delta}{T}, \text{ for } \delta \in [-10, 10].$$

This section computes the local power of the Sieve-IVX-Wald tests given the above local alternatives. It is obvious that $\delta = 0$ implies $B\,(0.5) = 0$, which corresponds to the empirical size of each test. Here, Figure 3.1 presents the local power plot for the time-varying coefficient $B\,(t/T) = \sin\,(\pi t/T) - 1$ in which the parameters $C = 0$, $\rho = -0.95$, and sample size $T = 200$. The plot includes the power comparison for different values of $\gamma \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$. The results show that as γ increases, the power of the Sieve-IVX-Wald test improves. Since the tests have slight size distortions for the case in which γ is large, this paper recommends a moderate value of γ in the empirical study, say $\gamma = 0.7$ or 0.8.

3.5.2 Multivariate regression

This subsection further explores the finite sample performance of the Sieves-IVX Wald test under the multivariate case. In the following discussion, we assume the number of predictors to be three. Particularly, the data generating process is imposed as follows

$$\begin{aligned} y_t &= B'(t/T)\, x_{t-1} + u_{0t}, \\ x_t &= R_T x_{t-1} + u_{xt}, \quad R_T = I_3 + \frac{C}{T^{\alpha}}, \end{aligned} \tag{3.13}$$

where y_t is a scalar and x_t is a 3×1 vector which denotes the three-dimensional predictors. In the simulation study, we consider the unit roots case ($C = diag\,(0, 0, 0)$) and mixed roots case ($C = diag\,(0, -10, -50)$, corresponding to the case of a unit root, a MIR, and a stationary predictor). We choose $\alpha = 0.7 \cdot I_3$ for both cases. The innovation u_{0t} admits the same GARCH (1,1) reprentation defined in (3.12). Denote the collection of the error term $\zeta = [\eta, u'_{xt}]'$, which is assumed to follow $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. In the simulaiton study, we set the variance of each error term to be unity and the correlation coefficients $\Sigma_{12} = -0.75$, $\Sigma_{13} = -0.75$, $\Sigma_{14} = -0.75$, $\Sigma_{23} = 0.7$, $\Sigma_{24} = 0.5$ and $\Sigma_{34} = 0.5$. The coefficient of the predictive regression $B\,(t/T)$ is a three-dimensional function. Similar to the univariate case, we employ a class of functional forms for each element of the slope coefficient.

We conduct the Monte Carlo simulation using $1,000$ repetitions with different values of C, different functional forms of time-varying coefficient $B\,(t/T)$, and sample size $T = 600$. For each of the simulated datasets, we examine the empirical size for the Wald test of coefficient at the midpoint of time. To be specific, we consider the joint Wald test under the null hypothesis that all three coefficients are jointly equal to zero at the midpoint of time. That is, H_0: $B\,(0.5) = (0, 0, 0)$. The Wald tests are constructed based on four estimators, including the proposed Sieve-IVX, Sieve-OLS, IVX, and OLS estimators.

Table 3.3 presents the empirical size of the Wald tests in the multivariate regression model. First, we focus on the Sieve-IVX-Wald test with various choices of γ. With the unit-root regressors ($C = diag\,(0, 0, 0)$), the size distortions caused by the $I(1)$ persistence are relatively lower and can be entirely removed by choosing a smaller value of γ. When the mixed-root regressors ($C = diag\,(0, -10, -50)$ are considered, we also find a slight oversizing phenomenon when γ is large. The empirical rejection rate peaks around 12% and doubles the chance of rejecting the null hypothesis. Therefore, this paper suggests choosing the value of γ to be 0.5 to cure the size distortion. On the other hand, even with a small value of γ, the IVX-Wald test only has size controls in the case of the time-invariant coefficient but becomes extremely oversized in the case of the time-varying coefficients, e.g., $B\,(t/T) = \sin\left(\frac{\pi t}{T}\right) - 1$. Similarly, the Sieve-OLS-Wald test also exhibits size distortions, especially when including more persistence regressors.

We also examine the power performance of the Sieve-IVX-Wald tests. Under the joint null hypothesis H_0: $B\,(0.5) = (0, 0, 0)$, we consider the following sequence of local alternatives to show the power performance of the Sieve-IVX-Wald tests:

$$B\,(0.5) = \frac{\delta}{\sqrt{T}}\,(1, 1, 1)\,, \text{ for } \delta \in [-10, 10], \tag{3.14}$$

with $\delta = 0$ corresponding to the empirical size of the test. Local power functions are computed with three $I(1)$ regressors and the sample size $T = 600$. Figure 3.2 presents the local power plots of the Sieve-IVX-Wald tests. The plot includes the power function for different values of γ and shows a similar pattern as in the univariate case.

3.6 Conclusion

Recent research has suggested that the stock return predictability via economic fundamentals shows time-varying feature. So the inference procedures dealing with constant parameters or piecewise-constant parameters can result in severe size distortions. Our motivation for this paper is to establish a predictability test, which is robust in terms of the parameter instability and the mixed-root phenomenon in regressors. In order to eliminate the spurious predictability induced by model misspecification and the endogenous regressors, our Sieve-IVX approach that combines the IVX estimation and the sieve approximation shows robust property. The tests rely on the self-generated instrument (Phillips & Magdalinos, 2009) and the sieve estimation (Andrews, 1991a; Newey, 1997; Phillips, 1998, 2002, 2007; Chen, 2007; Phillips & Liao, 2012; Belloni et al., 2015; Chen & Christensen, 2015, 2018; Dong & Gao, 2019; Dong et al., 2019, 2020; Li & Liao, 2020). Intuitively, the endogeneity problem of the predictive regression results from the spurious relationship induced by the near-unity persistent regressors, while the self-generated instrument eliminates the problem by controlling the persistence of fundamental variables. The employed sieve method attains the standard asymptotic normality. Since the limiting distributions of the corresponding statistics follow the pivotal distributions, the critical values of the chi-square distribution can be used for predictability inference. The above asymptotic theory holds for a broad class of innovations. Monte Carlo simulation results demonstrate that the proposed tests show size controls with finite sample, as well as satisfactory power performance in detecting temporary predictability.

This paper concludes with two suggestions for future research. First, the uniform inference procedure is called upon for testing predictability as the current method is capable of testing a finite set of grid points whose measure is zero on the whole time interval $[0, 1]$. Unfortunately, the extension from the predictive regression models with stationary regressors (Li & Liao, 2020) to those with the nonstationary regressors is not trivial, since the strong approximation to the non-Gaussian distribution is still an open question to the best knowledge of us. Second, the current framework only incorporates a finite number of regressors, while the discussion of large-dimensional regressors is of general interest. Very recently, few papers (Koo et al., 2020; Fan et al., 2021; Lee et al., 2020) provide relevant results of the large-dimensional predictive regression. Currently, the available results of large-dimensional predictive regression consider the mixed-root phenomenon where only the dimension of stationary regressors diverges, while the dimension of nonstationary regressors is fixed. The abovementioned results can be further strengthened by enlarging the dimensions of both stationary and nonstationary regressors and applying the sparsity restriction in parameters. This proposal again raises the technical difficulty of providing a strong approximation to non-Gaussian approximation. Therefore, to push the frontier of the predictive regression models in the context of the large-dimensional predictors, the strong approximation for the non-Gaussian argument of a diverging dimension remains to be developed.

Tables and Figures

Table 3.2: Empirical size under the univariate case ($d = 1$) and sample size $T = 200$

$\beta(t/T)$	Parameter	Est.	IVX					OLS
			$\gamma = 0.5$	$\gamma = 0.6$	$\gamma = 0.7$	$\gamma = 0.8$	$\gamma = 0.9$	
0	$C = 0$	Sieves	0.029	0.043	0.046	0.046	0.066	0.148
		Linear	0.053	0.068	0.07	0.073	0.065	0.064
	$C = -10$	Sieves	0.048	0.059	0.072	0.082	0.084	0.121
	$\alpha = 0.7$	Linear	0.067	0.082	0.088	0.081	0.074	0.068
$\frac{t}{T} - \frac{1}{2}$	$C = 0$	Sieves	0.032	0.044	0.058	0.059	0.076	0.183
		Linear	0.126	0.215	0.339	0.468	0.562	0.867
	$C = -10$	Sieves	0.050	0.059	0.073	0.080	0.086	0.130
	$\alpha = 0.7$	Linear	0.050	0.065	0.085	0.113	0.167	0.345
$\sin\left(\pi\frac{t}{T}\right) - 1$	$C = 0$	Sieves	0.031	0.044	0.058	0.058	0.062	0.121
		Linear	0.496	0.694	0.798	0.863	0.897	0.998
	$C = -10$	Sieves	0.068	0.081	0.071	0.074	0.080	0.083
	$\alpha = 0.7$	Linear	0.605	0.820	0.933	0.981	0.987	0.997

Table 3.3: Empirical size under multivariate case ($d = 3$) and sample size $T = 600$

$\beta(t/T)$	Parameter	Est.	IVX					OLS
			$\gamma = 0.5$	$\gamma = 0.6$	$\gamma = 0.7$	$\gamma = 0.8$	$\gamma = 0.9$	
0	$C = 0$	Sieves	0.013	0.030	0.047	0.073	0.095	0.442
		Linear	0.034	0.049	0.062	0.071	0.074	0.166
	$C = -10$	Sieves	0.036	0.050	0.061	0.064	0.078	0.147
	$\alpha = 0.7$	Linear	0.039	0.050	0.056	0.059	0.070	0.082
$\frac{t}{T} - \frac{1}{2}$	$C = 0$	Sieves	0.018	0.034	0.058	0.089	0.120	0.530
		Linear	0.636	0.834	0.928	0.982	0.993	0.999
	$C = -10$	Sieves	0.038	0.056	0.059	0.066	0.081	0.176
	$\alpha = 0.7$	Linear	0.317	0.502	0.670	0.800	0.869	0.977
$\sin\left(\pi\frac{t}{T}\right) - 1$	$C = 0$	Sieves	0.038	0.079	0.107	0.151	0.181	0.337
		Linear	0.891	0.967	0.99	0.997	0.999	1.000
	$C = -10$	Sieves	0.057	0.076	0.097	0.117	0.124	0.166
	$\alpha = 0.7$	Linear	0.947	0.969	0.977	0.990	0.995	1.000

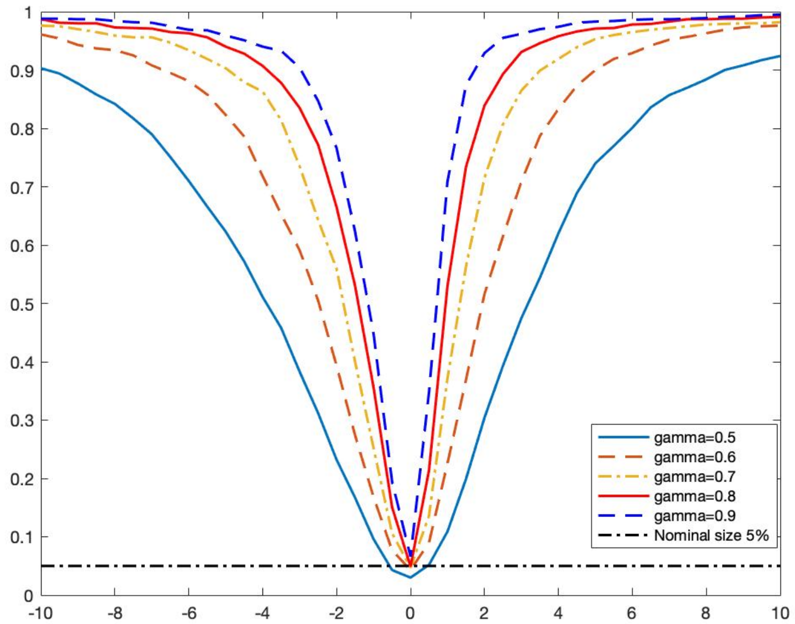

Figure 3.1: Local power function for testing with univariate unit root regressor

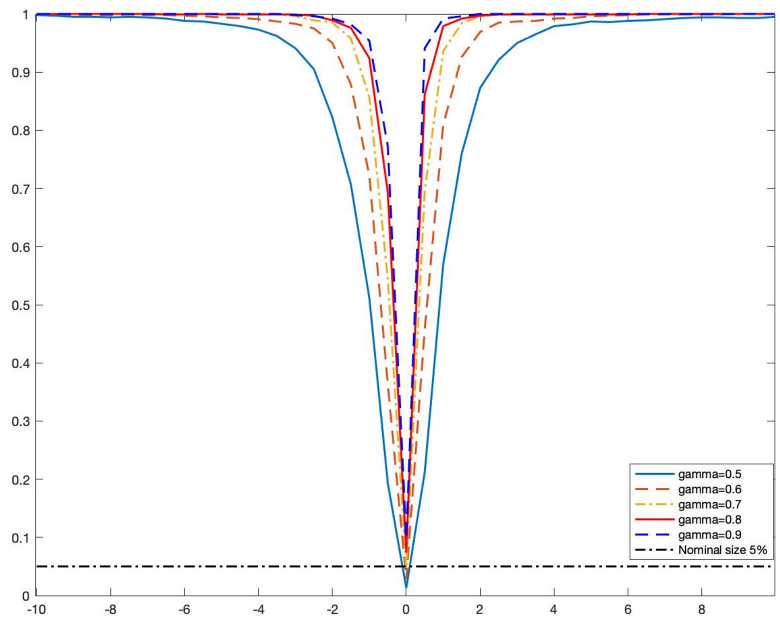

Figure 3.2: Local power function for testing with multivariate regressors

Reference

Ang, A., & Bekaert, G. (2007). Stock return predictability: Is it there?. The Review of Financial Studies, 20(3), 651-707.

Avramov, D. (2002). Stock return predictability and model uncertainty. Journal of Financial Economics, 64(3), 423-458.

Andrews, D. W. (1991a). Asymptotic normality of series estimators for nonparametric and semiparametric regression models. Econometrica: Journal of the Econometric Society, 307-345.

Andrews, D. W. (1991b). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica: Journal of the Econometric Society, 817-858.

Andersen, T. G., & Varneskov, R. T. (2021). Testing for Parameter Instability and Structural Change in Persistent Predictive Regressions (No. w28570). National Bureau of Economic Research.

Belloni, A., Chernozhukov, V., Chetverikov, D., & Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. Journal of Econometrics, 186(2), 345-366.

Bykhovskaya, A., & Phillips, P. C. B (2018). Boundary limit theory for functional local to unity regression. Journal of Time Series Analysis, 39(4), 523-562.

Bykhovskaya, A., & Phillips, P. C. B (2020). Point optimal testing with roots that are functionally local to unity. Journal of Econometrics, 219(2), 231-259.

Bandi, F. M., & Phillips, P. C. B (2003). Fully nonparametric estimation of scalar diffusion models. Econometrica, 71(1), 241-283.

Bekaert, G., Harvey, C. R., & Lundblad, C. (2007). Liquidity and expected returns: Lessons from emerging markets. The Review of Financial Studies, 20(6), 1783-1831.

Bierens, H. J., & Martins, L. F. (2010). Time-varying cointegration. Econometric Theory, 26(5), 1453-1490.

Breitung, J., & Demetrescu, M. (2015). Instrumental variable and variable addition based inference in predictive regressions. Journal of Econometrics, 187(1), 358-375.

Bossaerts, P., & Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: what do we learn?. The Review of Financial Studies, 12(2), 405-428.

Cai, B., Dong, C., & Gao, J. (2015). Orthogonal series estimation in nonlinear cointegrating models with endogeneity. Available at SSRN 2665110.

Cochrane, J. H. (2011). Presidential address: Discount rates. The Journal of Finance, 66(4), 1047-1108.

Campbell, J. Y. (2008). Estimating the equity premium. Canadian Journal of Economics, 41(1), 1-21.

Campbell, J. Y., & Shiller, R. J. (1988a). Stock prices, earnings, and expected dividends. the Journal of Finance, 43(3), 661-676.

Campbell, J. Y., & Shiller, R. J. (1988b). The dividend-price ratio and expectations of future dividends and discount factors. The Review of Financial Studies, 1(3), 195-228.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average?. The Review of Financial Studies, 21(4), 1509-1531.

Campbell, J. Y., & Vuolteenaho, T. (2004). Bad beta, good beta. American Economic Review, 94(5), 1249-1275.

Campbell, J. Y., & Yogo, M. (2006). Efficient tests of stock return predictability. Journal of Financial Economics, 81(1), 27-60.

Chan, N. H., & Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR (1) processes. The Annals of Statistics, 1050-1063.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. Handbook of Econometrics, 6, 5549-5632.

Chen, X., & Christensen, T. M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. Journal of Econometrics, 188(2), 447-465.

Chen, X., & Christensen, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. Quantitative Economics, 9(1), 39-84.

Chen, X., & Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. Journal of Econometrics, 135(1-2), 125-154.

Cai, Z., Li, Q., & Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. Journal of Econometrics, 148(2), 101-113.

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

Cai, Z., Wang, Y., & Wang, Y. (2015). Testing instability in a predictive regression model with nonstationary regressors. Econometric Theory, 953-980.

Dong, C., & Gao, J. (2019). Expansion and estimation of Lévy process functionals in nonlinear and nonstationary time series regression. Econometric Reviews, 38(2), 125-150.

Dong, C., Gao, J., & Peng, B. (2019). Estimation in a semiparametric panel data model with nonstationarity. Econometric Reviews, 38(8), 961-977.

Dong, C., Gao, J., & Peng, B. (2020). Varying-coefficient panel data models with nonstationarity and partially observed factor structure. Journal of Business and Economic Statistics, 1-12.

Dong, C., & Linton, O. (2018). Additive nonparametric models with time variable and both stationary and nonstationary regressors. Journal of Econometrics, 207(1), 212-236.

Dong, C., Linton, O., & Peng, B. (2020). A weighted sieve estimator for nonparametric time series models with nonstationary variables. Journal of Econometrics.

Dangl, T., & Halling, M. (2012). Predictive regressions with time-varying coefficients. Journal of Financial Economics, 106(1), 157-181.

Elliott, D. (1964). The evaluation and estimation of the coefficients in the Chebyshev series expansion of a function. Mathematics of Computation, 18(86), 274-284.

Elliott, G., Müller, U. K., & Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. Econometrica, 83(2), 771-811.

Fama, E. F. (1981). Stock returns, real activity, inflation, and money. The American Economic Review, 71(4), 545-565.

Fama, E. F. (1990). Stock returns, expected returns, and real activity. The Journal of Finance, 45(4), 1089-1108.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. Journal of Financial Economics, 22(1), 3-25.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. Journal of Financial Economics, 25(1), 23-49.

Fang, H., Gu, Q., Xiong, W., & Zhou, L. A. (2016). Demystifying the Chinese housing boom. NBER Macroeconomics Annual, 30(1), 105-166.

Fan, R., Lee, J. H., & Shin, Y. (2021). Predictive Quantile Regression with Mixed Roots and Increasing Dimensions. arXiv preprint arXiv:2101.11568.

Ferson, W. E., & Schadt, R. W. (1996). Measuring fund strategy and performance in changing economic conditions. The Journal of Finance, 51(2), 425-461.

Georgiev, I., Demetrescu, M., Rodrigues, P. M., & Taylor, A. R. (2021). Extensions to IVX Methods of Inference for Return Predictability. Essex Finance Centre Working Paper Series, No 68: 02-2021.

Georgiev, I., Harvey, D. I., Leybourne, S. J., & Taylor, A. R. (2018). Testing for parameter instability in predictive regression models. Journal of Econometrics, 204(1), 101-118.

Gao, J., & Phillips, P. C. B. (2013). Semiparametric estimation in triangular system equations with nonstationarity. Journal of Econometrics, 176(1), 59-79.

Gonzalo, J., & Pitarakis, J. Y. (2012). Regime-specific predictability in predictive regressions. Journal of Business and Economic Statistics, 30(2), 229-241.

Henkel, S. J., Martin, J. S., & Nardari, F. (2011). Time-varying short-horizon predictability. Journal of Financial Economics, 99(3), 560-580.

Juhl, T., & Xiao, Z. (2005). Partially linear models with unit roots. Econometric Theory, 21(5), 877-906.

Kostakis, A., Magdalinos, T., & Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. The Review of Financial Studies, 28(5), 1506-1553.

Koo, B., Anderson, H. M., Seo, M. H., & Yao, W. (2020). High-dimensional predictive regression in the presence of cointegration. Journal of Econometrics, 219(2), 456-477.

Kwiatkowski, D., Phillips, P. C. B, Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. Journal of econometrics, 54(1-3), 159-178.

Kothari, S. P., & Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. Journal of Financial Economics, 44(2), 169-203.

Li, D., Phillips, P. C. B, & Gao, J. (2016). Uniform consistency of nonstationary kernel-weighted sample covariances for nonparametric regression. Econometric Theory, 32(3), 655-685.

Li, D., Phillips, P. C. B, & Gao, J. (2020). Kernel-based inference in time-varying coefficient cointegrating regression. Journal of Econometrics, 215(2), 607-632.

Li, J., & Liao, Z. (2020). Uniform nonparametric inference for time series. Journal of Econometrics, 219(1), 38-51.

Lee, J. H. (2016). Predictive quantile regression with persistent covariates: IVX-QR approach. Journal of Econometrics, 192(1), 105-118.

Lee, J. H., Shi, Z., & Gao, Z. (2021). On LASSO for predictive regression. Journal of Econometrics.

Lettau, M., & Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence: The review of financial studies: Reconciling the return predictability evidence. The Review of Financial Studies, 21(4), 1607-1652.

Lamont, O. (1998). Earnings and expected returns. The Journal of Finance, 53(5), 1563-1587.

Li, Q., & Racine, J. S. (2007). Nonparametric econometrics: theory and practice. Princeton University Press.

Lin, Y., & Tu, Y. (2020). Robust inference for spurious regressions and cointegrations involving processes moderately deviated from a unit root. Journal of Econometrics, 219(1), 52-65.

Moon, H. R., & Phillips, P. C. B (2004). GMM estimation of autoregressive roots near unity with panel data. Econometrica, 72(2), 467-522.

Martins, L. F. (2018). Bootstrap tests for time varying cointegration. Econometric Reviews, 37(5), 466-483.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. Journal of Econometrics, 79(1), 147-168.

Newey, W. K., & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. Econometrica. 55 (3): 703–708.

Plazzi, A., Torous, W., & Valkanov, R. (2010). Expected returns and expected growth in rents of commercial real estate. The Review of Financial Studies, 23(9), 3469-3519.

Paye, B. S., & Timmermann, A. (2006). Instability of return prediction models. Journal of Empirical Finance, 13(3), 274-315.

Park, J. Y., & Hahn, S. B. (1999). Cointegrating regressions with time varying coefficients. Econometric Theory, 15(5), 664-703.

Park, J. Y., & Phillips, P. C. B (2001). Nonlinear regressions with integrated time series. Econometrica, 69(1), 117-161.

Pesaran, M. H., & Timmermann, A. (2002). Market timing and return prediction under model instability. Journal of Empirical Finance, 9(5), 495-510.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. Biometrika, 74(3), 535-547.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. Econometrica: Journal of the Econometric Society, 283-306.

Phillips, P. C. B. (1995). Fully modified least squares and vector autoregression. Econometrica, 63, 1023-1023.

Phillips, P. C. B. (1998). New tools for understanding spurious regressions. Econometrica, 1299-1325.

Phillips, P. C. B. (2002). New unit root asymptotics in the presence of deterministic trends. Journal of Econometrics, 111(2), 323-353.

Phillips, P. C. B. (2007). Unit root log periodogram regression. Journal of Econometrics, 138(1), 104-124.

Phillips, P. C. B. (2014). On confidence intervals for autoregressive roots and predictive regression. Econometrica, 82(3), 1177-1195.

Phillips, P. C. B. (2015). Halbert White jr. memorial jfec lecture: Pitfalls and possibilities in predictive regression. Journal of Financial Econometrics, 13(3), 521-555.

Phillips, P. C. B., Li, D., & Gao, J. (2017). Estimating smooth structural change in cointegration models. Journal of Econometrics, 196(1), 180-195.

Phillips, P. C. B., & Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. Journal of Econometrics, 177(2), 250-264.

Phillips, P. C. B., & Lee, J. H. (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. Journal of Econometrics, 192(2), 433-450.

Phillips, P. C. B., & Liao, Z. (2012). Series Estimation of Stochastic Processes: Recent Developments and Econometric Applications. Cowles Foundation Discussion Papers, (1871).

Phillips, P. C. B., & Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. Econometrica, 67(5), 1057-1111.

Phillips, P. C. B., & Moon, H. R. (2000). Nonstationary panel data analysis: An overview of some recent developments. Econometric reviews, 19(3), 263-286.

Phillips, P. C. B., Moon, H. R., & Xiao, Z. (2001). How to estimate autoregressive roots near unity. Econometric Theory, 17(1), 29-69.

Phillips, P. C. B., & Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. Journal of Econometrics, 136(1), 115-130.

Phillips, P. C. B., & Magdalinos, T. (2009). Econometric inference in the vicinity of unity. Singapore Management University, CoFie Working Paper, 7.

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, 75(2), 335-346.

Phillips, P. C. B., & Solo, V. (1992). Asymptotics for linear processes. The Annals of Statistics, 971-1001.

Pyun, S. (2019). Variance risk in aggregate stock returns and time-varying return predictability. Journal of Financial Economics, 132(1), 150-174.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. The Review of Financial Studies, 23(2), 821–862.

Stock, J. H. (1991). Confidence intervals for the largest autoregressive root in US macroeconomic time series. Journal of Monetary Economics, 28(3), 435-459.

Su, L., & Ju, G. (2018). Identifying latent grouped patterns in panel data models with interactive fixed effects. Journal of Econometrics, 206(2), 554-573.

Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration. Econometric theory, 10(1), 91-115.

Timmermann, A. (2008). Elusive return predictability. International Journal of Forecasting, 24(1), 1-18.

Tu, Y., & Wang, Y. (2021). Spurious functional-coefficient regression models and robust inference with marginal integration. Journal of Econometrics.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies, 21(4), 1455-1508.

Wang, Q., & Phillips, P. C. B (2009a). Structural nonparametric cointegrating regression. Econometrica, 77(6), 1901-1948.

Wang, Q., & Phillips, P. C. B (2009b). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. Econometric Theory, 25(3), 710-738.

Xiao, Z. (2009). Functional-coefficient cointegration models. Journal of Econometrics, 152(2), 81-92.

Yousuf, K., & Ng, S. (2020). Boosting high dimensional predictive regressions with time varying parameters. Journal of Econometrics.

Zaitsev, A. Y. (1998). Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. ESAIM: Probability and Statistics, 2, 41-108.

Appendix A

Technical Results for Chapter 1

Proof of the main results

In this appendix We show the technical proofs for IVX method under FLUR, FMER and FMSR cases.

Proof for FLUR

Lemma A.1. *If Assumptions 1.1 and 1.2 hold, as $T \to \infty$,*

$$\sup_{1\leq t\leq T} \mathbb{E}\left\|\eta^{(1)}_{T,t-1}\right\|^2 = O_p\left(T^{2\gamma+1}\right).$$

Proof of Lemma A.1: Set $x_0 = 0$. The inter-temporal summation of innovations is given as

$$x_{j-1} = \sum_{k=1}^{j-2}\left(\prod_{m=k+1}^{j-1} R_{Tm}\right) u_{xk} + u_{x,j-1}.$$

Based on the inter-temporal accumulations, we have

$$\begin{aligned}
&\left\|\eta^{(1)}_{T,t-1}\right\|^2 \\
= \;& \operatorname{tr}\left\{\sum_{i,j=1}^{t-1} R_{Tz}^{t-i-1} C\left(\frac{i}{T}\right)\left(x_{i-1}x'_{j-1}\right) C\left(\frac{j}{T}\right) R_{Tz}^{t-j-1}\right\} \\
\leq \;& M \sum_{i,j=1}^{t-1}\sum_{k=1}^{j}\sum_{l=1}^{i}\left\|R_{Tz}^{t-j-1} C\left(\frac{i}{T}\right)\right\|_\infty \left\|R_{Tz}^{t-i-1} C\left(\frac{j}{T}\right)\right\|_\infty \left\|R_{Tt}^{(j-k)}\right\|_\infty \left\|R_{Tt}^{(i-l)}\right\|_\infty \left\|u_{xk}u'_{xl}\right\|,
\end{aligned}$$

where

$$R_{Tt}^{(j-k)} := \prod_{m=k+1}^{j} R_{Tm} \text{ and } R_{Tt}^{(i-l)} := \prod_{m=l+1}^{i} R_{Tm},$$

and some $M < \infty$. If Assumptions 1.1 and 1.2 hold for any $i \in \{1, 2, ..., n\}$, we have

$$\frac{1}{T^{\gamma}} \sum_{j=1}^{T} \left\| R_{Tz}^{T-j-1} C\left(\frac{j}{T}\right) \right\|_{\infty} = \int_{0}^{+\infty} \left\| e^{rC_z} C(r) \right\|_{\infty} dr + o(1) < +\infty.$$

Therefore we justify the stochastic order of the term as,

$$\sup_{1 \le t \le T} \sum_{j=1}^{t} \left\| R_{Tz}^{t-j-1} C\left(\frac{j}{T}\right) \right\|_{\infty} = O_p\left(T^{\gamma}\right).$$

Similarly, by Assumptions 1.1 and 1.2, we have

$$\begin{aligned}
& \frac{1}{T} \sup_{1 \le t \le T} \sum_{j=1}^{t-1} \left\| R_{T,t-1}^{(t-j)} \right\|_{\infty} \\
= & \sup_{0 \le r \le 1} \frac{1}{T} \sum_{j=1}^{[Tr]} \left\| R_{T,t-1}^{([Tr]-j)} \right\|_{\infty} \\
= & \frac{1}{T} \sup_{0 \le r \le 1} \sum_{j=1}^{[Tr]} \left\| \exp\left(\sum_{m=j+1}^{[Tr]} C\left(\frac{m}{T}\right) \right) \right\|_{\infty} \\
= & \sup_{0 \le r \le 1} \sum_{j=1}^{[Tr]} \frac{1}{T} \left\| \exp\left(\sum_{m=1}^{[Tr]} C\left(\frac{m}{T}\right) \right) \right\|_{\infty} \left\| \exp\left(-\sum_{m=1}^{j} C\left(\frac{m}{T}\right) \right) \right\|_{\infty} \\
= & \sup_{0 \le r \le 1} \left\| \exp\left(\int_0^r C(s)\, ds \right) \right\|_{\infty} \cdot \left\| \int_0^r \exp\left(\int_0^p C(q)\, dq \right) dp \right\|_{\infty} + o_p(1) < \infty,
\end{aligned}$$

and

$$\sup_{1 \le t \le T} \sum_{j=1}^{t} \left\| R_{T,t-1}^{(t-j)} \right\|_{\infty} = O_p(T).$$

According to the formulation of innovations, we have

$$\sum_{l=-\infty}^{+\infty} \left\| \Gamma_{ux}(l) \right\| < +\infty,$$

where $\Gamma_{ux}(l) = \mathbb{E}\left(u_{xt} u_{xt-l}'\right)$. Combining the asymptotic results above, the stochastic order of IVX residual is justified as

$$\begin{aligned}\sup_{1\le t\le T}\mathbb{E}\left\|\eta_{T,t-1}^{(1)}\right\|^2 &\le \left(\sup_{1\le t\le T}\sum_{j=1}^{t-1}\left\|R_{Tz}^{t-j-1}C\left(\frac{j}{T}\right)\right\|\right)^2\left(\sup_{1\le t\le T}\sum_{k=1}^{t}\left\|R_{Tt}^{(t-k)}\right\|\right)\sum_{l=-\infty}^{+\infty}\left\|\Gamma_{ux}(l)\right\| \\ &= O_p\left(T^{1+2\gamma}\right).\end{aligned}$$

■

Proof of Lemma 1.1: (i) According to Lemma A.1 and the mds property of u_{0t}, it is easy to verify that

$$\sum_{t=1}^{T} z_{t-1}u_{0t} = O_p\left(T^{\frac{1+\gamma}{2}}\right) \text{ and } \sum_{t=1}^{T}\eta_{T,t-1}^{(1)}u_{0t} = O_p\left(T^{1+\gamma}\right).$$

(ii) Based on Lemma A.1,

$$\sum_{t=1}^{T} z_{t-1}z_{t-1}' = O_p\left(T^{1+\gamma}\right) \quad \text{and} \quad \sum_{t=1}^{T}\eta_{T,t-1}^{(1)}\left(\eta_{T,t-1}^{(1)}\right)' = O_p\left(T^{2+2\gamma}\right),$$

as $T\to\infty$. For the covariance term, we have

$$\begin{aligned}\left\|\sum_{t=1}^{T} z_{t-1}\left(\eta_{T,t-1}^{(1)}\right)'\right\| &\le \left(\sum_{t=1}^{T}\left\|z_{t-1}\right\|_2^2\right)^{\frac{1}{2}}\left(\sum_{t=1}^{T}\left\|\eta_{T,t-1}^{(1)}\right\|_2^2\right)^{\frac{1}{2}} \\ &= O_p\left(T^{\frac{3\gamma+3}{2}}\right),\end{aligned}$$

by Cauchy-Schwarz inequality. Therefore we have $\sum_{t=1}^{T} z_{t-1}\left(\eta_{T,t-1}^{(1)}\right)' = O_p\left(T^{\frac{3\gamma+3}{2}}\right)$. By combining above results, we can verify that

$$\begin{aligned}\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\tilde{z}_{t-1}\tilde{z}_{t-1}' &= \frac{1}{T^{1+\gamma}}\sum_{t=1}^{T} z_{t-1}z_{t-1}' + \frac{1}{T^{2+\gamma}}\left[\sum_{t=1}^{T} z_{t-1}\left(\eta_{T,t-1}^{(1)}\right)' + \sum_{t=1}^{T}\eta_{T,t-1}^{(1)}z_{t-1}'\right] \\ &\quad + \frac{1}{T^{3+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}\eta_{T,t-1}^{(1)\prime} \\ &= \frac{1}{T^{1+\gamma}}\sum_{t=1}^{T} z_{t-1}z_{t-1}' + o_p(1).\end{aligned}$$

(iii) We can decompose the IVX denominator into two terms. The first part follows the derivations as,

$$z_t x_t' = R_{Tz}z_{t-1}x_{t-1}'R_{Tt} + R_{Tz}z_{t-1}u_{xt} + u_{xt}x_{t-1}'R_{Tt} + u_{xt}u_{xt}'.$$

By taking summations across time horizon,

$$\begin{aligned}
&\left[-\frac{C_z}{T^\gamma}+o\left(\frac{1}{T^\gamma}\right)\right]\sum_{t=1}^{T} z_{t-1}x'_{t-1} \\
= \quad & R_{Tz}z_0x'_0 - z_Tx'_T + R_{Tz}\sum_{t=1}^{T} z_{t-1}u_{xt} + \sum_{t=1}^{T} u_{xt}x'_{t-1} + \sum_{t=1}^{T} u_{xt}u'_{xt}.
\end{aligned}$$

Therefore, we have the following approximation as,

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T} z_{t-1}x'_{t-1} = -\left[\frac{1}{T}\sum_{t=1}^{T} u_{xt}x'_{t-1} + \frac{1}{T}\sum_{t=1}^{T} u_{xt}u'_{xt} + o_p(1)\right]C_z^{-1}.$$

Similarly, for the second term involving $\eta^{(1)}_{T,t-1}$,

$$\eta^{(1)}_{T,t-1}x'_{t-1} = R_{Tz}\eta^{(1)}_{T,t-1}x'_{t-1}R_{Tt}+R_{Tz}\eta^{(1)}_{T,t-1}u'_{xt}+C\left(\frac{t}{T}\right)x_{t-1}x'_{t-1}R_{Tt}+C\left(\frac{t}{T}\right)x_{t-1}u'_{xt}.$$

We take summations across time horizon as

$$\begin{aligned}
&\left[-\frac{C_z}{T^\gamma}+o\left(\frac{1}{T^\gamma}\right)\right]\sum_{t=1}^{T} \eta^{(1)}_{T,t-1}x'_{t-1} \\
= \quad & R_{Tz}\eta^{(1)}_{T,0}x'_0 - \eta^{(1)}_{T,T}x'_T + R_{Tz}\sum_{t=1}^{T} \eta^{(1)}_{T,t-1}u_{xt} + \sum_{t=1}^{T} C\left(\frac{t}{T}\right)x_{t-1}x'_{t-1} + \sum_{t=1}^{T} C\left(\frac{t}{T}\right)x_{t-1}u'_{xt}.
\end{aligned}$$

Finally, we derive a similar approximation as follows,

$$\frac{1}{T^{2+\gamma}}\sum_{t=1}^{T} \eta^{(1)}_{T,t-1}x'_{t-1} = -\sum_{t=1}^{T} C\left(\frac{t}{T}\right)x_{t-1}x'_{t-1}C_z^{-1} + o_p(1).$$

■

Proof of Theorem 1.1: (i) is verified by PM;

(ii) can be obtained by applying Lemma 1.1. As $T\to\infty$,

$$\begin{aligned}
\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T} z_{t-1}x'_{t-1} &= -\left[\frac{1}{T}\sum_{t=1}^{T} u_{xt}x'_{t-1} + \frac{1}{T}\sum_{t=1}^{T} u_{xt}u'_{xt} + o_p(1)\right]C_z^{-1} \\
&\Rightarrow -\left(\int_0^1 dB_x(r)K'_c(r)\,dr + \Omega_{xx}\right)C_z^{-1},
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{T^{2+\gamma}}\sum_{t=1}^{T} \eta^{(1)}_{T,t-1}x'_{t-1} &= -\left[\frac{1}{T^2}\sum_{t=1}^{T} C\left(\frac{t}{T}\right)x_{t-1}x'_{t-1} + o_p(1)\right]C_z^{-1} \\
&\Rightarrow -\int_0^1 C(r)\,K_c(r)\,K'_c(r)\,drC_z^{-1}.
\end{aligned}$$

(iii) is easily derived based on (i) and (ii); (iv) is a natural extension of (iii).

■

Proof for FMER

Proof of Lemma 1.2: Since ε_j is assumed to be mds, we can apply the martingale central limit theorem.

(i) Stability condition: When $\frac{k_T}{T}+\frac{1}{k_T}\to 0$,

$$\begin{aligned}
& \mathbb{E}\left(\left\|\frac{1}{T^{\alpha/2}}\sum_{j=1}^{k_T}\left[\exp\left(-\frac{1}{T^{\alpha}}\sum_{l=k_T-j+1}^{k_T}C\left(\frac{k_T-l+1}{T}\right)\right)\right]F_x(1)\epsilon_j\right\|^2\right) \\
= & \mathbb{E}tr\left\{\frac{1}{T^{\alpha}}\sum_{j=1}^{k_T}F_x(1)\epsilon_j\epsilon_j^{'}F_x^{'}(1)\left[\exp\left(-\frac{2}{T^{\alpha}}\sum_{l=k_T-j+1}^{k_T}C\left(\frac{k_T-l+1}{T}\right)\right)\right]\right\} \\
\leq & \frac{1}{T^{\alpha}}\sum_{j=1}^{k_T}\left\|\exp\left(-\frac{2}{T^{\alpha}}\sum_{l=k_T-j+1}^{k_T}C\left(\frac{k_T-l+1}{T}\right)\right)\right\|_{\infty}\cdot\|\Omega_{xx}\| \\
= & \int_{1/T^{\alpha}}^{[Tr]/T^{\alpha}}\left\|\exp\left(-2\int_0^z C\left(x\frac{k_T}{T}\right)dx\right)\right\|_{\infty}dz\cdot\|\Omega_{xx}\|+o_p(1) \\
\to & \int_0^{+\infty}\left\|\exp\left(-2zC(0)\right)\right\|_{\infty}dz\cdot\|\Omega_{xx}\|.
\end{aligned}$$

Therefore, we have

$$\frac{1}{T^{\alpha}}\tilde{x}_{k_T}\tilde{x}_{k_T}^{\prime}\Rightarrow\int_0^{+\infty}e^{-C(0)p}\Omega_{xx}e^{-C(0)p}dp,$$

where $\int_0^{+\infty}e^{-C(0)p}\Omega_{xx}e^{-C(0)p}dp$ is a constant matrix. We show the stability condition.

(ii) Lindeberg condition: For arbitrary $\delta > 0$, as $T \to +\infty$,

$$\frac{1}{T^\alpha}\sum_{j=1}^{k_T}\mathbb{E}\left\{\begin{array}{c}\left\|\exp\left(-\frac{1}{T^\alpha}\sum_{l=k_T-j+1}^{k_T}C\left(\frac{k_T-l+1}{T}\right)\right)F_x(1)\varepsilon_j\right\|^2 \\ \cdot\mathbf{1}\left[\|\varepsilon_j\| > \delta T^{\frac{\alpha}{2}}\exp\left(\frac{1}{T^\alpha}\sum_{l=k_T-j+1}^{k_T}C\left(\frac{[Tr]-l+1}{T}\right)\right)\right]\end{array}\right\}$$

$$\leq \frac{1}{T^\alpha}\sum_{j=1}^{k_T}\left\|\exp\left(-\frac{2}{T^\alpha}\sum_{l=k_T-j+1}^{k_T}C\left(\frac{k_T-l+1}{T}\right)\right)\right\|_\infty \mathbb{E}\|\varepsilon_j\|^2$$
$$\mathbf{1}\left\{\|\varepsilon_j\| > \delta T^{\alpha/2}\exp\left[\frac{1}{T^\alpha}\sum_{l=k_T-j+1}^{k_T}C\left(\frac{k_T-l+1}{T}\right)\right]\right\}$$
$$\leq \frac{1}{T^\alpha}\sum_{j=1}^{k_T}\left\|\exp\left(-\frac{2}{T^\alpha}\sum_{l=k_T-j+1}^{k_T}C\left(\frac{k_T-l+1}{T}\right)\right)\right\|_\infty \mathbb{E}\|\varepsilon_j\|^2\mathbf{1}\left[\|\varepsilon_j\| > \delta T^{\alpha/2}\right]$$
$$\leq \frac{T}{T^\alpha}\frac{1}{T}\sum_{j=1}^{[Tr]}\left\|\exp\left(-\frac{2}{T^\alpha}\sum_{l=[Tr]-j+1}^{[Tr]}C\left(\frac{T[r]-l+1}{T}\right)\right)\right\|_\infty \mathbb{E}\|\varepsilon_j\|^2\mathbf{1}\left[\|\varepsilon_j\|^2 > \delta^2 T^\alpha\right]$$
$$= \frac{T}{T^\alpha}\int_0^r\left\|\exp\left(-\frac{2T}{T^\alpha}\int_0^s C(a)\,da\right)\right\|_\infty ds\cdot\mathbb{E}\|\varepsilon_j\|^2\mathbf{1}\left[\|\varepsilon_j\| > \delta^2T^\alpha\right] + o_p(1)$$
$$\leq \frac{1}{2\|C(0)\|_\infty}\mathbb{E}\left[\|\varepsilon_j\|^2\mathbf{1}\left(\|\varepsilon_j\|^2 > \delta^2T^\alpha\right)\right] \to 0.$$

Therefore we successfully verify the Lindeberg condition. Summarizing (i) and (ii), we apply the martingale central limit theorem and have the following limit theory

$$\bar{x}_{k_T} \xrightarrow{d} \mathcal{N}\left(0, \int_0^{+\infty} e^{-C(0)p}\Omega_{xx}e^{-C(0)p}dp\right).$$

■

Proof of Lemma 1.3: The proof follows PL. For $k_T \leq j-1 \leq T$,

$$\bar{x}_{j-1} = \frac{1}{T^{\alpha/2}}\hat{\hat{x}}_{j-1} = \frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\exp\left[-\frac{1}{T^\alpha}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]u_{x,i-1}, \tag{A.1}$$

where

$$\hat{\hat{x}}_{j-1} := \exp\left[\frac{1}{T^\alpha}\sum_{i=1}^{j-1}C\left(\frac{i}{T}\right)\right]x_{j-1}. \tag{A.2}$$

By (A.1), (A.2) and Beveridge-Nelson decomposition in Phillips and Solo

(1992), we get the following expansion,

$$\begin{aligned}\frac{1}{T^{\alpha/2}}x_{j-1} &= \sum_{i=1}^{j-1}\frac{1}{T^{\alpha/2}}\exp\left[-\frac{1}{T^{\alpha/2}}\sum_{i=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]F_x(1)\,\varepsilon_{i-1} \\ &\quad -\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\exp\left[-\frac{1}{T^{\alpha/2}}\sum_{i=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]\Delta\tilde{\varepsilon}_{i-1}.\end{aligned}$$

We define $\phi_T(r) := \frac{1}{\sqrt{T}}S_{[Tr]} = \frac{1}{\sqrt{T}}\sum_{i=1}^{[Tr]}\varepsilon_i$. According to Zaitsev (1998), for some $q \geq 4$, there is a Brownian motion adapted to the expanded probability space by a strong approximation. We construct a Brownian motion $B(r)$ for some $q \geq 4$ with the following uniform approximations

$$\sup_{0\leq r\leq 1}\left(\left\|\frac{1}{\sqrt{T}}S_{[Tr]} - B(r)\right\|\right) = O_{a.s.}\left(T^{-\frac{1}{2}+\frac{1}{q}+\varepsilon}\right).$$

For each $t = 1, 2, ..., T$, we define

$$\frac{e_t}{\sqrt{T}} := B\left(\frac{t}{T}\right) - B\left(\frac{t-1}{T}\right) \text{ and } d_t = R_{Tt}d_{t-1} + v_t,$$

where $v_t = F_x(1)\,e_t - \Delta\tilde{e}_t$. Similarly, we have

$$\begin{aligned}&\frac{1}{T^{\alpha/2}}\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}C\left(\frac{i}{T}\right)\right]d_{j-1} \\ = &\sum_{i=1}^{j-1}\frac{1}{T^{\alpha/2}}\exp\left[-\frac{1}{T^{\alpha/2}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]F_x(1)\,e_{i-1} - \\ &\sum_{i=1}^{j-1}\frac{1}{T^{\alpha/2}}\exp\left[-\frac{1}{T^{\alpha/2}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]\Delta\tilde{e}_{i-1}.\end{aligned}$$

By triangle inequality, we bound the difference between the empirical process and its limiting process as follows,

$$
\begin{aligned}
& \left\| \frac{\hat{\hat{x}}_{j-1}}{T^{\alpha/2}} - X_{C(0)} \right\| \\
\leq & \left\| \begin{array}{l} \sum_{i=1}^{j-1} \frac{1}{T^{\alpha/2}} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i}^{j-1} C\left(\frac{j-l}{T}\right) \right] F_x(1)\,\varepsilon_i - \\ \sum_{i=1}^{j-1} \frac{1}{T^{\alpha/2}} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i}^{j-1} C\left(\frac{j-l}{T}\right) \right] F_x(1)\, e_i \end{array} \right\| \\
& + \left\| \sum_{i=1}^{j-1} \frac{1}{T^{\alpha/2}} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i}^{j-1} C\left(\frac{j-l}{T}\right) \right] F_x(1)\, e_i - X_{C(0)} \right\| \\
& + \left\| \frac{1}{T^{\alpha/2}} \sum_{i=1}^{j-1} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i}^{j-1} C\left(\frac{j-l}{T}\right) \right] \Delta\tilde{\varepsilon}_i \right\| \\
& + \left\| \frac{1}{T^{\alpha/2}} \sum_{i=1}^{j-1} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i}^{j-1} C\left(\frac{j-l}{T}\right) \right] \Delta\tilde{e}_i \right\| \\
= & : (B.1) + (B.2) + (B.3) + (B.4).
\end{aligned}
$$

(i) For the term of $(B.1)$, we intend to show

$$
\sup_{k_T \leq j-1 \leq T} \left\| \frac{1}{T^{\alpha/2}} \sum_{i=1}^{j-1} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i}^{j-1} C\left(\frac{j-l}{T}\right) \right] \left(F_x(1)\,\varepsilon_i - F_x(1)\, e_i\right) \right\| = o_{a.s}(1).
$$

First, for the empirical process, we have

$$
\begin{aligned}
& \frac{1}{T^{\alpha/2}} \sum_{i=1}^{j-1} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i}^{j-1} C\left(\frac{j-l}{T}\right) \right] F_x(1)\,\varepsilon_i \\
= & \frac{1}{T^{\frac{\alpha-1}{2}}} \sum_{i=1}^{j-1} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i}^{j-1} C\left(\frac{j-l}{T}\right) \right] F_x(1) \left[\phi_T\left(\frac{i}{T}\right) - \phi_T\left(\frac{i-1}{T}\right) \right] \\
= & \frac{\sqrt{T}}{T^{\alpha/2}} \sum_{i=1}^{j-1} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i}^{j-1} C\left(\frac{j-l}{T}\right) \right] F_x(1)\, \phi_T\left(\frac{i}{T}\right) \\
& - \frac{\sqrt{T}}{T^{\alpha/2}} \sum_{i=1}^{j-2} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i+1}^{j-1} C\left(\frac{j-l}{T}\right) \right] F_x(1)\, \phi_T\left(\frac{i}{T}\right) \\
= & \frac{\sqrt{T}}{T^{\alpha/2}} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=1}^{j-1} C\left(\frac{j-l}{T}\right) \right] F_x(1)\, \phi_T\left(\frac{j-1}{T}\right) \\
& + \frac{\sqrt{T}}{T^{\alpha/2}} \sum_{i=1}^{j-2} \exp \left[-\frac{1}{T^{\alpha}} \sum_{l=j-i+1}^{j-1} C\left(\frac{j-l}{T}\right) \right] \left\{ \exp \left[\frac{1}{T^{\alpha}} C\left(\frac{i-1}{T}\right) \right] - I_n \right\} F_x(1)\, \phi_T\left(\frac{i}{T}\right).
\end{aligned}
$$

Second, for the limiting process, we have

$$\begin{aligned}
&\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]F_x(1)\,e_i \\
=\ &\frac{\sqrt{T}}{T^{\alpha/2}}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=1}^{j-1}C\left(\frac{j-l}{T}\right)\right]F_x(1)\,B\left(\frac{j-1}{T}\right) \\
&+\frac{\sqrt{T}}{T^{\alpha/2}}\sum_{i=1}^{j-2}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i+1}^{j-1}C\left(\frac{j-l}{T}\right)\right]\left\{\exp\left[\frac{1}{T^{\alpha}}C\left(\frac{i-1}{T}\right)\right]-I_n\right\}F_x(1)\,B\left(\frac{j}{T}\right).
\end{aligned}$$

Therefore, the distance between empirical process and limiting process is bounded as follows,

$$\begin{aligned}
&\left\|\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]\left[F_x(1)\,\varepsilon_i-F_x(1)\,e_i\right]\right\| \\
\leq\ &\frac{\sqrt{T}}{T^{\alpha/2}}\left\|F_x(1)\right\|_{\infty}\left[\sup_{0\leq r\leq 1}\left\|\phi_T(r)-B(r)\right\|\right]\cdot \\
&\underbrace{\sum_{i=1}^{j-2}\left\|\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i+1}^{j-1}C\left(\frac{j-l}{T}\right)\right]\cdot\left\{\exp\left[\frac{1}{T^{\alpha}}C\left(\frac{i}{T}\right)\right]-I_n\right\}\right\|_{\infty}}_{(B.1.1)} \\
&+\frac{\sqrt{T}}{T^{\alpha/2}}\left\|F_x(1)\right\|_{\infty}\left[\sup_{0\leq r\leq 1}\left\|\phi_T(r)-B(r)\right\|\right]\underbrace{\left\|\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=1}^{j-1}C\left(\frac{j-l}{T}\right)\right]\right\|_{\infty}}_{(B.1.2)}.
\end{aligned}$$

When $k_T\leq j-1\leq T$, for the term of $(B.1.1)$, we have

$$\begin{aligned}
&\sum_{i=1}^{j-2}\left\|\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i+1}^{j-1}C\left(\frac{j-l}{T}\right)\right]\left[\exp\left[\frac{C(i/T)}{T^{\alpha}}\right]-I_n\right]\right\|_{\infty} \\
=\ &\sum_{i=1}^{j-2}\left\|\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=1}^{i-1}C\left(\frac{l}{T}\right)\right]\left[\exp\left[\frac{C(i/T)}{T^{\alpha}}\right]-I_n\right]\right\|_{\infty} \\
\approx\ &\int_0^{\frac{j-2}{T^{\alpha}}}\left\|\exp\left[-\int_0^z C\left(x\frac{T^{\alpha}}{T}\right)dx\right]C\left(\frac{zT^{\alpha}}{T}\right)\right\|_{\infty}dz+o(1) \\
\approx\ &\int_0^{+\infty}\left\|\exp\left[-\int_0^z C(0)\,dx\right]C(0)\right\|_{\infty}dz \\
=\ &\int_0^{+\infty}\left\|\exp\left[-zC(0)\right]C(0)\right\|_{\infty}dz=O(1), \qquad \text{(A.3)}
\end{aligned}$$

where $j-1>T^{\alpha}$ holds. Similarly, we have

$$\left\|\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=1}^{j-1}C\left(\frac{j-l}{T}\right)\right]\right\|_{\infty}\sim\left\|\exp\left[-\int_{0}^{+\infty}C\left(\frac{zT^{\alpha}}{T}\right)dz\right]\right\|_{\infty}=O\left(1\right), \tag{A.4}$$

due to the rate restriction $j-1>T^{\alpha}$. Moreover, with rate restriction $\alpha q>2$, the Skorohod embedding theorem illustrates the uniform approximations of Brownian motion as

$$\frac{\sqrt{T}}{T^{\alpha/2}}\left[\sup_{0\leq r\leq 1}\left\|\phi_{T}\left(r\right)-B\left(r\right)\right\|\right]=o_{a.s.}\left(1\right). \tag{A.5}$$

Combining (A.3) (A.4) and (A.5), we can justify that $(B.1)=o_{a.s.}(1)$.

(ii) We are supposed to check that the following two arguments, $(B.3)$ and $(B.4)$, are asymptotically negligible as,

$$\sup_{k_T\leq j-1\leq T}\left\|\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]\Delta\tilde{\varepsilon}_{i-1}\right\| = o_{a.s.}\left(1\right), \tag{A.6}$$

$$\sup_{k_T\leq j-1\leq T}\left\|\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]\Delta\tilde{e}_{i-1}\right\| = o_{a.s.}\left(1\right). \tag{A.7}$$

Since (A.7) and (A.6) share identical derivations, without loss of generality, we concentrate on (A.6). We decompose

$$\sup_{k_T\leq j-1\leq T}\left\|\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]\Delta\tilde{\varepsilon}_{i-1}\right\|,$$

into three terms as $(C.1)$, $(C.2)$ and $(C.3)$ as,

$$\begin{aligned}
&\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]\Delta\tilde{\varepsilon}_{i-1}\\
=\;&\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\left\{\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]-\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i+1}^{j-1}C\left(\frac{j-l}{T}\right)\right]\right\}\tilde{\varepsilon}_{i}\\
&+\frac{1}{T^{\alpha/2}}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]\tilde{\varepsilon}_{j-1}-\frac{1}{T^{\alpha/2}}\exp\left[-\frac{1}{T^{\alpha}}C\left(\frac{1}{T}\right)\right]\tilde{\varepsilon}_{0}\\
=\;&(C.1)+(C.2)-(C.3).
\end{aligned}$$

For $(C.3)$, based on the summability assumptions on innovations, we have

$$P\left(\left\|\frac{1}{T^{\alpha/2}}\tilde{\varepsilon}_{t}\right\|>\varepsilon\right)\leq\frac{E\left\|\tilde{\varepsilon}_{t}\right\|^{q}}{T^{\alpha q/2}\varepsilon^{q}}=O\left(\frac{1}{T^{\alpha q/2}}\right)=o\left(1\right).$$

With $\alpha q > 2$, Borel-Cantelli lemma applies as

$$\sum_{t=1}^{\infty} \Pr\left(\left\|\frac{1}{T^{\alpha/2}}\tilde{\varepsilon}_t\right\| > \varepsilon\right) \leq \sum_{t=1}^{\infty} O\left(\frac{1}{T^{\alpha q/2}}\right) < \infty. \tag{A.8}$$

Based on the fact that $\left\|\exp\left[-\frac{1}{T^\alpha}C\left(\frac{1}{T}\right)\right]\right\| = o(1)$ and (A.8), we have the following argument

$$\left\|\exp\left[-\frac{1}{T^\alpha}C\left(\frac{1}{T}\right)\right]\frac{1}{T^{\alpha/2}}\tilde{\varepsilon}_0\right\| = o_{a.s.}(1).$$

Similarly, for $(C.2)$, with $T^\alpha \leq k_T \leq j-1 \leq T$, we have

$$\exp\left[-\frac{1}{T^\alpha}\sum_{l=1}^{j-1} C\left(\frac{j-l}{T}\right)\right] = o(1).$$

By Borel-Cantelli lemma, the following argument holds

$$\left\|\frac{1}{T^{\alpha/2}}\exp\left[-\frac{1}{T^\alpha}\sum_{l=1}^{j-1} C\left(\frac{j-l}{T}\right)\right]\tilde{\varepsilon}_{j-1}\right\| = o_{a.s.}(1).$$

For $(C.1)$, if $k_T \leq j-1 \leq T$ holds, we have

$$\begin{aligned}
&\left\|\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\left\{\exp\left[-\frac{1}{T^\alpha}\sum_{l=j-i}^{j-1} C\left(\frac{j-l}{T}\right)\right] - \exp\left[-\frac{1}{T^\alpha}\sum_{l=j-i+1}^{j-1} C\left(\frac{j-l}{T}\right)\right]\right\}\tilde{\varepsilon}_i\right\| \\
= &\left\|\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\left\{\exp\left[-\frac{1}{T^\alpha}\sum_{l=j-i+1}^{j-1} C\left(\frac{j-l}{T}\right)\right]\left[\exp\left(\frac{1}{T^\alpha}C\left(\frac{i}{T}\right)\right) - I_n\right]\right\}\tilde{\varepsilon}_i\right\| \\
\sim &\left\|\frac{1}{T^{3\alpha/2}}\sum_{i=1}^{j-1}\left\{\exp\left[-\frac{1}{T^\alpha}\sum_{l=j-i+1}^{j-1} C\left(\frac{j-l}{T}\right)\right]C\left(\frac{i}{T}\right)\right\}\tilde{\varepsilon}_i\right\| \\
\sim &\ o_{a.s.}(1),
\end{aligned} \tag{A.9}$$

where the following fact holds,

$$\frac{1}{T^\alpha}\sum_{i=1}^{j-1}\left\{\exp\left[-\frac{1}{T^\alpha}\sum_{l=j-i+1}^{j-1} C\left(\frac{j-l}{T}\right)\right]C\left(\frac{i}{T}\right)\right\} \sim \int_0^{+\infty}\exp\left[-qC(0)\right]C(0)\,dq < +\infty.$$

Similarly, if $k_T \leq j-1 \leq T$, it shows that

$$\left\|\frac{1}{T^{3\alpha/2}}\sum_{i=1}^{j-1}\left\{\exp\left[-\frac{1}{T^\alpha}\sum_{l=j-i+1}^{j-1} C\left(\frac{j-l}{T}\right)\right]C\left(\frac{i}{T}\right)\right\}\tilde{\varepsilon}_{i-1}\right\| \sim O_{a.s.}\left(\frac{1}{T^{\alpha/2}}\right). \tag{A.10}$$

Combining (A.9) and (A.10), we have the aysmptotic negligible term as

$$\sup_{k_T\leq j-1\leq T}\left\|\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]\Delta\tilde{\varepsilon}_{i-1}\right\|=o_{a.s.}(1).$$

Similar derivations apply for $\Delta\tilde{e}_{i-1}$ as

$$\sup_{k_T\leq j-1\leq T}\left\|\frac{1}{T^{\alpha/2}}\sum_{i=1}^{j-1}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]\Delta\tilde{e}_{i-1}\right\|=o_{a.s.}(1).$$

(iii) To prove that

$$\sup_{k_T\leq j-1\leq T}\left\|\sum_{i=1}^{j-1}\frac{1}{T^{\alpha/2}}\exp\left[-\frac{1}{T^{\alpha}}\sum_{l=j-i}^{j-1}C\left(\frac{j-l}{T}\right)\right]F_x(1)\,e_i-X_{C(0)}\right\|=o_{a.s.}(1),$$

where the martingale central limit theorem can be applied to show pointwise convergence. Based on pointwise convergence, the uniform approximations can be justified as what we did for Lemma 1.2. ■

Proof of Lemma 1.4: For IVX residual $\eta_{T,t}^{(1)}=\sum_{j=1}^{t}R_{Tz}C\left(\frac{j+1}{T}\right)x_j$, we justify the following uniform approximation as

$$\begin{aligned}&\frac{1}{T^{\frac{\alpha}{2}+(\alpha\wedge\gamma)}}\exp\left(-\frac{1}{T^{a}}\sum_{i=1}^{t}C\left(\frac{i}{T}\right)\right)\eta_{T,t}^{(1)}\\ =\;&\frac{1}{T^{\frac{\alpha}{2}+(\alpha\wedge\gamma)}}\exp\left(-\frac{1}{T^{a}}\sum_{i=1}^{t}C\left(\frac{i}{T}\right)\right)\sum_{j=k_T+1}^{t}R_{Tz}^{t-j}C\left(\frac{j}{T}\right)x_{j-1}+o_p(1),\end{aligned}$$

where the following asymptotic negligibility holds as,

$$\begin{aligned}&\left\|\frac{1}{T^{\frac{\alpha}{2}+(\alpha\wedge\gamma)}}\exp\left(-\frac{1}{T^{a}}\sum_{i=1}^{t}C\left(\frac{i}{T}\right)\right)\sum_{j=1}^{k_T}R_{Tz}^{t-j}C\left(\frac{j}{T}\right)x_{j-1}\right\|\\ \leq\;&\sup_{j-1}\left\|\frac{1}{T^{\alpha/2}}\exp\left(-\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}C\left(\frac{i}{T}\right)\right)x_{j-1}\right\|\frac{1}{T^{\alpha\wedge\gamma}}\left\|\sum_{j=1}^{k_T}R_{Tz}^{t-j}\exp\left(-\frac{1}{T^{a}}\sum_{i=j}^{t}C\left(\frac{i}{T}\right)\right)\right\|_{\infty}\\ \leq\;&\sup_{j-1}\left\|\frac{1}{T^{\alpha/2}}\exp\left(-\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}C\left(\frac{i}{T}\right)\right)x_{j-1}\right\|\frac{1}{T^{\alpha\wedge\gamma}}\left\|\sum_{j=1}^{k_T}R_{Tz}^{t-j}R^{t-j}\right\|_{\infty}=o_p(1),\end{aligned}\tag{A.11}$$

and $R:=\exp\left(\frac{C}{T^{\alpha}}\right)$ with $C=c\cdot I_n$ and $c_i<c$ for all $i=1,2,...,n$. The last line of asymptotic negligibility in (A.11) is due to the following two facts:

$$\sup_{j-1}\left\|\frac{1}{T^{\alpha/2}}\exp\left(-\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}C\left(\frac{i}{T}\right)\right)x_{j-1}\right\|=O_p(1),$$

and

$$\begin{aligned}\left\|\sum_{j=1}^{k_T} R_{Tz}^{t-j} R^{t-j}\right\|_\infty &= O\left(\frac{1}{T^{\alpha\wedge\gamma}}\left\|R_{Tz}\right\|_\infty^{t-1}\left\|R\right\|_\infty^{-t}\frac{\left\|R_{Tz}\right\|_\infty^{k_T}\left\|R\right\|_\infty^{k_T}-1}{\left\|R_{Tz}\right\|_\infty^{-1}\left\|R\right\|_\infty^{-1}}\right)\\ &= O\left(\frac{\left\|R_{Tz}\right\|_\infty^{t-k_T-1}\left\|R\right\|_\infty^{k_T-t}-\left\|R\right\|_\infty^{-t}\left\|R_{Tz}\right\|_\infty^{t-1}}{T^{\alpha\wedge\gamma}\left(\left\|R_{Tz}\right\|_\infty^{-1}\left\|R\right\|_\infty^{-1}-1\right)}\right)\\ &= o(1). \qquad\qquad (A.12)\end{aligned}$$

We discuss the validity of (A.12) as follows: When $\frac{T^\alpha}{k_T}+\frac{T^\gamma}{k_T}\to 0$, we have $\left\|R\right\|_\infty^{-k_T}\to 0$, $\left\|R_{Tz}\right\|_\infty^{k_T}\to 0$, and

$$\begin{aligned}T^{\alpha\wedge\gamma}\left(\left\|R_{Tz}\right\|_\infty^{-1}\left\|R\right\|_\infty-1\right) &= T^{\alpha\wedge\gamma}\left(\frac{\frac{1}{T^\alpha}\max c_i-\frac{1}{T^\gamma}\max c_{zi}}{1+\frac{1}{T^\gamma}\max c_{zi}}\right)\\ &\to \begin{cases}-\max\left(c_{zi}\right) & \text{if } \alpha>\gamma,\\ \max\left(c_i\right) & \text{if } \alpha<\gamma,\\ \max\left(c_i\right)-\max\left(c_{zi}\right) & \text{if } \alpha=\gamma.\end{cases}\end{aligned}$$

Therefore we verify the boundedness of the denominator in (A.12). To justify the asymptotic negligibility of the numerator in (A.12), we introduce the following decomposition as

$$\left\|R_{Tz}\right\|_\infty^{t-k_T}\left\|R\right\|_\infty^{k_T-t}=O\left(\exp\left(\max\left(c_{zi}\right)\frac{t-k_T}{T^\gamma}\right)\exp\left(-\max\left(c_i\right)\frac{t-k_T}{T^\alpha}\right)\right),$$

where the following rate restriction $\frac{t-k_T}{T^\gamma}+\frac{t-k_T}{T^\alpha}\to 0$ holds. Therefore, for any $t\in[k_T,T]$, we have $\left\|R_{Tz}\right\|_\infty^{t}\left\|R\right\|_\infty^{-t}=o(1)$ and $\left\|R_{Tz}\right\|_\infty^{t-k_T}\left\|R\right\|_\infty^{k_T-t}=o(1)$. Based on the arguments above, we justify the asymptotic negligibility of (A.11) as

$$T^{-\frac{\alpha+2(\alpha\wedge\gamma)}{2}}\exp\left[-\frac{1}{T^\alpha}\sum_{i=1}^{t}C\left(\frac{i}{T}\right)\sum_{j=1}^{k_T}C\left(\frac{j+1}{T}\right)x_j\right]=o_p(1). \qquad (A.13)$$

Based on (A.13), we apply the approximation and justify the asymptotics of

IVX residuals. We provide details as follows,

$$\begin{aligned}
&\frac{1}{T^{\frac{\alpha}{2}+(\alpha\wedge\gamma)}}\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=1}^{t}C\left(\frac{i}{T}\right)\right]\eta_{T,t}^{(1)} \\
=\; &\frac{1}{T^{\frac{\alpha}{2}+(\alpha\wedge\gamma)}}\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=1}^{t}C\left(\frac{i}{T}\right)\right]\sum_{j=k_T+1}^{t}R_{Tz}^{t-j}C\left(\frac{j}{T}\right)x_{j-1}+o_p(1) \\
=\; &\frac{1}{T^{\frac{\alpha}{2}+(\alpha\wedge\gamma)}}\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=1}^{t}C\left(\frac{i}{T}\right)\right]\sum_{j=k_T+1}^{t}R_{Tz}^{t-j}C\left(\frac{j}{T}\right)\exp\left[\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}C\left(\frac{i}{T}\right)\right] \\
&\cdot\left\{\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}C\left(\frac{i}{T}\right)\right]x_{j-1}\right\}+o_p(1) \\
=\; &\frac{1}{T^{(\alpha\wedge\gamma)}}\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=1}^{t}C\left(\frac{i}{T}\right)\right]\sum_{j=1}^{t}R_{Tz}^{t-j}C\left(\frac{j}{T}\right)\exp\left[\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}C\left(\frac{i}{T}\right)\right]X_{C(0)}+o_p(1).
\end{aligned}$$

Note that

$$\begin{aligned}
&\frac{1}{T^{(\alpha\wedge\gamma)}}\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=1}^{t}C\left(\frac{i}{T}\right)\right]\sum_{j=1}^{t}R_{Tz}^{t-j}C\left(\frac{j}{T}\right)\exp\left[\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}C\left(\frac{i}{T}\right)\right] \\
=\; &\frac{1}{T^{(\alpha\wedge\gamma)}}\mathrm{diag}\left\{\begin{array}{c}\sum_{j=1}^{t}\exp\left(\frac{c_z}{T^{\gamma}}\right)^{t-j}\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=1}^{t}c_1\left(\frac{i}{T}\right)\right]c_1\left(\frac{j}{T}\right)\exp\left[\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}c_1\left(\frac{i}{T}\right)\right],\ldots, \\ \sum_{j=1}^{t}\exp\left(\frac{c_z}{T^{\gamma}}\right)^{t-j}\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=1}^{t}c_n\left(\frac{i}{T}\right)\right]c_n\left(\frac{j}{T}\right)\exp\left[\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}c_n\left(\frac{i}{T}\right)\right]\end{array}\right\}.
\end{aligned}$$

For any persistent regressor $1\leq m\leq n$, we have

$$\begin{aligned}
&\frac{1}{T^{(\alpha\wedge\gamma)}}\sum_{j=1}^{t}\exp\left(\frac{c_z}{T^{\gamma}}\right)^{t-j}\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=1}^{t}c_m\left(\frac{i}{T}\right)\right]c_m\left(\frac{j}{T}\right)\exp\left[\frac{1}{T^{\alpha}}\sum_{i=1}^{j-1}c_m\left(\frac{i}{T}\right)\right] \\
=\; &\frac{1}{T^{(\alpha\wedge\gamma)}}\sum_{j=1}^{t}\exp\left[\frac{(t-j)\,c_z}{T^{\gamma}}\right]c_m\left(\frac{j}{T}\right)\exp\left[-\frac{1}{T^{\alpha}}\sum_{i=j}^{t}c_m\left(\frac{i}{T}\right)\right] \\
=\; &\frac{1}{T^{(\alpha\wedge\gamma)}}\sum_{j=1}^{t}\exp\left[\frac{1}{T^{\alpha+\gamma}}\left((t-j)\,c_zT^{\alpha}-T^{\gamma}\sum_{i=j}^{t}c_m\left(\frac{i}{T}\right)\right)\right]c_m\left(\frac{j}{T}\right).
\end{aligned}$$

(i) If $\alpha>\gamma$,

$$\begin{aligned}
&\frac{1}{T^{(\alpha\wedge\gamma)}}\sum_{j=1}^{t}\exp\left[\frac{1}{T^{\alpha+\gamma}}\left((t-j)\,c_zT^{\alpha}-T^{\gamma}\sum_{i=j}^{t}c_m\left(\frac{i}{T}\right)\right)\right]c_m\left(\frac{j}{T}\right) \\
=\; &\frac{1}{T^{(\alpha\wedge\gamma)}}\sum_{j=1}^{t}\exp\left[\frac{(t-j)\,c_z}{T^{\gamma}}-O\left(\frac{t-j}{T^{\alpha}}\right)\right]c_m\left(\frac{j}{T}\right) \\
\sim\; &\frac{1}{T^{(\alpha\wedge\gamma)}}\sum_{j=1}^{t}\exp\left[\frac{(t-j)\,c_z}{T^{\gamma}}\right]c_m\left(\frac{j}{T}\right)=\sum_{j=1/T^{(\alpha\wedge\gamma)}}^{t/T^{(\alpha\wedge\gamma)}}\exp\left[\frac{(t-j)\,c_z}{T^{\gamma}}\right]c_m\left(\frac{j}{T}\right) \quad \text{(A.14)}
\end{aligned}$$

Since $\frac{1}{T^{(\alpha\wedge\gamma)}} \leq j \leq \frac{t}{T^{(\alpha\wedge\gamma)}}$ holds, we have $c_m\left(\frac{j}{T}\right) \to c_m(0)$ as $T \to \infty$. Equation (A.14) can be computed as,

$$\sum_{j=1/T^{(\alpha\wedge\gamma)}}^{t/T^{(\alpha\wedge\gamma)}} \exp\left[\frac{(t-j)\,c_z}{T^\gamma}\right] c_m\left(\frac{j}{T}\right) \sim \sum_{j=1/T^{(\alpha\wedge\gamma)}}^{t/T^{(\alpha\wedge\gamma)}} \exp\left[\frac{(t-j)\,c_z}{T^\gamma}\right] c_m(0) \to -\frac{c_m(0)}{c_z}.$$

(ii) If $\alpha < \gamma$,

$$\begin{aligned}
& \frac{1}{T^{(\alpha\wedge\gamma)}} \sum_{j=1}^{t} \exp\left[\frac{1}{T^{\alpha+\gamma}}\left[(t-j)\,c_z T^\alpha - T^\gamma \sum_{i=j}^{t} c_m\left(\frac{i}{T}\right)\right]\right] c_m\left(\frac{j}{T}\right) \\
= \; & \frac{1}{T^{(\alpha\wedge\gamma)}} \sum_{j=1}^{t} \exp\left[O\left(\frac{t-j}{T^\gamma}\right) - \frac{1}{T^\alpha}\sum_{i=j}^{t} c_m\left(\frac{i}{T}\right)\right] c_m\left(\frac{j}{T}\right) \\
\sim \; & \frac{1}{T^{(\alpha\wedge\gamma)}} \sum_{j=1}^{t} \exp\left[-\frac{1}{T^\alpha}\sum_{i=j}^{t} c_m\left(\frac{i}{T}\right)\right] c_m\left(\frac{j}{T}\right) \\
\sim \; & \sum_{j=1/T^{(\alpha\wedge\gamma)}}^{t/T^{(\alpha\wedge\gamma)}} \exp\left[-\frac{1}{T^\alpha}\sum_{i=j}^{t} c_m\left(\frac{i}{T}\right)\right] c_m(0) \\
= \; & \sum_{j=1/T^{(\alpha\wedge\gamma)}}^{t/T^{(\alpha\wedge\gamma)}} \exp\left[-\sum_{i=j/T^\alpha}^{t/T^\alpha} c_m\left(\frac{i}{T}\right)\right] c_m(0) \\
= \; & \sum_{j=1/T^{(\alpha\wedge\gamma)}}^{t/T^{(\alpha\wedge\gamma)}} \exp\left[-\sum_{i=0}^{(t-j)/T^\alpha} c_m\left(\frac{t-i}{T}\right)\right] c_m(0) \\
= \; & \sum_{j=1/T^{(\alpha\wedge\gamma)}}^{(t-1)/T^{(\alpha\wedge\gamma)}} \exp\left[-\sum_{i=0}^{(t-j)/T^\alpha} c_m\left(\frac{t-i}{T}\right)\right] c_m(0) \\
= \; & \int_0^{(t-1)/T^{(\alpha\wedge\gamma)}} \exp\left[\int_0^j c_m\left(\frac{t}{T}z\right) dz\right] dj \cdot c_m(0) + o(1) \\
\to \; & \int_0^\infty \exp\Big[-j\cdot\, c_m(0)\Big]\, dj \cdot c_m(0) = \frac{c_m(0)}{c_m(0)},
\end{aligned}$$

where we have $c_m\left(\frac{j}{T}\right) \to c_m(0)$ due to the following rate restriction $\frac{1}{T^{(\alpha\wedge\gamma)}} \leq j \leq \frac{t}{T^{(\alpha\wedge\gamma)}}$.

(iii) If $\alpha = \gamma$,

$$\begin{aligned}
& \frac{1}{T^\alpha} \sum_{j=1}^{t} \exp\left[\frac{1}{T^\alpha}\left[(t-j)\,c_z - \sum_{i=0}^{t-j} c_m\left(\frac{t-i}{T}\right)\right]\right] c_m\left(\frac{j}{T}\right) \\
= \; & \int_0^{(t-1)/T^{(\alpha\wedge\gamma)}} \exp\left[j\cdot c_z - \int_0^j c_m\left(\frac{t}{T}z\right) dz\right] dj \cdot c_m(0)\,[1+o(1)] \\
\to \; & \int_0^\infty \exp\left[j\cdot c_z - j\cdot c_m(0)\right] dj \cdot c_m(0) = \frac{c_m(0)}{-c_z + c_m(1)},
\end{aligned}$$

where we have $c_m\left(\frac{j}{T}\right) \to c_m(0)$ due to the following rate restriction $\frac{1}{T^{(\alpha\wedge\gamma)}} \leq j \leq \frac{t}{T^{(\alpha\wedge\gamma)}}$. We conclude this proof. ■

Proof of Lemma 1.5: Due to the dominance of IVX residual in numerator and denominator, with k_T defined in Lemma 1.4, we have the following approximation

$$\begin{aligned}
& \frac{1}{T^{(\alpha\wedge\gamma)}}\sum_{t=1}^{T} u_{0t}\tilde{z}_{t-1}' \exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T} C\left(\frac{j}{T}\right)\right] \\
= & \frac{1}{T^{\alpha+(\alpha\wedge\gamma)}}\sum_{t=1}^{T} u_{0t}\eta_{T,t-1}^{(1)\prime} \exp\left[-\frac{1}{T^{a}}\sum_{j=1}^{T} C\left(\frac{j}{T}\right)\right] + o_p(1) \\
= & \frac{1}{T^{\alpha+(\alpha\wedge\gamma)}}\sum_{t=k_T}^{T} u_{0t}\eta_{T,t-1}^{(1)\prime} \exp\left[-\frac{1}{T^{a}}\sum_{j=1}^{T} C\left(\frac{j}{T}\right)\right] + o_p(1). \quad \text{(A.15)}
\end{aligned}$$

The first equality comes from the dominance of the IVX residual in the numerator. The second equality is due to the following derivation

$$\begin{aligned}
& \left\|\frac{1}{T^{\alpha+(\alpha\wedge\gamma)}}\sum_{t=1}^{k_T-1} u_{0t}\eta_{T,t-1}^{(1)\prime} \exp\left[-\frac{1}{T^{a}}\sum_{j=1}^{T} C\left(\frac{j}{T}\right)\right]\right\| \\
= & \left\|\frac{1}{T^{\alpha/2}}\sum_{t=1}^{k_T-1} u_{0t}\left\{\frac{1}{T^{\alpha/2+(\alpha\wedge\gamma)}}\eta_{T,t-1}^{(1)\prime} \exp\left[-\frac{1}{T^{a}}\sum_{j=1}^{t-1} C\left(\frac{j}{T}\right)\right]\right\}' \exp\left[-\frac{1}{T^{\alpha}}\sum_{j=t}^{T} C\left(\frac{j}{T}\right)\right]\right\| \\
\leq & \sup_{1\leq t\leq T}\|u_{0t}\|_{\infty} \sup_{1\leq t\leq T}\left\|\frac{1}{T^{\alpha/2+(\alpha\wedge\gamma)}}\eta_{T,t-1}^{(1)\prime} \exp\left[-\frac{1}{T^{a}}\sum_{j=1}^{t-1} C\left(\frac{j}{T}\right)\right]\right\| \\
& \cdot\frac{1}{T^{\alpha/2}}\sum_{t=1}^{k_T-1}\left\|\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=t}^{T} C\left(\frac{j}{T}\right)\right]\right\|_{\infty} \\
\leq & O_p(1)\frac{1}{T^{\alpha/2}}\sum_{t=1}^{k_T}\left\|\exp\left(-\frac{1}{T^{\alpha}}\sum_{j=t}^{T}\underline{C}\right)\right\|_{\infty} \\
= & O_p(1)\frac{1}{T^{\alpha/2}}\left\|\exp\left(-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}\underline{C}\right)\sum_{t=1}^{k_T}\exp\left(\frac{1}{T^{\alpha}}\sum_{j=1}^{t}\underline{C}\right)\right\|_{\infty} \\
= & O_p\left(T^{\alpha/2}\left\|\exp\left(\frac{k_T-T}{T^{\alpha}}\underline{C}\right)\right\|_{\infty}\right) = o_p(1),
\end{aligned}$$

where we define $\underline{C} := \underline{c}\cdot I_n$ with $\inf_{1\leq i\leq n}\inf_{r\in[0,\infty)} c_i(r) > \underline{c}$. As $\frac{k_T-T}{T^{\alpha}} \to -\infty$, we have

$$O_p\left(T^{\alpha/2}\left\|\exp\left(\frac{k_T-T}{T^{\alpha}}\underline{C}\right)\right\|_{\infty}\right) = o(1).$$

Therefore,

$$\frac{1}{T^{\alpha+(\alpha\wedge\gamma)}}\sum_{t=1}^{k_T-1}u_{0t}\eta_{T,t-1}^{(1)\prime}\exp\left[-\frac{1}{T^a}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right]=o_p\left(1\right).$$

(i) Based on (A.15), we have

$$\begin{aligned}
&\frac{1}{T^{\alpha+(\alpha\wedge\gamma)}}\sum_{t=1}^{T}u_{0t}\eta_{T,t-1}^{(1)\prime}\exp\left[-\frac{1}{T^a}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right]\\
=\ &\frac{1}{T^{\alpha+(\alpha\wedge\gamma)}}\sum_{t=k_T}^{T}u_{0t}\eta_{T,t-1}^{(1)\prime}\exp\left[-\frac{1}{T^a}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right]+o_p\left(1\right)\\
=\ &\frac{1}{T^{\alpha/2}}\sum_{t=k_T}^{T}u_{0t}\left\{\frac{1}{T^{\alpha/2+(\alpha\wedge\gamma)}}\exp\left[-\frac{1}{T^a}\sum_{j=1}^{t-1}C\left(\frac{j}{T}\right)\right]\eta_{T,t-1}^{(1)}\right\}'\exp\left[-\frac{1}{T^a}\sum_{j=t}^{T}C\left(\frac{j}{T}\right)\right]+\\
&o_p\left(1\right)\\
=\ &\frac{1}{T^{\alpha/2}}\sum_{t=k_T}^{T}u_{0t}\left[C_{\alpha\gamma}X_{C(0)}+o_p\left(1\right)\right]'\exp\left[-\frac{1}{T^a}\sum_{j=t}^{T}C\left(\frac{j}{T}\right)\right]+o_p\left(1\right)\\
=\ &\frac{1}{T^{\alpha/2}}\sum_{t=k_T}^{T}u_{0t}\left[C_{\alpha\gamma}X_{C(0)}\right]'\exp\left[-\frac{1}{T^a}\sum_{j=t}^{T}C\left(\frac{j}{T}\right)\right]+o_p\left(1\right)\\
=\ &\frac{1}{T^{\alpha/2}}\sum_{t=1}^{T}u_{0t}\left[C_{\alpha\gamma}X_{C(0)}\right]'\exp\left[-\frac{1}{T^a}\sum_{j=t}^{T}C\left(\frac{j}{T}\right)\right]+o_p\left(1\right)\\
=\ &\frac{1}{T^{\alpha/2}}\sum_{t=1}^{T}u_{0t}\left[C_{\alpha\gamma}X_{C(0)}\right]'\exp\left[-\frac{1}{T^a}\sum_{j=0}^{T-t}C\left(\frac{T-j}{T}\right)\right]+o_p\left(1\right)\\
\Rightarrow\ &\mathcal{MN}\left(0,V\right),
\end{aligned}$$

where $0\leq j\leq\frac{T-t}{T^{\alpha}}$ and $1-\frac{T-t}{T^{\alpha+1}}\leq\frac{T-j}{T}\leq 1$. The variance matrix V follows

$$V:=\int_0^{+\infty}e^{-pC(1)}C_{\alpha\gamma}X_{C(0)}X_{C(0)}'C_{\alpha\gamma}e^{-pC(1)}dp\cdot\Omega_{00}.$$

(ii) Since the IVX denominator follows the decomposition as,

$$\sum_{t=1}^{T}\tilde{z}_{t-1}x_{t-1}'=\sum_{t=1}^{T}z_{t-1}x_{t-1}'+\frac{1}{T^a}\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}x_{t-1}'.$$

By the dominance of exponential rate in IVX residual, we have the following approximation

$$
\begin{aligned}
& \frac{1}{T^{\alpha+(\alpha\wedge\gamma)}}\sum_{t=1}^{T}\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right]\tilde{z}_{t-1}x_{t-1}'\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right] \\
= \; & \frac{1}{T^{\alpha+(\alpha\wedge\gamma)}}\sum_{t=1}^{T}\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right]z_{t-1}x_{t-1}'\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right]+ \\
& \frac{1}{T^{2\alpha+(\alpha\wedge\gamma)}}\sum_{t=1}^{T}\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right]\eta_{T,t-1}^{(1)}x_{t-1}'\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right] \\
= \; & \frac{1}{T^{2\alpha+(\alpha\wedge\gamma)}}\sum_{t=1}^{T}\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right]\eta_{T,t-1}^{(1)}x_{t-1}'\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right]+o_p(1).
\end{aligned}
$$

For the leading term, we have

$$
\begin{aligned}
& \frac{1}{T^{2\alpha+(\alpha\wedge\gamma)}}\sum_{t=1}^{T}\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right]\eta_{T,t-1}^{(1)}x_{t-1}'\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{T}C\left(\frac{j}{T}\right)\right] \\
= \; & \frac{1}{T^{\alpha}}\sum_{t=k_T}^{T}\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=t}^{T}C\left(\frac{j}{T}\right)\right]\left\{\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{t-1}C\left(\frac{j}{T}\right)\right]\frac{x_{t-1}}{T^{\alpha/2}}\right\} \\
& \cdot\left\{\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=1}^{t-1}C\left(\frac{j}{T}\right)\right]\frac{\eta_{T,t-1}^{(1)}}{T^{\alpha/2+(\alpha\wedge\gamma)}}\right\}'\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=t}^{T}C\left(\frac{j}{T}\right)\right]+o_p(1) \\
= \; & \frac{1}{T^{\alpha}}\sum_{t=1}^{T}\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=0}^{T-t}C\left(\frac{T-j}{T}\right)\right]X_{C(0)}X_{C(0)}'\exp\left[-\frac{1}{T^{\alpha}}\sum_{j=0}^{T-t}C\left(\frac{T-j}{T}\right)\right]C_{\alpha\gamma}+o_p(1) \\
= \; & \int_{1/T^{\alpha}}^{T/T^{\alpha}}\exp\left[-\int_0^p C\left(q\frac{T-T^{\alpha}}{T}\right)dq\right]X_{C(0)}X_{C(0)}'\exp\left[-\int_0^p C\left(q\frac{T-T^{\alpha}}{T}\right)dq\right]C_{\alpha\gamma}dp+ \\
& o_p(1) \\
\Rightarrow \; & \int_0^{+\infty}e^{-pC(1)}X_{C(0)}X_{C(0)}'e^{-pC(1)}dp\cdot C_{\alpha\gamma}=:\Phi.
\end{aligned}
$$

(iii) is proved by continuous mapping theorem based on (i) (ii). ■

Proof for FMSR

Proof of Lemma 1.6: Denote the autocovariance matrix of u_{xt} as

$$\Gamma_{u_x}(h)=\mathbb{E}\left(u_{x_t}u_{x_{t-h}}'\right).$$

We have

$$\mathbb{E}\left(x_j x_i'\right)=\sum_{k=2}^{j}\sum_{l=2}^{i}R_{Tt}^{(j-k)}\Gamma_{u_x}(k-l)\,R_{Tt}^{(i-l)},$$

where

$$R_{Tt}^{(j-k)} := \prod_{m=k+1}^{j} R_{Tm} \text{ and } R_{Tt}^{(i-l)} := \prod_{m=l+1}^{i} R_{Tm}.$$

The asymptotic upper bound of IVX residual is as

$$\begin{aligned}
\mathbb{E}\left\|\eta_{T,t-1}^{(1)}\right\|^2 &= \mathbb{E}tr\left\{\sum_{i,j=1}^{t-1} R_{Tz}^{t-i-1} C\left(\frac{i}{T}\right)\left(x_{i-1}x_{j-1}'\right) C\left(\frac{j}{T}\right) R_{Tz}^{t-j-1}\right\} \\
&\leq \sum_{i,j=1}^{t-1}\sum_{k=2}^{j}\sum_{l=2}^{i}\left\|R_{Tz}^{t-j-1}C\left(\frac{j}{T}\right)\right\|_\infty\left\|R_{Tz}^{t-i-1}C\left(\frac{i}{T}\right)\right\|_\infty\left\|R_{Tt}^{(j-k)}\right\|_\infty \\
&\quad \left\|R_{Tt}^{(i-l)}\right\|_\infty\left\|\Gamma_{u_x}(k-l)\right\|.
\end{aligned}$$

If Assumptions 1.1 and 1.2 hold, we have

$$\begin{aligned}
\frac{1}{T^\gamma}\sum_{j=1}^{t}\left\|R_{Tz}^{t-j-1}C\left(\frac{j}{T}\right)\right\|_\infty &= \int_0^{+\infty}\left\|e^{rC_z}C(r)\right\|_\infty dr + o_p(1) \\
&\leq \left(\int_0^{+\infty}\left\|e^{rC_z}\right\|_\infty^2 dr\right)^{1/2}\left(\int_0^{+\infty}\left\|C(r)\right\|_\infty^2 dr\right)^{1/2} \\
&< +\infty.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\frac{1}{T^\alpha}\sup_{0\leq r\leq 1}\sum_{j=1}^{t}\left\|R_{Tt}^{(t-j)}\right\|_\infty \\
=\ &\sup_{0\leq r\leq 1}\sum_{j=1/T^\alpha}^{[Tr]/T^\alpha}\left\|\exp\left[\sum_{l=1}^{[Tr]-j}\frac{1}{T^\alpha}C\left(\frac{[Tr]-l+1}{T}\right)\right]\right\|_\infty \\
\leq\ &\frac{1}{T^\alpha}\sup_{0\leq r\leq 1}\sum_{j=1}^{[Tr]-\left[\sqrt{T^{\alpha+1}}\right]-1}\left\|\exp\left[\frac{1}{T^\alpha}\sum_{l=1}^{[Tr]-j}C\left(\frac{[Tr]-l+1}{T}\right)\right]\right\|_\infty \\
&+\frac{1}{T^\alpha}\sup_{0\leq r\leq 1}\sum_{j=[Tr]-\left[\sqrt{T^{\alpha+1}}\right]}^{[Tr]}\left\|\exp\left[\frac{1}{T^\alpha}\sum_{l=1}^{[Tr]-j}C\left(\frac{[Tr]-l+1}{T}\right)\right]\right\|_\infty \\
=\ &:(D.1)+(D.2).
\end{aligned}$$

For $(D.1)$, if $j \leq [Tr]-\left[\sqrt{T^{\alpha+1}}\right]$, we have

$$\left\|\exp\left[\frac{1}{T^\alpha}\sum_{l=1}^{[Tr]-j}C\left(\frac{[Tr]-l+1}{T}\right)\right]\right\|_\infty \leq \left\|\exp\left(-\sqrt{T^{1-\alpha}}C\right)\right\|_\infty,$$

where $C := cI_n$ with $c := \sup_{1\le m\le n}\sup_{1\le t\le T} c_m(\frac{t}{T})$. Due to the dominant exponential rate, we have

$$\begin{aligned}
& \frac{1}{T^{\alpha}} \sum_{j=1}^{[Tr]-\left[\sqrt{T^{\alpha+1}}\right]-1} \left\| \exp\left[\frac{1}{T^{\alpha}} \sum_{l=1}^{[Tr]-j} C\left(\frac{[Tr]-l+1}{T}\right)\right] \right\|_{\infty} \\
\le \;& \frac{[Tr]-\left[\sqrt{T^{\alpha+1}}\right]}{T^{\alpha}} \left\| \exp\left(-\sqrt{T^{1-\alpha}}C\right)\right\|_{\infty} = o_p(1),
\end{aligned}$$

as $T \to +\infty$. Next, we consider the second term. If $j \ge [Tr]-\sqrt{T^{\alpha+1}}$holds, then we have $\frac{j+1}{T} \ge \frac{[Tr]-\sqrt{T^{\alpha+1}}}{T} \sim r$. The term $C\left(\frac{[Tr]-l+1}{T}\right) \sim C(r)$ holds uniformly for any $l = 1, 2, \ldots, [Tr]-j$. Therefore, for $(D.2)$, we have

$$\begin{aligned}
& \frac{1}{T^{\alpha}} \sum_{j=[Tr]-\left[\sqrt{T^{\alpha+1}}\right]}^{[Tr]} \left\| \exp\left[\frac{1}{T^{\alpha}} \sum_{l=1}^{[Tr]-j} C\left(\frac{[Tr]-l+1}{T}\right)\right] \right\|_{\infty} \\
\sim \;& \frac{1}{T^{\alpha}} \sum_{j=[Tr]-\left[\sqrt{T^{\alpha+1}}\right]}^{[Tr]} \left\| \exp\left[2C(r)\frac{[Tr]-j}{T^{\alpha}}\right] \right\|_{\infty} \\
\sim \;& \int_0^{\frac{\sqrt{T^{\alpha+1}}}{T^{\alpha}}} \left\|\exp\left[C(r)z\right]\right\|_{\infty} dz \\
\to \;& \int_0^{+\infty} \left\|\exp\left[C(r)z\right]\right\|_{\infty} dz < +\infty.
\end{aligned}$$

It illustrates that

$$\sup_{1\le t\le T} \sum_{j=1}^{t-1} \left\| R_{Tt}^{(j-k)} \right\|_{\infty} = O_p\left(T^{\alpha}\right).$$

Therefore we discuss the stochastic order of IVX residual in the following two cases : (i) $\gamma < \alpha$; (ii) $\gamma > \alpha$. If $\gamma < \alpha$ holds, then we have

$$\begin{aligned}
\sup_{1\le t\le T} \mathbb{E}\left\|\eta_{T,t}^{(1)}\right\|^2 &\le \left(\sup_{1\le t\le T}\sum_{i=1}^{t} \|R_{Tz}\|_{\infty}^{t-i}\right)^2 \left(\sup_{1\le t\le T}\sum_{k=1}^{i}\left\|R_{Tt}^{(j-k)}\right\|_{\infty}\right) \sum_{l=-\infty}^{+\infty} \|\Gamma_{u_x}(l)\| \\
&= O_p\left(T^{\alpha+2\gamma}\right).
\end{aligned}$$

If $\gamma > \alpha$ holds, letting $j-k=m$, we have

$$\sup_{1\le t\le T} \mathbb{E}\left\|\eta_{T,t}^{(1)}\right\|^2 \le \sum_{i,j=1}^{t}\sum_{m=0}^{j-2}\sum_{l=1}^{i} \|R_{Tz}\|_{\infty}^{2t-j-i} \left\|R_{Tt}^{(j-k)}\right\|_{\infty} \left\|R_{Tt}^{(i-l)}\right\|_{\infty} \|\Gamma_{u_x}(k-l)\|.$$

By the fact that $\|R_{Tz}\|^{t-j} \leq 1$,

$$\begin{aligned} \sup_{1\leq t\leq T} \mathbb{E}\left\|\eta_{T,t}^{(1)}\right\|^2 &\leq \left(\sup_{1\leq t\leq T}\sum_{i=1}^{t}\|R_{Tz}\|_{\infty}^{t-i}\right)\left(\sup_{1\leq t\leq T}\sum_{j=1}^{t-1}\left\|R_{Tt}^{(j-k)}\right\|_{\infty}\right)^2\sum_{l=-\infty}^{+\infty}\|\Gamma_{u_x}(l)\| \\ &= O_p\left(T^{2\alpha+\gamma}\right). \end{aligned}$$

We complete the proof. ■

Proof of Lemma 1.7 : Since $\left\langle \frac{1}{\sqrt{T^{\alpha}}} x_{[Tr]} \right\rangle \to \int_0^{\infty} e^{zC(r)}\Omega_{xx}e^{zC(r)}dz$ holds, we have

$$\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T} x_{t-1}x_{t-1}' \Rightarrow \int_0^1\left[\int_0^{\infty} e^{zC(r)}\Omega_{xx}e^{zC(r)}dz\right]dr.$$

(i) According to the decomposition of $\tilde{z}_{t-1}$ with rate restriction $\alpha<\gamma$, if $x_0=0$ holds, we have

$$\begin{aligned} \frac{1}{T^{\frac{1+\alpha}{2}}}\left(\sum_{t=1}^{T}\tilde{z}_{t-1}u_{0t}-\sum_{t=1}^{T}x_{t-1}u_{0t}\right) &= \frac{1}{T^{\frac{1+\alpha}{2}}}\sum_{t=1}^{T}\left(\frac{C_z}{T^{\gamma}}\eta_{T,t-1}^{(1)}-R_{Tz}^{t}x_0\right)u_{0t} \\ &= \frac{1}{T^{\frac{1+\alpha}{2}}}\sum_{t=1}^{T}\frac{C_z}{T^{\gamma}}\eta_{T,t-1}^{(1)}u_{0t}-\frac{1}{T^{\frac{1+\alpha}{2}}}\sum_{t=1}^{T}R_{Tz}^{t}x_0u_{0t} \\ &= \frac{1}{T^{\frac{1+\alpha}{2}}}\sum_{t=1}^{T}\frac{C_z}{T^{\gamma}}\eta_{T,t-1}^{(1)}u_{0t}+o_p(1) \\ &= \frac{1}{T^{\frac{1+\alpha}{2}+\gamma}}\sum_{t=1}^{T}C_z\eta_{T,t-1}^{(1)}u_{0t}+o_p(1). \qquad \text{(A.16)} \end{aligned}$$

The leading term of (A.16) is asymptotically negligible, since

$$\left\|\frac{1}{T^{\frac{1+\alpha}{2}+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}u_{0t}\right\| \leq \sup_{1\leq t\leq T}\left\|\frac{u_{0t}}{\sqrt{T}}\right\|_{\infty}\sup_{1\leq t\leq T}\left\|\frac{\eta_{T,t-1}^{(1)}}{T^{\alpha+\frac{\gamma}{2}}}\right\|\frac{1}{T^{\frac{\gamma-\alpha}{2}}} = O_p\left(\frac{1}{T^{\frac{\gamma-\alpha}{2}}}\right) = o_p(1).$$

(ii) By the decomposition of $\tilde{z}_{t-1}$ with rate restriction $\alpha<\gamma$, if $x_0=0$, we have the following argument for the sample moment

$$\begin{aligned} &\frac{1}{T^{1+\alpha}}\left(\sum_{t=1}^{T}\tilde{z}_{t-1}x_{t-1}'-\sum_{t=1}^{T}x_{t-1}x_{t-1}'\right) \\ =\ &\frac{1}{T^{1+\alpha}}\left[\frac{C_z}{T^{\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}x_{t-1}'-\sum_{t=1}^{T}R_{Tz}^{t}x_0x_{t-1}'\right] \\ \leq\ &\frac{C_z}{T^{1+\alpha+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}x_{t-1}'+O_p\left(\frac{1}{T^{1-\gamma+\frac{\alpha}{2}}}\right). \qquad \text{(A.17)} \end{aligned}$$

For the leading term of (A.17), we achieve the asymptotic upper bound as follows

$$\left\|\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}x_{t-1}'\right\| \leq \sum_{t=1}^{T}\left(\mathbb{E}\left\|\eta_{T,t-1}^{(1)}\right\|^{2}\right)^{1/2}\left(\mathbb{E}\left\|x_{t-1}\right\|^{2}\right)^{1/2} = O_p\left(T^{1+\alpha+\frac{\gamma}{2}+\frac{\alpha}{2}}\right).$$

Therefore, $\frac{C_z}{T^{1+\alpha+\gamma}}\left\|\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}x_{t-1}'\right\| = O_p\left(\frac{1}{T^{\frac{\gamma-\alpha}{2}}}\right) = o_p(1)$.

(iii) The sample moment, $\sum_{t=1}^{T}\tilde{z}_{t-1}\tilde{z}_{t-1}'$, follows the decomposition as,

$$\begin{aligned}
&\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}\tilde{z}_{t-1}\tilde{z}_{t-1}' \\
=\ &\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}\left(x_{t-1}-R_{Tz}^{t}x_0+\frac{C_z}{T^{\gamma}}\eta_{T,t-1}^{(1)}\right)\left(x_{t-1}-R_{Tz}^{t}x_0+\frac{C_z}{T^{\gamma}}\eta_{T,t-1}^{(1)}\right)' \\
=\ &\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}\begin{bmatrix} x_{t-1}x_{t-1}'+R_{Tz}^{t}x_0x_0'R_{Tz}^{t}+\frac{C_z}{T^{2\gamma}}\eta_{T,t-1}^{(1)}\eta_{T,t-1}^{(1)\prime}C_z \\ -x_{t-1}x_0'R_{Tz}^{t}+x_{t-1}\eta_{T,t-1}^{(1)\prime}\frac{C_z}{T^{\gamma}}-R_{Tz}^{t}x_0x_{t-1}' \\ -R_{Tz}^{t}x_0\eta_{T,t-1}^{(1)\prime}\frac{C_z}{T^{\gamma}}+\frac{C_z}{T^{\gamma}}\eta_{T,t-1}^{(1)}x_{t-1}'-\frac{C_z}{T^{\gamma}}\eta_{T,t-1}^{(1)}x_0'R_{Tz}^{t}\end{bmatrix} \quad \text{(A.18)}
\end{aligned}$$

It is sufficient to show that the spectral norm of each term in (A.18) above is $o_p(1)$. Therefore it is sufficient to show $\left\|\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}R_{Tz}^{t}x_0x_0'R_{Tz}^{t}\right\|$, $\left\|\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}R_{Tz}^{t}x_0x_0'R_{Tz}^{t}\right\|$, $\left\|\frac{1}{T^{1+\alpha+2\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}\eta_{T,t-1}^{(1)\prime}\right\|$ and the remaining terms to be asymptotically negligible by Cauchy-Schwarz inequality.

For $\left\|\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}R_{Tz}^{t}x_0x_0'R_{Tz}^{t}\right\|$, we have

$$\left\|\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}R_{Tz}^{t}x_0x_0'R_{Tz}^{t}\right\| \leq \left\|x_0\right\|_{\infty}^{2}\frac{1}{T^{1+\alpha}}\sum_{t=1}^{T}\left\|R_{Tz}\right\|^{2t} = O_p\left(\frac{1}{T^{1+\alpha-\gamma}}\right).$$

For $\left\|\frac{1}{T^{1+\alpha+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}x_0'R_{Tz}^{t}\right\|$, we have

$$\left\|\frac{1}{T^{1+\alpha+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}x_0'R_{Tz}^{t}\right\| \leq \left\|x_0\right\|_{\infty}\frac{\left\|\eta_{T,t-1}^{(1)}\right\|}{T^{\alpha+\frac{\gamma}{2}}}\frac{1}{T^{1+\frac{\gamma}{2}}}\sum_{t=1}^{T}\left\|R_{Tz}^{t}\right\|_{\infty} = O_p\left(\frac{1}{T^{\frac{1-\gamma}{2}}}\right).$$

Similarly, we have

$$\left\|\frac{1}{T^{1+\alpha+2\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(1)}\eta_{T,t-1}^{(1)\prime}\right\| = O_p\left(\frac{1}{T^{\gamma-\alpha}}\right).$$

Then we complete the proof. ■

Kernel Estimation

For simplicity, we adopt a scalar FLUR model in the scalar form ($n = 1$),

$$x_s = R_{Ts} x_{s-1} + u_{xs},$$

where $r_s := \frac{s}{T}$, $R_{Ts} := \exp\left(\frac{C(r_s)}{T}\right)$ and $s = 1, 2, ..., T$. We intend to estimate R_{Tt} for any $t = 1, 2, ..., T$ as

$$\hat{R}_{Tt} = \arg\min_{R_{Tt}} \sum_{s=1}^{T} \left(x_s - R_{Tt} x_{s-1}\right)^2 \overline{K}_h(r_s - r), \tag{A.19}$$

where $r_s := \frac{s}{T}$, $r := \frac{t}{T}$ and $\overline{K}_h(\cdot) := \frac{1}{h}\overline{K}\left(\frac{\cdot}{h}\right)$. We define h as the bandwidth for local average. To establish consistent estimation for R_{Tt}, we impose the following assumptions.

Assumption A.1. *(i) The distance function $C(r)$ is 2nd-order differentiable in r for all $r \in [0, 1]$.*

(ii) The kernel function $\overline{K}(\cdot)$ is symmetric and bounded with a compact support $[-1, 1]$. The kernel function is Lipschitz continuous.

(iii) Bandwidth $h \to 0$ has the rate restriction $T\sqrt{h} \to \infty$.

(iv) The error term u_{xt} follows a stationary linear process as

$$u_{xt} = \sum_{j=0}^{\infty} F_{xj} \epsilon_{t-j} \text{ with } \sum_{j=0}^{\infty} j \left|F_{xj}\right| < \infty,$$

where $\epsilon_t \sim$i.i.d.$(0, \Sigma)$ with finite $p \geq 4$ moment. The two-sided long-run variance $\Omega_{xx} := \left(\sum_{j=0}^{\infty} F_{xj}\right) \Sigma \left(\sum_{j=0}^{\infty} F_{xj}\right)$.

(v) We define $\nu(K) := \int \overline{K}(v)dv$ and $\mu_j(K) := \int v^j \overline{K}(v)dr$.

We provide several remarks here: First, we employ the symmetric kernel function $\overline{K}(\cdot)$. The symmetry of kernel functions results in the fact $\int v\overline{K}(v) = 0$ as a standard requirement for discussing the approximation bias. Second, different from stationary cases which require $Th \to \infty$, we impose $T\sqrt{h} \to \infty$ for our discussions. Smaller bandwidth results from the instability of nonstationarity.

By minimizing (A.19), we define the estimator for the slope as,

$$\begin{aligned}
\hat{R}_{Tt} - R_{Tt} &= \left(\sum_{s=1}^{T} x_{s-1}^2 \overline{K}_h(r_s - r)\right)^{-1} \sum_{s=1}^{T} x_{s-1} u_{xs} \overline{K}_h(r_s - r) \\
&\quad + \left(\sum_{s=1}^{T} x_{s-1}^2 \overline{K}_h(r_s - r)\right)^{-1} \sum_{s=1}^{T} x_{s-1}^2 \overline{K}_h(r_s - r)\left(R_{Ts} - R_{Tt}\right) \\
&=: \left(S_{nT,t}\right)^{-1} T_{nT,t} + \left(S_{nT,t}\right)^{-1} R_{nT,t},
\end{aligned}$$

where $S_{nT,t} := \sum_{s=1}^{T} x_{s-1}^2 \overline{K}_h(r_s - r)$, $T_{nT,t} := \sum_{s=1}^{T} x_{s-1} u_{xs} \overline{K}_h(r_s - r)$ and $R_{nT,t} := \sum_{s=1}^{T} x_{s-1}^2 \overline{K}_h(r_s - r)\left(R_{Ts} - R_{Tt}\right)$. We can prove the consistency of $\widehat{R}_{Tt}$ by the following theorem.

Theorem A.1. *Let Assumption A.1 hold. As $T \to \infty$, for any t,*

$$T\sqrt{h}\left(\hat{R}_{Tt} - R_{Tt}\right) \Rightarrow \mathcal{MN}\left(0, \frac{\Omega_{xx}\nu(K)}{rK_c^2(1)}\right),$$

where $K_c(r)$ follows the stochastic process as $dK_c(r) = C(r)K_c(r)dr + \Omega_{xx}^{\frac{1}{2}} dW_x(r)$.

We perceive that for each $t \in \{1, 2, .., T\}$, $\hat{R}_{Tt}$ is converging to R_{Tt} at a slower rate than the rate of least squares. Different from the time-invariant case in Phillips (1987), the asymptotic normality of $\hat{R}_{Tt}$ locates at the origin. The reason why asymptotic normality happens is due to the fact that $\sqrt{\frac{h}{T}} \sum_{s=1}^{T} \overline{K}_h(r_s - r) u_{xs}$ is asymptotically orthogonal to $\frac{1}{\sqrt{T}} \sum_{s=1}^{T} u_{xs}$. Surprisingly, serial correlations of innovations do not introduce asymptotic bias.

For the stationary case, there exists an asymptotically non-negligible term of the order $O_p(h^2)$, which contributes to the approximation bias. For a nonstationary time series model as FLUR, the asymptotic bias of order $\min\{O_p\left(\frac{\sqrt{h}}{T}\right)$, $O_p\left(T^{-\frac{3}{2}+\frac{1}{q}+\delta}\right)\}$ is diminishing at a rate faster than the limiting normality, $O_p\left(\frac{1}{T\sqrt{h}}\right)$. Therefore, no bias-corrected procedure is on request.

For the given h, we define the standard leave-one-out estimate as

$$\widehat{R}_{Tt}(\phi|h) = \left[\sum_{s=1, s\neq t}^{T} y_{s-1}^2 \overline{K}_h\left(r_s - \phi\right)\right]^{-1} \left[\sum_{s=1, s\neq t}^{T} y_{s-1} y_s \overline{K}_h\left(r_s - \phi\right)\right].$$

The cross-validation function is then written as

$$\mathbf{CV}_T(h) = \frac{1}{T}\sum_{t=1}^{T}\left[y_t - \widehat{R}_{Tt}(\frac{t}{T}|h)y_{t-1}\right]^2.$$

We can find an optimal bandwidth by mininmizing the cross-validation function,

$$\widehat{h}_{\mathbf{CV}} = \arg\min_h \mathbf{CV}_T(h).$$

We collect technical proofs for kernel estimations in the following lemmas.

Lemma A.2. *Let Assumption A.1 hold. As $T \to \infty$, we have*

$$\frac{1}{T^2}S_{nT,t} \Rightarrow rK_c^2(1).$$

Proof of Lemma A.2: Define $r(T) := [(r-h)T]$. We have,

$$\begin{aligned}\sum_{s=1}^{T} x_{s-1}^2 \overline{K}_h(r_s - r) &= \sum_{s=1}^{T} x_{r(T)}^2 \overline{K}_h(r_s - r) + 2\sum_{s=1}^{T} x_{r(T)}\left(x_{s-1} - x_{r(T)}\right)\overline{K}_h(r_s - r) \\ &\quad + \sum_{s=1}^{T}\left(x_{s-1} - x_{r(T)}\right)^2 \overline{K}_h(r_s - r). \qquad \text{(A.20)}\end{aligned}$$

For the first term of (A.20), according to Bykhovskaya and Phillips (2018),

$$\begin{aligned}\frac{1}{T^2}\sum_{s=1}^{T} x_{r(T)}^2 \overline{K}_h(r_s - r) &= \frac{r(T)}{T}\left(\frac{x_{r(T)}^2}{r(T)}\right)\frac{1}{T}\sum_{s=1}^{T}\overline{K}_h(r_s - r) \\ &\Rightarrow rK_c^2(1)\int \overline{K}(v)\,dv \\ &= rK_c^2(1), \qquad \text{(A.21)}\end{aligned}$$

where the condition $\int \overline{K}(v)\,dv = 1$ holds. For the second term of (A.20),

$$x_{s-1} - x_{r(T)} = \sum_{i=r(T)+1}^{s-1}\left(\prod_{m=i+1}^{s-1}\exp\left(\frac{C(m/T)}{T}\right)\right)u_{xi}, \qquad \text{(A.22)}$$

where $s \in [r(T) - [hT], r(T) + [hT]]$ holds. Based on (A.22),

$$\begin{aligned}&\sup_{r(T)-[hT]\le s-1\le r(T)+[hT]}\left|\frac{x_{s-1} - x_{r(T)}}{\sqrt{2[hT]}}\right| \\ =\ &\sup_{r(T)-[hT]\le s-1\le r(T)+[hT]}\left|\frac{\sum_{i=r(T)+1}^{s-1}\left(\prod_{m=i+1}^{s-1}\exp\left(\frac{C(m/T)}{T}\right)\right)u_{xi}}{\sqrt{2[hT]}}\right| \\ \Rightarrow\ &\sup_{0\le d\le 1}\left|K_c(d)\right|. \qquad \text{(A.23)}\end{aligned}$$

Equivalently, we have

$$\sup_{r(T)-[hT]\leq s-1\leq r(T)+[hT]} \left|x_{s-1}-x_{r(T)}\right| = O_p\left(\sqrt{[hT]}\right).$$

By Assumption A.1 (ii) and the fact that $\overline{K}_h(\cdot)$ is defined on a compact support, we can justify the asymptotic negligibility as,

$$\begin{aligned}
& \left|\sum_{s=1}^{T} x_{r(T)}\left(x_{s-1}-x_{r(T)}\right)\overline{K}_h\left(r_s-r\right)\right| \\
\leq\ & \left|x_{r(T)}\right|\left|x_{s-1}-x_{r(T)}\right|\left|\sum_{s=r(T)-[hT]+1}^{r(T)+[hT]}\overline{K}_h\left(r_s-r\right)\right| \\
=\ & O_p\left(\sqrt{T}\right)\times O_p\left(T\right)\times O_p\left(\sqrt{Th}\right) \\
=\ & O_p\left(\sqrt{T^4h}\right)=o_p\left(T^2\right).
\end{aligned} \tag{A.24}$$

Similarly, for the third term of (A.20),

$$\begin{aligned}
& \left|\sum_{s=1}^{T}\left(x_{s-1}-x_{r(T)}\right)^2\overline{K}_h\left(r_s-r\right)\right| \\
\leq\ & \left|x_{s-1}-x_{r(T)}\right|^2\left|\sum_{s=r(T)-[hT]+1}^{r(T)+[hT]}\overline{K}_h\left(r_s-r\right)\right| \\
\leq\ & O_p\left(\sqrt{Th}\right)\times O_p\left(\sqrt{Th}\right)\times O_p\left(T\right) \\
=\ & O_p\left(T^2h\right)=o_p\left(T^2\right).
\end{aligned} \tag{A.25}$$

Combining (A.21), (A.24) and (A.25), we prove the lemma. ■

Lemma A.3. *Let Assumption A.1 hold. As* $T\to\infty$,

(i) $\sqrt{\frac{h}{T}}\sum_{s=1}^{T}\overline{K}_h(r_s-r)u_{xs}\xrightarrow{p}Z_x:=\mathcal{N}\left(0,\Omega_{xx}\nu(K)\right)$.

(ii) $\sqrt{\frac{h}{T}}\sum_{s=1}^{T}\overline{K}_h(r_s-r)u_{xs}$ *is asymptotically orthogonal to* $\frac{1}{\sqrt{T}}\sum_{s=1}^{T}u_{xs}$.

Proof of Lemma A.3: (i) Following the steps in Proposition A.2 of Phillips et al. (2017), we complete the proof.

(ii) We have justified the asymptotic normality of $\sqrt{\frac{h}{T}}\sum_{s=1}^{T}\overline{K}_h(r_s-r)u_{xs}$. By Beveridge-Nelson-Phillips decomposition in Phillips and Solo (1992), $\frac{1}{\sqrt{T}}\sum_{s=1}^{T}u_{xs}$ asymptotically converges to a Gaussian process. Therefore, the asymptotical orthogonality between two Gaussian processes is given by

$$\begin{aligned}
& \mathbb{E}\left(\sqrt{\frac{h}{T}}\sum_{s=1}^{T}\overline{K}_h(r_s-r)u_{xs}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{xt}\right) \\
= & \mathbb{E}\left(\frac{\sqrt{h}}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\overline{K}_h(r_s-r)u_{xs}u_{xt}\right) \\
= & \frac{\sqrt{h}}{T}\sum_{s=1}^{T}\sum_{l=-\infty}^{\infty}\overline{K}_h(r_s-r)\mathbb{E}\left(u_{xs}u_{x,s-l}\right) \\
= & \left(\frac{1}{T}\sum_{s=1}^{T}\overline{K}_h(r_s-r)\right)\left(\sum_{l=-\infty}^{\infty}\mathbb{E}\left(u_{xs}u_{x,s-l}\right)\right)\sqrt{h} \\
= & O_p\left(\sqrt{h}\right),
\end{aligned}$$

where $\sum_{l=-\infty}^{\infty}\mathbb{E}\left(u_{xs}u_{x,s-l}\right)=O_p(1)$ due to Assumption A.1 (iv). ■

Lemma A.4. *Let Assumption A.1 hold. As* $T\to\infty$*,*

$$\sqrt{\frac{h}{T}}T_{nT,t}\Rightarrow\sqrt{r}K_c(1)Z_x,$$

where $Z_x=_d\mathcal{N}\left(0,\Omega_{xx}\nu(K)\right).$

Proof of Lemma A.4: Asymptotically, we have the approximations as

$$\begin{aligned}
& \sqrt{\frac{h}{T}}\sum_{s=1}^{T}x_{s-1}u_{xs}\overline{K}_h(r_s-r) \\
= & \sqrt{\frac{h\cdot r(T)}{T}}\sum_{s=1}^{T}\frac{x_{r(T)}}{\sqrt{r(T)}}u_{xs}\overline{K}_h(r_s-r) \qquad \text{(A.26)} \\
& +\sqrt{\frac{h\cdot r(T)}{T}}\sum_{s=1}^{T}\frac{x_{s-1}-x_{r(T)}}{\sqrt{r(T)}}u_{xs}\overline{K}_h(r_s-r) \\
= & \sqrt{\frac{h\cdot r(T)}{T}}\frac{x_{r(T)}}{\sqrt{r(T)}}\sum_{s=1}^{T}u_{xs}\overline{K}_h(r_s-r)+O_p(\sqrt{h}) \\
\Rightarrow & \sqrt{r}K_c(1)Z_x. \qquad \text{(A.27)}
\end{aligned}$$

For the second equality of (A.27), we apply the following argument

$$\begin{aligned}
& \left|\sqrt{\frac{h\cdot r(T)}{T}}\sum_{s=1}^{T}\frac{x_{s-1}-x_{r(T)}}{\sqrt{r(T)}}u_{xs}\overline{K}_h(r_s-r)\right| \\
\leq & r\left(1+o(1)\right)\sqrt{\frac{1}{hT^2}}\times O_p(\sqrt{Th})O_p(\sqrt{Th}) \\
= & O_p(\sqrt{h}).
\end{aligned}$$

For the last line, we employ Lemma B.2 and Bykhovskaya and Phillips (2018). ■

Lemma A.5. *Let Assumption A.1 hold. As $T \to \infty$ and $h \to 0$,*

$$\frac{1}{T} R_{nT,t} \Rightarrow \frac{rh^2}{2} K_c^2(1)\mu_2(K)C^{(2)}(r) + O_p(T^{-\frac{1}{2}+\frac{1}{q}+\delta}) + O_p(\sqrt{h}),$$

where $C^{(2)}(\cdot)$ is the 2nd-order derivative of $C(\cdot)$.

Proof of Lemma A.5: It is shown that

$$\begin{aligned}
& \frac{1}{T}\sum_{s=1}^{T} x_{s-1}^2 \overline{K}_h(r_s - r)\left(R_{Ts} - R_{Tt}\right) \\
= & \frac{1}{T^2}\sum_{s=1}^{T} x_{s-1}^2 \overline{K}_h(r_s - r)\left(C(r_s) - C(r)\right) \\
= & \frac{1}{T^2}\sum_{s=1}^{T} x_{r(T)}^2 \overline{K}_h(r_s - r)\left(C(r_s) - C(r)\right) \\
& + \frac{1}{T^2}\sum_{s=1}^{T} \left(x_{s-1}^2 - x_{r(T)}^2\right) \overline{K}_h(r_s - r)\left(C(r_s) - C(r)\right) \\
= & \frac{1}{T^2}\sum_{s=1}^{T} x_{r(T)}^2 \overline{K}_h(r_s - r)\left(C(r_s) - C(r)\right) + O_p(\sqrt{h}) \\
= & rK_c^2(1)\int \overline{K}_h(r_s - r)\left(C(r_s) - C(r)\right) dr_s + O_p(T^{-\frac{1}{2}+\frac{1}{q}+\delta}) + O_p(\sqrt{h}) \\
= & rK_c^2(1)\int \overline{K}_h(v)\left(C(r + hv) - C(r)\right) dv + O_p(T^{-\frac{1}{2}+\frac{1}{q}+\delta}) + O_p(\sqrt{h}) \\
= & rK_c^2(1)\int \overline{K}_h(v)\left(C(r) - C(r) + C^{(1)}(r)hv + \frac{1}{2}C^{(2)}(r)h^2v^2 + O(h^3)\right) dv + \\
& O_p(T^{-\frac{1}{2}+\frac{1}{q}+\delta}) + O_p(\sqrt{h}) \\
= & \frac{rh^2}{2} K_c^2(1)\mu_2(K)C^{(2)}(r) + O_p(T^{-\frac{1}{2}+\frac{1}{q}+\delta}) + O_p(\sqrt{h}) \\
= & O_p(T^{-\frac{1}{2}+\frac{1}{q}+\delta}) + O_p(\sqrt{h}).
\end{aligned}$$

The third equality is due to (A.23). The fourth equality is due to the uniform strong approximation to Brownian motions for each $\delta > 0$ (See Lemma 3.1 of Phillips (2007)). The last equality is due to the symmetry of the kernel function. ■

Proof of Theorem A.1: Compared with the stationary case, the estimation bias in this model is of lower order. It diminishes at a faster rate than the limiting normality,

$$\frac{\sum_{s=1}^{T} x_{s-1}^2 \overline{K}_h \left(r_s - r\right)\left(R_{Ts} - R_{Tt}\right)}{\sum_{s=1}^{T} x_{s-1}^2 \overline{K}_h \left(r_s - r\right)}$$
$$= \min\left\{O_p\left(\frac{h}{T}\right), O_p\left(T^{-\frac{3}{2}+\frac{1}{q}+\delta}\right)\right\} << O_p\left(\frac{1}{T\sqrt{h}}\right).$$

Therefore the asymptotical normality dominates. We complete our proof. ■

Appendix B

Technical Results for Chapter 2

Proof of the main results

Orthogonal basis

Proof of Lemma 2.1: (a) We deal with term, $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\exp\left(\frac{\sum_{j=t}^{[Tr]}C(j/T)}{T}\right)u_{xt}$. By a strong approximation, we can enlarge the original probability space and construct a Brownian motion vector as $\omega := BM\left(\Sigma\right)$. With $\mathbb{E}\left\|u_{xt}\right\|^{p} < \infty$, we have

$$\sup_{0\leq r\leq 1}\left\|\eta_T(r)-\omega(r)\right\| = O_{a.s.}\left(T^{-\alpha}\right),$$

where $\eta_T(r) := \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}u_{xt}$. Define $\frac{e_t}{\sqrt{T}} = \omega\left(\frac{t}{T}\right)-\omega\left(\frac{t-1}{T}\right)$, whose distribution is i.i.d. normal with $\mathbb{E}\left(e_t e_t^{'}\right) = \Sigma$. We generate another FLUR process z_t as

$$z_t = R_{Tt}z_{t-1}+e_t,\ t=1,2,...,T.$$

For each $r\in[0,1]$, as $T\to\infty$, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\exp\left(\frac{\sum_{j=t}^{[Tr]}C\left(j/T\right)}{T}\right)e_t \Rightarrow K_c(r).$$

For $\eta_T(r)$, we have

$$
\begin{aligned}
& \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \exp\left(\frac{\sum_{j=t}^{[Tr]} C(j/T)}{T}\right) u_{xt} \\
= & \sum_{t=1}^{[Tr]} \exp\left(\frac{\sum_{j=t}^{[Tr]} C(j/T)}{T}\right)\left(\eta_T\left(\frac{t}{T}\right) - \eta_T\left(\frac{t-1}{T}\right)\right) \\
= & \sum_{t=1}^{[Tr]} \exp\left(\frac{\sum_{j=t}^{[Tr]} C(j/T)}{T}\right)\eta_T\left(\frac{t}{T}\right) - \sum_{t=1}^{[Tr]} \exp\left(\frac{\sum_{j=t}^{[Tr]} C(j/T)}{T}\right)\eta_T\left(\frac{t-1}{T}\right) \\
= & \exp\left(\frac{C([Tr]/T)}{T}\right)\eta_T\left(\frac{[Tr]}{T}\right) + \\
& \sum_{t=1}^{[Tr]-1}\left(\exp\left(\frac{\sum_{j=t}^{[Tr]} C(j/T)}{T}\right) - \exp\left(\frac{\sum_{j=t+1}^{[Tr]} C(j/T)}{T}\right)\right)\eta_T\left(\frac{t}{T}\right) \\
= & \exp\left(\frac{C([Tr]/T)}{T}\right)\eta_T\left(\frac{[Tr]}{T}\right) + \sum_{t=1}^{[Tr]-1} \exp\left(\frac{\sum_{j=t+1}^{[Tr]} C(j/T)}{T}\right)\frac{C(t/T)}{T}\eta_T\left(\frac{t}{T}\right).
\end{aligned}
$$

Similarly, for $\omega(\cdot)$, we justify that

$$
\begin{aligned}
& \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \exp\left(\frac{\sum_{j=t}^{[Tr]} C(j/T)}{T}\right) e_t \\
= & \exp\left(\frac{C([Tr]/T)}{T}\right)\omega\left(\frac{[Tr]}{T}\right) + \sum_{t=1}^{[Tr]-1} \exp\left(\frac{\sum_{j=t+1}^{[Tr]} C(j/T)}{T}\right)\frac{C(t/T)}{T}\omega\left(\frac{t}{T}\right).
\end{aligned}
$$

Then we can argue that

$$
\begin{aligned}
& \sup_{0\le r\le 1}\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \exp\left(\frac{\sum_{j=t}^{[Tr]} C(j/T)}{T}\right) u_{xt} - \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \exp\left(\frac{\sum_{j=t}^{[Tr]} C(j/T)}{T}\right) e_t \right\| \\
\le & \sup_{0\le r\le 1}\|\eta_T(r) - \omega(r)\| \sup_{0\le r\le 1}\left\|\exp\left(\frac{C([Tr]/T)}{T}\right)\right\|_\infty \\
& + \sup_{0\le r\le 1}\left\|\sum_{t=1}^{[Tr]-1} \exp\left(\frac{\sum_{j=t+1}^{[Tr]} C(j/T)}{T}\right)\frac{C(t/T)}{T}\left(\eta_T\left(\frac{t}{T}\right) - \omega\left(\frac{t}{T}\right)\right)\right\| \\
\le & M \sup_{0\le r\le 1}\|\eta_T(r) - \omega(r)\| + \sup_{0\le r\le 1}\left|\int_0^r \exp\left(\int_p^r \bar{c}da\right)\bar{c}dp\right| \sup_{0\le r\le 1}\|\eta_T(r) - \omega(r)\| \\
= & O_p\left(T^{-\alpha}\right),
\end{aligned}
$$

where $M(>0)$ is some constant. The notation $\|\cdot\|_\infty$ denotes the infinity norm. The proof for Lemma 2.1 concludes. ■

The Proof of Lemma 2.2: According to Lemma 1 in Bierens and Martins (2010), we have $g(t)\left(:= \overline{g}\left(\frac{t}{T}\right)\right)$ which is a qth-order differentiable function with

$\overline{g}^{(q)}(x) = \frac{d^q \overline{g}(x)}{(dx)^q}$ and $\int_0^1 \left\| \overline{g}^{(q)}(x) \right\|^2 dx < \infty$. For any $k \geq 1$, we have

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \| g(t) - g_{k,T}(t) \|^2 \leq \frac{\int_0^1 \left\| \overline{g}^{(q)}(x) \right\|^2 dx}{\pi^{2q} (k+1)^{2q}}.$$

Here $g_{k,T}(\cdot)$ is a k-dimensional approximation to $g(\cdot)$ using orthogonal polynomials. ■

Time-series estimator

Lemma B.1. *Let Assumptions 2.1 and 2.2 hold. As $T \to \infty$, we have*

(a)

$$\begin{aligned} &\frac{1}{T^2} \sum_{t=1}^{T} y_{k,t-1} y_{k,t-1}^{'} \\ = &\int_0^1 \left[\left(f_k \left(\frac{[Tr]}{T} \right) f_k^{'} \left(\frac{[Tr]}{T} \right) \right) \otimes \left(K_{T,c}^{*}(r) K_{T,c}^{*'}(r) \right) \right] dr; \end{aligned}$$

(b)

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} y_{k,t-1} = \int_0^1 \left[f_k \left(\frac{[Tr]}{T} \right) \otimes K_{T,c}^{*}(r) \right] dr;$$

(c)

$$\frac{1}{T} \sum_{t=1}^{T} y_{k,t-1} u_{xk,t}^{'} = \int_0^1 \left[f_k \left(\frac{[Tr]}{T} \right) \otimes K_{T,c}^{*}(r) \right] dW_{xT}^{'}(r);$$

(d)

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^{T} y_{k,t-1} \left(u_{xk,t} - u_{xt} \right)^{'} \\ = &\int_0^1 \left[f_k \left(\frac{[Tr]}{T} \right) \otimes \left(K_{T,c}^{*}(r) K_{T,c}^{*'}(r) \right) \right] \left(R_T \left(\frac{[Tr]}{T} \right) - R_T^{(k)} \left(\frac{[Tr]}{T} \right) \right) dr. \end{aligned}$$

The Proof of Lemma B.1: We define $t = [Tr]$. For (a), we have the following argument as

$$\begin{aligned} & \frac{1}{T^2}\sum_{t=1}^{T}\left[f_k\left(\frac{[Tr]}{T}\right)\otimes y_{t-1}\right]\left[f_k\left(\frac{[Tr]}{T}\right)\otimes y_{t-1}\right]' \\ = & \frac{1}{T}\sum_{t=1}^{T}\left(f_k\left(\frac{[Tr]}{T}\right)f_k'\left(\frac{[Tr]}{T}\right)\right)\otimes\left(\frac{y_{t-1}}{\sqrt{T}}\frac{y_{t-1}'}{\sqrt{T}}\right) \\ = & \int_0^1\left[\left(f_k\left(\frac{[Tr]}{T}\right)f_k'\left(\frac{[Tr]}{T}\right)\right)\otimes\left(K_{T,c}^*(r)\,K_{T,c}^{*'}(r)\right)\right]dr. \end{aligned}$$

The argument for (b) is trivial. For (c), we have

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T}\left[f_k\left(\frac{[Tr]}{T}\right)\otimes y_{t-1}\right]u_{xt}' \\ = & \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[f_k\left(\frac{[Tr]}{T}\right)\otimes y_{t-1}\right]\frac{u_{xt}'}{\sqrt{T}} \\ = & \int_0^1\left[f_k\left(\frac{[Tr]}{T}\right)\otimes K_{T,c}^*(r)\right]dW_{xT}'(r), \end{aligned}$$

where $W_{xT}'(r)$ is the empirical process driven by $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{xt}'$. For (d), we have

$$\begin{aligned} & \frac{1}{T^2}\sum_{t=1}^{T}f_k\left(\frac{[Tr]}{T}\right)\otimes y_{t-1}\left(u_{xk,t}-u_{xt}\right)' \\ = & \frac{1}{T^2}\sum_{t=1}^{T}f_k\left(\frac{[Tr]}{T}\right)\otimes y_{t-1}y_{t-1}'\left(R_T\left(\frac{t}{T}\right)-R_T^{(k)}\left(\frac{t}{T}\right)\right) \\ = & \int_0^1\left[f_k\left(\frac{[Tr]}{T}\right)\otimes\left(K_{T,c}^*(r)\,K_{T,c}^{*'}(r)\right)\right]\left(R_T\left(\frac{[Tr]}{T}\right)-R_T^{(k)}\left(\frac{[Tr]}{T}\right)\right)dr. \end{aligned}$$

■

We define the following sample moments. For each value of k, T, we have

$$\begin{aligned} M_{Tk} &:= \int_0^1\left[\left(f_k\left(\frac{[Tr]}{T}\right)f_k'\left(\frac{[Tr]}{T}\right)\right)\otimes\left(K_{T,c}^*(r)K_{T,c}^{*'}(r)\right)\right]dr; \\ N_{Tk} &:= \int_0^1\left[f_k\left(\frac{[Tr]}{T}\right)\otimes K_{T,c}^*(r)\right]dW_{xT}'(r); \\ O_{Tk} &:= \int_0^1\left[f_k\left(\frac{[Tr]}{T}\right)\otimes\left(K_{T,c}^*(r)K_{T,c}^{*'}(r)\right)\right]\left(R_T\left(\frac{[Tr]}{T}\right)-R_T^{(k)}\left(\frac{[Tr]}{T}\right)\right)dr. \end{aligned}$$

If k is finite, as $T \to \infty$, we have

$$\begin{aligned}
M_k &:= \int_0^1 \left[\left(f_k(r) f_k^{'}(r) \right) \otimes \left(K_c(r) K_c^{'}(r) \right) \right] dr; \\
N_k &:= \int_0^1 \left[f_k(r) \otimes K_c(r) \right] dW_x^{'}(r); \\
O_k &:= \int_0^1 \left[f_k(r) \otimes \left(K_c(r) K_c^{'}(r) \right) \right] \left(R_T(r) - R_T^{(k)}(r) \right) dr.
\end{aligned}$$

Based on the above notations, we can justify the following intermediate results.

Lemma B.2. *Let Assumptions 2.1 and 2.2 hold. Note that α is defined in Lemma 2.2. As $T \to \infty$, we have*

(a)

$$\begin{aligned}
\|M_k\| &= O_p(1),\ \left\|M_k^{-1}\right\| = O_p(1),\ M_k = O_p(1), \\
M_k^{-1} &= O_p(1),\ \text{and}\ \|M_k - M_{Tk}\| = O_p(T^{-\alpha}k);
\end{aligned}$$

(b)

$$\|N_k\| = O_p(\sqrt{k})\ \text{and}\ \|N_k - N_{Tk}\| = O_p(T^{-\alpha}\sqrt{k});$$

(c)

$$\|O_k\| = O_p(k^{\frac{1}{2}-q})\ \text{and}\ \|O_{Tk}\| = O_p(k^{\frac{1}{2}-q}).$$

The Proof of Lemma B.2: This proof follows the approach in Park and Hahn (1999). If we want to prove $M_k = O_p(1)$, it is equivalent to justify that

$$\sup_{k \geq 1} \int_0^1 \left(f_k(r) f_k^{'}(r) \right) \otimes \left(K_c(r) K_c^{'}(r) \right) dr = O_p(1).$$

By orthogonality of cosine functions, we can verify that

$$\underline{M} \leq \lambda_{\min} \left(\int_0^1 f_k(r) f_k^{'}(r) dr \right) \leq \lambda_{\max} \left(\int_0^1 f_k(r) f_k^{'}(r) dr \right) \leq \overline{M},$$

for some constant $\underline{M} < \infty$ and $\overline{M} < \infty$. Define $s_{\max} := \sup_{1 \leq i \leq n} \sup_{0 \leq r \leq 1} |K_{c,i}(r)|$. By Cauchy-Schwarz inequality, we can derive that

$$\lambda_{\max} \left(\int_0^1 \left(f_k(r) f_k^{'}(r) \right) \otimes \left(K_c(r) K_c^{'}(r) \right) dr \right) \leq s_{\max}^2 \overline{M},\ \forall k \geq 1.$$

To show that $M_k^{-1} = O_p(1)$, it is sufficient to show $M_k > 0$ asymptotically. For any $\delta > 0$, we define $K_c^{\delta}(\cdot) := \left(K_{c,1}^{\delta}(\cdot), K_{c,2}^{\delta}(\cdot), ..., K_{c,n}^{\delta}(\cdot)\right)$ and $K_{c,j}^{\delta}(r) := K_{c,j}(r)\mathbf{1}_{\{|K_{c,j}(r)|>\delta\}} + \delta\mathbf{1}_{\{|K_{c,j}(r)|\leq\delta\}}$. Thus it is shown that

$$\lambda_{\min}\left(\int_0^1 \left(f_k(r)f_k^{'}(r)\right) \otimes \left(K_c^{\delta}(r)K_c^{\delta'}(r)\right) dr\right) \geq \delta^2\underline{M}.$$

Define $J_\delta = \{r \in [0,1] \mid \inf_{1\leq j\leq n}|K_{c,j}(r)| \leq \delta\}$. Based on the definition of J_δ, we can derive the following argument as,

$$\begin{aligned}
&\int_0^1 \left(f_k(r)f_k^{'}(r)\right) \otimes \left(K_c^{\delta}(r)K_c^{\delta'}(r)\right) dr - \int_0^1 \left(f_k(r)f_k^{'}(r)\right) \otimes \left(K_c(r)K_c^{'}(r)\right) dr \\
\leq\ & 2\int_{J_\delta} \left(f_k(r)f_k^{'}(r)\right) \otimes \left(K_c(r)K_c^{'}(r)\right) dr.
\end{aligned}$$

Additionally, for each $i \in \{1,2,...,n\}$, we have $\sup_{0\leq r\leq 1}|c_i(r)| < \bar{c}$ and the differentiability of $c_i(\cdot)$. Based on Lou and Ouyang (2017), the occupation formula holds with

$$\liminf_{\epsilon\downarrow 0}\frac{1}{\delta}\int_{I\subseteq[0,1]} \Pr\left\{\|K_c(s) - K_c(u)\| \leq \delta\right\} ds < \infty \ a.s.$$

for any $u \in I$. Besides, $\forall i = 1,2,...,n$, the local time $L_i(1,0)$ exists almost surely due to Lou and Ouyang (2017). Therefore, we have

$$\liminf_{\delta\downarrow 0}\frac{\Sigma_i}{2\delta}\int_0^1 \mathbf{1}_{\{|K_{c,i}(r)|\leq\delta\}}dr = L_i(1,0) \ \ a.s., \tag{B.1}$$

where Σ_i is the variance of driven Brownian motion in the ith entry. Based on (B.1), the following two equations hold:

$$\frac{1}{\delta}\int_0^1 \mathbf{1}_{(J_\delta)} = O(1),$$

and

$$\left\|\int_{J_\delta}\left(f_k(r)f_k^{'}(r)\right) dr\right\| = O_p(\delta).$$

Therefore for any $\delta > 0$, we have

$$\begin{aligned} & \lambda_{\min}\left(\int_0^1 \left(f_k(r)f_k^{'}(r)\right) \otimes \left(K_c(r)K_c^{'}(r)\right) dr\right) \\ = & \lambda_{\min}\left(\int_0^1 \left(f_k(r)f_k^{'}(r)\right) \otimes \left(K_c^{\delta}(r)K_c^{\delta'}(r)\right) dr\right) + O_p\left(\delta^3\right) \\ \geq & \delta^2 \left(1 + o_p(1)\right) > 0. \end{aligned}$$

These arguments show that M_k is bounded from zero. This shows $M_k^{-1} = O_p(1)$.

If we impose norm operators, it is trivial to have

$$\|M_k\| = O_p(1), \ \ \left\|M_k^{-1}\right\| = O_p(1).$$

To justify that $\|M_{Tk} - M_k\| = O_p(kT^{-\alpha})$, we apply the Skorohod embedding theorem and obtain

$$\int_0^1 \phi_i\left(\frac{[Tr]}{T}\right)\phi_j\left(\frac{[Tr]}{T}\right)K_{T,c}^*(r)K_{T,c}^{*'}(r)dr = \int_0^1 \phi_i(r)\phi_j(r)K_c(r)K_c^{'}(r)dr + O_p\left(T^{-\alpha}\right),$$

uniformly in $i, j \in \{1, 2, ..., k\}$. The proof for (a) is complete.

For (b), notice that

$$\left\|\int_0^1 \phi_i\left(\frac{[Tr]}{T}\right)\phi_j\left(\frac{[Tr]}{T}\right)K_{Tc}^*(r)\left\langle dW_{xT}(r), dW_{xT}^{'}(r)\right\rangle K_{Tc}^{*'}(r)\right\| = O_p\left(k\right),$$

uniformly in $i, j \in \{1, 2, ..., k\}$. Then we have

$$\left\|\int_0^1 \phi_i\left(\frac{[Tr]}{T}\right)K_{Tc}^*(r)dW_{xT}^{`}(r)\right\| = O_p\left(\sqrt{k}\right),$$

uniformly in $i \in \{1, 2, ..., k\}$. For the asymptotic distance between N_{Tk} and N_k, it suffices to establish that

$$\left\|\int_0^1 \phi_i\left(\frac{[Tr]}{T}\right)K_{Tc}^*(r)dW_{xT}'(r) - \int_0^1 \phi_i\left(r\right)K_c(r)dW_x'(r)\right\| = O_p\left(\sqrt{k}T^{-\alpha}\right),$$

uniformly in $i \geq 1$. It directly follows

$$\begin{aligned} & \left\| \int_0^1 \phi_i \left(\frac{[Tr]}{T} \right) K_{Tc}^*(r) dW_{xT}'(r) - \int_0^1 \phi_i(r) K_c(r) dW_x'(r) \right\| \\ \leq & \underbrace{\left\| \int_0^1 \left(\phi_i \left(\frac{[Tr]}{T} \right) K_{Tc}^*(r) - \phi_i(r) K_c(r) \right) dW_{xT}'(r) \right\|}_{(\mathbf{I.1})} \\ & + \underbrace{\left\| \int_0^1 \phi_i(r) K_c(r) \left(dW_x'(r) - dW_{xT}'(r) \right) \right\|}_{(\mathbf{I.2})}. \end{aligned}$$

Following Lemma 3.1, we have $(\mathbf{I.1}) = O_p\left(\sqrt{k} T^{-\alpha}\right)$. Following Lemma 2.1, for $(\mathbf{I.2})$, we have

$$\begin{aligned} & \left\| \int_0^1 \phi_i(r) K_c(r) dW_{xT}'(r) - \int_0^1 \phi_i(r) K_c(r) dW_x'(r) \right\| \\ = & \left\| \phi_i(1) K_c(1) \left(W_{xT}'(1) - W_x'(1) \right) \right\| + \left\| \int_0^1 \phi_i(r) \left(W_x(r) - W_{xT}(r) \right) dK_c'(r) \right\| \\ & + \left\| \int_0^1 \left(W_{xT}'(r) - W_x'(r) \right) K_c(r) \, d\phi_i(r) \right\| \\ = & O_p\left(\sqrt{k} T^{-\alpha} \right), \end{aligned}$$

uniformly in $i \in \{1, 2, ..., k\}$.

For (c), we follow the previous notations as $s_{\max} = \sup_{1 \leq j \leq n} \sup_{0 \leq r \leq 1} |K_{c,i}(r)|$. We have the following argument,

$$\begin{aligned} & \left\| \int_0^1 \phi_i(r) K_c(r) K_c'(r) \left(R_T(r) - R_T^{(k)}(r) \right) dr \right\| \\ \leq & O_p\left(\sqrt{k} \right) \left(\int_0^1 \left\| R_T(r) - R_T^{(k)}(r) \right\|^2 dr \right)^{\frac{1}{2}} \\ = & O_p\left(k^{\frac{1}{2} - q} \right), \end{aligned}$$

uniformly in $i \in \{1, 2, ..., k\}$. ■

Remark: The proof of Lemma B.2 illustrates the necessity of smoothness on basis function since the integration-by-part formula is applied.

Lemma B.3. *(Stochastic order of estimation bias) Let Assumptions 2.1 and 2.2 hold. As* $T \to \infty$*, we have*

(a)

$$\prod \widehat{R}_T^{(k)} = \prod R_T^{(k)} + O_p(T^{-1}k);$$

(b)

$$A_{Tk}^{-\frac{1}{2}} \left(\prod \widehat{R}_T^{(k)} - \prod R_T^{(k)}\right) = \left(S_k M_k^{-1} S_k^{'}\right)^{-\frac{1}{2}} S_k M_k^{-1} N_k + o_p(1).$$

The Proof of Lemma B.3: This paper follows the proof of Park and Hahn (1999). Define $M_{Tk}^o = \frac{Y_{Tk,-1}Y_{Tk,-1}^{'}}{T^2}$ and $N_{Tk}^o = \frac{Y_{Tk,-1}^{'}U_{xk}}{T}$. We have

$$T\left(\prod\left(\widehat{R}_T^{(k)}\right) - \prod\left(R_T^{(k)}\right)\right) = TS_k\left(\widehat{\beta}_{Tk} - \beta_k\right) = S_k\left(M_{Tk}^o\right)^{-1} N_{Tk}^o. \quad \text{(B.2)}$$

For (a), it is enough to justify that $S_k\left(M_{Tk}^o\right)^{-1} N_{Tk}^o = S_k M_k^{-1} N_k + O_p\left(T^{-\alpha}k^2\right) + O_p\left(Tk^{1-q}\right)$. It follows

$$\begin{aligned}\left\|\left(M_{Tk}^o\right)^{-1} - \left(M_k\right)^{-1}\right\| &\leq \left\|\left(M_{Tk}^o\right)^{-1}\right\| \left\|\left(M_k\right)^{-1}\right\| \left\|\left(M_{Tk}^o - M_k\right)\right\| \\ &= O_p(1)\, O_p(1)\, O_p\left(T^{-\alpha}k\right) = O_p\left(T^{-\alpha}k\right), \quad \text{(B.3)}\end{aligned}$$

and

$$\left\|M_{Tk}^o - M_k\right\| \leq \left\|M_{Tk}^o - M_{Tk}\right\| + \left\|M_{Tk} - M_k\right\| = O_p\left(T^{-\alpha}k\right),$$

as illustrated in Lemma B.1 and B.2. Since $\left\|\left(M_k\right)^{-1}\right\| = O_p(1)$ and $\left\|M_{Tk} - M_k\right\| = o_p(1)$, we have $\left\|\left(M_{Tk}\right)^{-1}\right\| = O_p(1)$. Similarly, for the numerator, we have

$$\left\|N_{Tk}^o - N_k\right\| \leq \left\|N_{Tk}^o - N_{Tk}\right\| + \left\|N_{Tk} - N_k\right\| = O_p\left(\sqrt{k}T^{-\alpha}\right) + O_p\left(k^{\frac{1}{2}-q}T\right). \quad \text{(B.4)}$$

Combining (B.3) and (B.4), we have

$$\begin{aligned}&\left\|S_k\left(M_{Tk}^o\right)^{-1} N_{Tk}^o - S_k M_k^{-1} N_k\right\| \\ =\ &\left\|S_k\left(M_{Tk}^o\right)^{-1} N_{Tk}^o - S_k M_k^{-1} N_k^o + S_k M_k^{-1} N_k^o - S_k M_k^{-1} N_k\right\| \\ \leq\ &\left\|S_k\right\| \left(\left\|\left(M_{Tk}^o\right)^{-1} - \left(M_k\right)^{-1}\right\| \left\|N_{Tk}^o\right\| + \left\|\left(M_k\right)^{-1}\right\| \left\|N_{Tk}^o - N_{Tk}\right\|\right) \\ =\ &O_p\left(\sqrt{k}\right)\left(O_p\left(T^{-\alpha}k\right) O_p\left(\sqrt{k}\right) + O_p(1)\, O_p\left(T^{-\alpha}k^{-1}\right) + O_p(1)\, O_p\left(k^{\frac{1}{2}-q}T\right)\right) \\ =\ &O_p\left(T^{-\alpha}k^2\right) + O_p\left(Tk^{1-q}\right). \quad \text{(B.5)}\end{aligned}$$

For (b), at first, we show that

$$\left(T^2 A_{Tk}\right)^{-\frac{1}{2}} = \left(S_k \left(M_{Tk}^o\right)^{-1} S_k^{'}\right)^{-\frac{1}{2}} = \left(S_k \left(M_k\right)^{-1} S_k^{'}\right)^{-\frac{1}{2}} + O_p\left(T^{-\alpha} k^{\frac{1}{2}}\right). \tag{B.6}$$

This argument is based on the following facts

$$\lambda_{\min}\left(S_k \left(M_{Tk}^o\right)^{-1} S_k^{'}\right) \geq \frac{\lambda_{\min}\left(S_k S_k^{'}\right)}{\lambda_{\max}\left(M_{Tk}^o\right)} \geq \frac{k\underline{M}}{\lambda_{\max}\left(M_{Tk}^o\right)},$$

and

$$\lambda_{\min}\left(S_k \left(M_k\right)^{-1} S_k^{'}\right) \geq \frac{\lambda_{\min}\left(S_k S_k^{'}\right)}{\lambda_{\max}\left(M_k\right)} \geq \frac{k\underline{M}}{\lambda_{\max}\left(M_k\right)}.$$

Therefore we have

$$\left(S_k \left(M_{Tk}^o\right)^{-1} S_k^{'}\right)^{-\frac{1}{2}} = O_p\left(k^{-\frac{1}{2}}\right),$$

and

$$\left(S_k \left(M_k\right)^{-1} S_k^{'}\right)^{-\frac{1}{2}} = O_p\left(k^{-\frac{1}{2}}\right).$$

It is shown that

$$\begin{aligned} \left\|S_k \left(\left(M_{Tk}^o\right)^{-1} - \left(M_k\right)^{-1}\right) S_k^{'}\right\| &\leq \left\|S_k\right\|^2 \left\|\left(M_{Tk}^o\right)^{-1} - \left(M_k\right)^{-1}\right\| \\ &= O_p(k)\, O_p\left(T^{-\alpha} k\right) = O_p\left(T^{-\alpha} k^2\right). \end{aligned}$$

Therefore, it is easily derived that

$$\begin{aligned} &\left(S_k \left(M_{Tk}^o\right)^{-1} S_k^{'}\right)^{-\frac{1}{2}} - \left(S_k \left(M_k\right)^{-1} S_k^{'}\right)^{-\frac{1}{2}} \\ =\ & \left(S_k \left(M_{Tk}^o\right)^{-1} S_k^{'}\right)^{-\frac{1}{2}} \left(S_k \begin{pmatrix} \left(M_k\right)^{-1} \\ -\left(M_{Tk}^o\right)^{-1} \end{pmatrix} S_k^{'}\right) \begin{pmatrix} \left(S_k \left(M_k\right)^{-1} S_k^{'}\right)^{\frac{1}{2}} \\ +\left(S_k \left(M_{Tk}^o\right)^{-1} S_k^{'}\right)^{\frac{1}{2}} \end{pmatrix}^{-1} \cdot \\ & \left(S_k \left(M_k\right)^{-1} S_k^{'}\right)^{-\frac{1}{2}} \\ =\ & O_p\left(T^{-\alpha} k^{\frac{1}{2}}\right). \end{aligned}$$

Based on (B.2) and (B.6), we have

$$S_k \left(M_k\right)^{-1} S_k^{'} = O_p(k), \quad \left(S_k \left(M_k\right)^{-1} S_k^{'}\right)^{-\frac{1}{2}} = O_p(k^{-\frac{1}{2}}).$$

To ensure (b), we need the following asymptotic negligibility conditions,

$$T^{-1}k = o\left(1\right),\ Tk^{\frac{1}{2}-q} = o\left(1\right),\ T^{-\alpha}\left(\sqrt{k}\right)^{3} = o\left(1\right),\ T^{-\alpha}k^{\frac{1}{2}} = o\left(1\right).$$

These are guaranteed by Assumption 2.2 (ii). Then (b) follows directly (B.6), (B.5) and rate restrictions in Assumption 2.2 (ii). ■

Lemma B.4. *(Stochastic order of approximation error) Let Assumptions 2.1 and 2.2 hold. As $T \to \infty$, we have*

(a)

$$\prod R_T^{(k)} = \prod R_T + O_p(k^{-q});$$

(b)

$$A_{Tk}^{-\frac{1}{2}}\left(\prod R_T^{(k)} - \prod R_T\right) = O_p(Tk^{-q-1}) = o_p(1).$$

The Proof of Lemma B.4: (a) can be derived easily since the approximation error is diminishing; For (b), note that

$$A_{Tk}^{-\frac{1}{2}} = O_p\left(\frac{T}{\sqrt{k}}\right),$$

where $A_{Tk} := S_k\left(Y_{Tk,-1}'Y_{Tk,-1}\right)^{-1}S_k'$. It is essential for the derivations. By Assumption 2.2 (ii), we have $O_p\left(Tk^{-q-1}\right) = o_p(1)$. The proof for Lemma B.4 is complete. ■

The Proof of Theorem 2.1: As illustrated in Lemma B.3 and B.4, the estimation bias has a higher order than approximation error. With triangle inequality, we prove that

$$\begin{aligned}\left\|\prod \widehat{R}_T^{(k)} - \prod R_T\right\| &\le \left\|\prod \widehat{R}_T^{(k)} - \prod R_T^{(k)}\right\| + \left\|\prod R_T^{(k)} - \prod \left(R_T\right)\right\| \\ &= O_p\left(kT^{-1}\right) + O_p\left(k^{-q}\right) \\ &= o_p(1).\end{aligned}$$

Besides,

$$\begin{aligned} A_{Tk}^{-\frac{1}{2}}\left\|\prod \widehat{R}_T^{(k)}-\prod R_T\right\| &\leq A_{Tk}^{-1}\left\|\prod \widehat{R}_T^{(k)}-\prod R_T^{(k)}\right\|+A_{Tk}^{-1}\left\|\prod R_T^{(k)}-\prod (R_T)\right\| \\ &= O_p(1)+O_p\left(Tk^{-q-1}\right), \end{aligned}$$

where the Assumption 2.2 ensures $O_p\left(Tk^{-q-1}\right)=o_p(1)$. Therefore, we have the following approximation for (b) as

$$A_{Tk}^{-\frac{1}{2}}\left(\prod \widehat{R}_T^{(k)}-\prod R_T\right)=\left(S_k M_k^{-1} S_k^{'}\right)^{-\frac{1}{2}} S_k M_k^{-1} N_k+o_p(1). \qquad \text{(B.7)}$$

■

The Proof of Theorem 2.2: Based on the approximation (B.7), we have

$$A_{Tk}^{-\frac{1}{2}}\left(\prod \widehat{R}_T^{(k)}-\prod R_T\right)=\left(S_k M_k^{-1} S_k^{'}\right)^{-\frac{1}{2}} S_k M_k^{-1} N_k+o_p(1).$$

If Assumptions 2.1 and 2.2 hold, as $T\to\infty$, we have

$$A_{Tk}^{-\frac{1}{2}}\left(\prod \widehat{R}_T^{(k)}-\prod R_T\right)\Rightarrow \mathbf{Z},$$

where

$$\begin{aligned} \mathbf{Z} \ &: = \lim_{k\to\infty}\left[S_k\left(\int_0^1\left(f_k(r)f_k^{'}(r)\right)\otimes\left(K_c(r)K_c^{'}(r)\right)dr\right)^{-1} S_k\right]^{-\frac{1}{2}} \\ &\cdot S_k\left(\int_0^1\left(f_k(r)f_k^{'}(r)\right)\otimes\left(K_c(r)K_c^{'}(r)\right)dr\right)^{-1}\left(\int_0^1 f_k(r)\otimes K_c(r)dW_x^{'}(r)\right). \end{aligned}$$

■

Panel estimation and test

The Proof of Theorem 2.3: Let Assumptions 2.1-2.3 hold. According to the definition of panel sieve estimator, we have

$$
\begin{aligned}
& A_{MTk}^{-\frac{1}{2}}\left(\prod \widehat{R}_T^{(k)} - \prod R_T\right) \\
= & \left[S_k\left(\frac{1}{M}\sum_{m=1}^{M} Y'_{mTk,-1}Y_{mTk,-1}\right)^{-1} S'_k\right]^{-\frac{1}{2}} \left[S_k\left(\frac{1}{M}\sum_{m=1}^{M} Y'_{mTk,-1}Y_{mTk,-1}\right)^{-1}\right] \\
& \cdot \left[\frac{1}{\sqrt{M}}\sum_{m=1}^{M} Y'_{mTk,-1}U_{mxk}\right] \\
= & \left[S_k\left(\frac{1}{MT^2}\sum_{m=1}^{M} Y'_{mTk,-1}Y_{mTk,-1}\right)^{-1} S'_k\right]^{-\frac{1}{2}} \left[S_k\left(\frac{1}{MT^2}\sum_{m=1}^{M} Y'_{mTk,-1}Y_{mTk,-1}\right)^{-1}\right] \\
& \cdot \left[\frac{1}{\sqrt{M}T}\sum_{m=1}^{M} Y'_{mTk,-1}U_{mxk}\right] + o_p(1).
\end{aligned}
$$

Rate restrictions in Assumption 2.1-2.3 make sure that approximation error is dominated by estimation error. When Assumption 2.1-2.3 hold, as $(M,T)_{seq} \to \infty$, we have

$$
\begin{aligned}
& A_{MTk}^{-\frac{1}{2}}\left(\prod \widehat{R}_T^{(k)} - \prod R_T\right) \\
\Rightarrow & \left[S_k\left(\frac{1}{M}\sum_{m=1}^{M}\int_0^1 \left(f_k(r)\, f'_k(r)\right) \otimes \left(K_{mc}^2(r)\right) dr\right)^{-1} S'_k\right]^{-\frac{1}{2}} \cdot \\
& \left[S_k\left(\frac{1}{M}\sum_{m=1}^{M}\int_0^1 \left(f_k(r)\, f'_k(r)\right) \otimes \left(K_{mc}^2(r)\right) dr\right)^{-1}\right] \cdot \\
& \frac{1}{\sqrt{M}}\sum_{m=1}^{M}\int_0^1 f_k(r) \otimes K_{mc}(r)\, dW'_x(r) \\
\stackrel{d}{\to} & N(0, I_d).
\end{aligned}
$$

For large T asymptotics, we apply the invariance principle. For large M asymptotics, we apply the Lindeberg-Lévy central limit theorem. To justify

$$
A_{MTk}^{-\frac{1}{2}}\left(\prod\left(\widehat{R}_T\right) - \prod R_T\right) \stackrel{d}{\to} N(0, I_d), \text{ as } (M,T) \to \infty,
$$

we are supposed to verify the following arguments. If Assumptions 2.1-2.3 hold, as $(M,T) \to \infty$, we need to show that

(1)

$$\frac{S_k}{MT^2}\left(\sum_{m=1}^{M} Y'_{mTk,-1} Y_{mTk,-1}\right)^{-1} S'_k$$
$$\xrightarrow{p} S_k \left(\mathbb{E}\lim_{k\to\infty}\int_0^1 \left(f_k(r) f'_k(r)\right) \otimes \left(K^2_{mc}(r)\right) dr\right)^{-1} S'_k;$$

(2)

$$\frac{S_k}{MT^2}\left(\sum_{m=1}^{M} Y'_{mTk,-1} Y_{mTk,-1}\right)^{-1} \xrightarrow{p} S_k \left(\mathbb{E}\lim_{k\to\infty}\int_0^1 \left(f_k(r) f'_k(r)\right) \otimes \left(K^2_{mc}(r)\right) dr\right)^{-1};$$

(3)

$$\frac{1}{\sqrt{M}T}\left(\sum_{m=1}^{M} Y'_{mTk,-1} U_{mxk}\right) \xrightarrow{d} N\left(0, \Sigma \cdot \mathbb{E}\lim_{k\to\infty}\int_0^1 \left(f_k(r) f'_k(r)\right) \otimes \left(K^2_{mc}(r)\right) dr\right).$$

For (1), we need to show the sequential limit as $(M,T)_{seq} \to \infty$ and verify the sufficient conditions in Theorem 1 of Phillips and Moon (1999, PM hereafter). Obviously,

$$\begin{aligned} & S_k\left(\frac{1}{MT^2}\sum_{m=1}^{M} Y'_{mTk,-1} Y_{mTk,-1}\right)^{-1} S'_k \\ \Rightarrow & \left(\frac{1}{M}\sum_{m=1}^{M}\int_0^1 \left(f_k(r) f'_k(r)\right) \otimes \left(K^2_{mc}(r)\right) dr\right)^{-1} S'_k \\ \xrightarrow{p} & S_k\left(\mathbb{E}\int_0^1 \left(f_k(r) f'_k(r)\right) \otimes \left(K^2_{mc}(r)\right) dr\right)^{-1} S'_k, \end{aligned}$$

as $(M,T)_{seq} \to \infty$. For Theorem 1 (i) in PM, we show it by

$$\begin{aligned} & \limsup_{M,T} \frac{1}{MT^2}\sum_{m=1}^{M} \mathbb{E}\left(Y'_{mTk,-1} Y_{mTk,-1}\right) \\ = & \limsup_{M,T} \frac{1}{M}\sum_{m=1}^{M}\int_0^1 \left(f_k(r) f'_k(r)\right) \otimes \mathbb{E}\left(K^*_{mT,c}(r)\right)^2 dr = O_p(1), \end{aligned}$$

where $t = [Tr]$ and $K^*_{mT,c}(r)$ is the m-th entry of term $K^*_{T,c}(r)$. For Theorem

1 (ii) in PM, we have

$$
\begin{aligned}
&\limsup_{M,T}\frac{1}{M}\sum_{m=1}^{M}\left\|\frac{\mathbb{E}\left(Y_{mTk,-1}^{\prime}Y_{mTk,-1}\right)}{T^2}-\int_0^1\left(f_k\left(r\right)f_k^{\prime}\left(r\right)\right)\otimes\mathbb{E}\left(K_{m,c}\left(r\right)\right)^2dr\right\| \\
=\;&\limsup_{M,T}\frac{1}{M}\sum_{m=1}^{M}\left\|\begin{array}{c}\int_0^1\left(f_k\left(\frac{[Tr]}{T}\right)f_k^{\prime}\left(\frac{[Tr]}{T}\right)\right)\otimes\mathbb{E}\left(K_{mT,c}^{*}\left(r\right)\right)^2dr \\ -\int_0^1\left(f_k\left(r\right)f_k^{\prime}\left(r\right)\right)\otimes\mathbb{E}\left(K_{m,c}\left(r\right)\right)^2dr\end{array}\right\| \\
\leq\;&\limsup_{M,T}\frac{1}{M}\sum_{m=1}^{M}\left\|\int_0^1\left(f_k\left(r\right)f_k^{\prime}\left(r\right)\right)\otimes\left[\mathbb{E}\left(K_{mT,c}^{*}\left(r\right)\right)^2-\mathbb{E}\left(K_{m,c}\left(r\right)\right)^2\right]dr\right\| \\
&+\limsup_{M,T}\frac{1}{M}\sum_{m=1}^{M}\left\|\int_0^1\left(f_k\left(\frac{[Tr]}{T}\right)f_k^{\prime}\left(\frac{[Tr]}{T}\right)-\left(f_k\left(r\right)f_k^{\prime}\left(r\right)\right)\right)\otimes\mathbb{E}\left(K_{mT,c}\left(r\right)\right)^2dr\right\| \\
=\;&O_p\left(T^{-\alpha}\right)+O_p\left(kT^{-1}\right)
\end{aligned}
$$

for any $k>1$. For Theorem 1 (iii) in PM, for any $\eta>0$, we have

$$
\begin{aligned}
&\limsup_{M,T}\frac{1}{M}\sum_{m=1}^{M}\left\|\int_0^1\left(f_k\left(\frac{[Tr]}{T}\right)f_k^{\prime}\left(\frac{[Tr]}{T}\right)\right)\otimes\mathbb{E}\left(K_{mT,c}^{*}\left(r\right)\right)^2dr\right\| \\
&\cdot\mathbf{1}\left\{\left\|\int_0^1\left(f_k\left(\frac{[Tr]}{T}\right)f_k^{\prime}\left(\frac{[Tr]}{T}\right)\right)\otimes\mathbb{E}\left(K_{mT,c}^{*}\left(r\right)\right)^2dr\right\|>M\eta\right\} \\
\leq\;&\limsup_{M,T}B_1\sup_{0\leq r\leq 1}\left\|\mathbb{E}\left(K_{mT,c}^{*}\left(r\right)\right)^2dr\right\|\mathbf{1}\left\{B_1\sup_{0\leq r\leq 1}\left\|\mathbb{E}\left(K_{mT,c}^{*}\left(r\right)\right)^2dr\right\|>M\eta\right\} \\
=\;&o_p\left(1\right),
\end{aligned}
$$

where B_1 is some finite positive constant value and $K_{mT,c}^{*}\left(r\right)$ is stochastically bounded for each $r\in[0,1]$. Similarly, for Theorem 1 (iv) in PM, we have

$$
\begin{aligned}
&\limsup_{M,T}\frac{1}{M}\sum_{m=1}^{M}\left\|\int_0^1\left(f_k\left(r\right)f_k^{\prime}\left(r\right)\right)\otimes\mathbb{E}\left(K_{m,c}\left(r\right)\right)^2dr\right\| \\
&\cdot\mathbf{1}\left\{\left\|\int_0^1\left(f_k\left(r\right)f_k^{\prime}\left(r\right)\right)\otimes\mathbb{E}\left(K_{m,c}\left(r\right)\right)^2dr\right\|>M\eta\right\}=o_p\left(1\right),
\end{aligned}
$$

for any $k>1$. Combining (i) (ii) (iii) and (iv) of Theorem 1, we have

$$
\frac{S_k}{MT^2}\left(\sum_{m=1}^{M}Y_{mTk,-1}^{\prime}Y_{mTk,-1}\right)^{-1}S_k^{\prime}\xrightarrow{p}S_k\left(\mathbb{E}\int_0^1\left(f_k\left(r\right)f_k^{\prime}\left(r\right)\right)\otimes\left(K_{mc}^2\left(r\right)\right)dr\right)^{-1}S_k^{\prime},
$$

under joint convergence framework. Since the difference between $\frac{S_k}{MT^2}\left(\sum_{m=1}^{M}Y_{mTk,-1}^{\prime}Y_{mTk,-1}\right)^{-1}$ and $\frac{S_k}{MT^2}\left(\sum_{m=1}^{M}Y_{mTk,-1}^{\prime}Y_{mTk,-1}\right)^{-1}$ is a non-random value, the joint convergence of $\frac{S_k}{MT^2}\left(\sum_{m=1}^{M}Y_{mTk,-1}^{\prime}Y_{mTk,-1}\right)^{-1}$ can be demonstrated in a similar way as $\frac{S_k}{MT^2}\left(\sum_{m=1}^{M} Y_{mTk,-1}^{'} Y_{mTk,-1}\right)^{-1} S_k^{'}$.

For (3), under the convergence framework, we are supposed to justify

$$\frac{1}{\sqrt{M}T}\left[S_k\left(\mathbb{E}\int_0^1\left(f_k(r)\,f_k^{'}(r)\right)\otimes\left(K_{mc}^2(r)\right)dr\right)^{-1}S_k^{'}\right]^{-\frac{1}{2}}$$
$$\cdot S_k\left(\mathbb{E}\int_0^1\left(f_k(r)\,f_k^{'}(r)\right)\otimes\left(K_{mc}^2(r)\right)dr\right)^{-1}\left(\sum_{m=1}^{M} Y_{mTk,-1}^{'}U_{mxk}\right)\xrightarrow{d} N(0,I_d),$$

under Assumptions 2.1-2.3. Following the Theorem 2 of PM, we firstly show the sequential limit and then justify uniform integrability in Theorem 2 of PM. In order to simplify the notation, we define $\Gamma_k := \mathbb{E}\lim_{k\to\infty}\int_0^1\left(f_k(r)\,f_k^{'}(r)\right)\otimes\left(K_{mc}^2(r)\right)dr$. If Assumptions 2.1-2.3 hold, under sequential asymptotics $(M,T)_{seq}\to\infty$, we have

$$\frac{1}{\sqrt{M}T}\left[S_k\Gamma_k^{-1}S_k^{'}\right]^{-\frac{1}{2}}S_k\Gamma_k^{-1}\left(\sum_{m=1}^{M} Y_{mTk,-1}^{'}U_{mxk}\right)\xrightarrow{d} N(0,I_d). \tag{B.8}$$

Equation (B.8) holds due to the invariance principle and Lindeberg-Lévy central limit theorem. For any $\eta>0$, we have

$$\begin{aligned}
&\frac{1}{MT^2}\left\|\begin{array}{c}\left[S_k\Gamma_k^{-1}S_k^{'}\right]^{-\frac{1}{2}}S_k\Gamma_k^{-1}\mathbb{E}\left[\left(\sum_{m=1}^{M} Y_{mTk,-1}^{'}U_{mxk}\right)\left(\sum_{m=1}^{M} Y_{mTk,-1}^{'}U_{mxk}\right)^{'}\right]\\ \cdot\left\{\left[S_k\Gamma_k^{-1}S_k^{'}\right]^{-\frac{1}{2}}S_k\Gamma_k^{-1}\right\}^{'}\end{array}\right\|\\
&\cdot\mathbf{1}\left\{\left\|\left[S_k\Gamma_k^{-1}S_k^{'}\right]^{-\frac{1}{2}}S_k\Gamma_k^{-1}\left(Y_{mTk,-1}^{'}U_{mxk}\right)\right\|>\sqrt{M}T\eta\right\}\\
\leq\;&\frac{1}{MT^2}B_2\sum_{m=1}^{M}\mathbb{E}\left\|Y_{mT,-1}^{'}Y_{mT,-1}\right\|\mathbf{1}\left\{\left\|B_2\left(\frac{Y_{mT,-1}^{'}}{\sqrt{T}}\frac{Y_{mT,-1}}{\sqrt{T}}\right)\right\|>MT\eta\right\}\\
=\;&o_p(1),
\end{aligned}$$

for some positive constant $B_2<\infty$, where $Y_{mT}:=(y_{m1},y_{m2},...,y_{mT})^{'}$ and $Y_{mT,-1}:=(y_{m0},y_{m0},...,y_{m0})^{'}$. The inequality is based on the invariance principle in BM and the continuous mapping theorem as $(M,T)\to\infty$. Therefore, the joint convergence coincides with sequential asymptotics. We conclude the proof.

■

The Proof of Theorem 2.4: We need to justify the following three results:

(1) the consistency of the variance estimator $\widehat{\Sigma}$;

(2) the stochastic boundness of proposed test statistics under the null hypothesis;

(3) the asymptotic divergence of the test statistics under the alternative hypothesis.

First, we collect sieve regression residuals $\widehat{u}_{mt} = y_{mt} - \widehat{R}_{Tt}^{(k)} y_{m,t-1}$ for $t = 1, 2, ..., T$ and $m = 1, 2, ..., M$. The sieve estimator $\widehat{R}_{Tt}^{(k)}$ is based on orthogonal trigonometric basis. Under joint asymptotics, we have

$$
\begin{aligned}
\widehat{\Sigma} &= \frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} \widehat{u}_{mt}^2 \\
&= \frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} \left(\left(R_{Tt} - \widehat{R}_{Tt}^{(k)} \right) y_{m,t-1} + u_{mt} \right)^2 \\
&= \frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} (u_{mt})^2 + \frac{1}{MT} \Big(R_{Tt} - \widehat{R}_{Tt}^{(k)} \Big)^2 \sum_{m=1}^{M} \sum_{t=1}^{T} (y_{m,t-1})^2 + \\
&\quad \frac{2\Big(R_{Tt} - \widehat{R}_{Tt}^{(k)} \Big)}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} y_{m,t-1} u_{mt} \\
&= \frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} (u_{mt})^2 + O_p \left(\frac{k^2}{T} \right) + O_p \left(\frac{k}{\sqrt{M} T} \right) \\
&= \frac{1}{MT} \sum_{m=1}^{M} \sum_{t=1}^{T} (u_{mt})^2 + o_p (1) \xrightarrow{p} \Sigma.
\end{aligned}
$$

Second, with the time-invariant term $\prod R_T$ as the true value under H_0, we show that

$$
\begin{aligned}
& \prod \widehat{R}_T^{(k)} - \prod \left(\widetilde{R}_T \right) \\
= \ & \left(\prod \widehat{R}_T^{(k)} - \prod R_T \right) - \left(\prod \left(\widetilde{R}_T \right) - \prod R_T \right) \\
= \ & O_p \left(\frac{k}{\sqrt{M} T} \right) + O_p \left(\frac{1}{\sqrt{M} T} \right).
\end{aligned}
$$

Then we have

$$
\begin{aligned}
& A_{MTk}^{-\frac{1}{2}} \prod \widehat{R}_T^{(k)} - \prod \left(\widetilde{R}_T\right) \\
= \; & A_{MTk}^{-\frac{1}{2}} \left(\prod \widehat{R}_T^{(k)} - \prod R_T\right) - A_{MTk}^{-\frac{1}{2}} \left(\prod \left(\widetilde{R}_T\right) - \prod R_T\right) \\
= \; & O_p(1) + O_p\left(\frac{1}{k}\right).
\end{aligned}
$$

Under the alternative hypothesis, we have

$$
\begin{aligned}
& \prod \widehat{R}_T^{(k)} - \prod \left(\widetilde{R}_T\right) \\
= \; & \left(\prod \widehat{R}_T^{(k)} - \prod R_T\right) - \left(\prod \left(\widetilde{R}_T\right) - \prod R_T\right) \\
= \; & \left(\prod \widehat{R}_T^{(k)} - \prod R_T\right) - \left(\prod \left(\widetilde{R}_T\right) - \prod R_T\right) - \left(\prod R_T - \prod (R_T)\right) \\
= \; & O_p\left(\frac{k}{\sqrt{MT}}\right) + O_p\left(\frac{1}{\sqrt{MT}}\right) + O_p\left(\frac{1}{T}\right),
\end{aligned}
$$

where $\prod R_T$ is the pseudo-true value for the probability limit of pooled least square estimator $\prod \left(\widetilde{R}_T\right)$. Then,

$$
\begin{aligned}
& A_{MTk}^{-\frac{1}{2}} \prod \widehat{R}_T^{(k)} - \prod \left(\widetilde{R}_T\right) \\
= \; & A_{MTk}^{-\frac{1}{2}} \left(\prod \widehat{R}_T^{(k)} - \prod R_T\right) - A_{MTk}^{-\frac{1}{2}} \left(\prod \left(\widetilde{R}_T\right) - \prod R_T\right) - A_{MTk}^{-\frac{1}{2}} \left(\prod (R_T^*) - \prod (R_T)\right) \\
= \; & O_p(1) + O_p\left(\frac{1}{k}\right) + O_p\left(\frac{\sqrt{M}}{k}\right).
\end{aligned}
$$

By Assumption 2.3, we have $k/\sqrt{M} \to 0$. So the test statistics diverge under the alternative hypothesis, which illustrates the consistency. We conclude the proof.

■

Appendix C

Technical Results for Chapter 3

Proof of the main results

Theoretical derivations are all collected here.

Proof of Proposition3.1.

$$\widehat{\beta}_k = \beta_k + \left(\widetilde{Z}'_{k,-1} X_{k,-1}\right)^{-1} \widetilde{Z}'_{k,-1} U_0 + \left(\widetilde{Z}'_{k,-1} X_{k,-1}\right)^{-1} \sum_{t=1}^{T} \widetilde{z}_{k,t-1} x'_{t-1} \left(B - B^{(k)}\right) \left(\frac{t}{T}\right). \tag{C.1}$$

Define

$$\begin{aligned} A_1 : &= \left(\sum_{t=1}^{T} \tilde{z}_{k,t-1} x'_{k,t-1}\right)^{-1} \sum_{t=1}^{T} \tilde{z}_{k,t-1} u_{0t}, \\ A_2 : &= \left(\sum_{t=1}^{T} \tilde{z}_{k,t-1} x'_{k,t-1}\right)^{-1} \sum_{t=1}^{T} \tilde{z}_{k,t-1} x'_{t-1} \left(B - B^{(k)}\right) \left(\frac{t}{T}\right). \end{aligned}$$

(i) For the term A_1,

$$
\begin{aligned}
& \|A_1\|^2 \\
= \quad & \left\| \left(\sum_{t=1}^{T} \tilde{z}_{k,t-1} x'_{k,t-1} \right)^{-1} \sum_{t=1}^{T} \tilde{z}_{k,t-1} u_{0t} \right\|^2 \\
= \quad & tr \left[\left(\sum_{t=1}^{T} \tilde{z}_{k,t-1} x'_{k,t-1} \right)^{-1} \left(\sum_{t=1}^{T} \tilde{z}_{k,t-1} \tilde{z}'_{k,t-1} u_{0t}^2 \right) \left(\sum_{t=1}^{T} x_{k,t-1} \tilde{z}'_{k,t-1} \right)^{-1} \right] \\
= \quad & tr \left[\begin{array}{c} \left(\frac{1}{T^{1+\alpha\wedge\gamma}} \sum_{t=1}^{T} \tilde{z}_{k,t-1} x'_{k,t-1} \right)^{-1} \left(\frac{1}{T^{2+2(\alpha\wedge\gamma)}} \sum_{t=1}^{T} \tilde{z}_{k,t-1} \tilde{z}'_{k,t-1} u_{0t}^2 \right) \\ \cdot \left(\frac{1}{T^{1+\alpha\wedge\gamma}} \sum_{t=1}^{T} x_{k,t-1} \tilde{z}'_{k,t-1} \right)^{-1} \end{array} \right] \\
\leq \quad & \underline{c}_{\Phi}^{-2} \bar{c}_{\Phi} \lambda_{\max} \left(\frac{1}{T^{1+\alpha\wedge\gamma}} \sum_{t=1}^{T} \tilde{z}_{t-1} \tilde{z}'_{t-1} \right) \lambda_{\min} \left(\frac{1}{T^{1+\alpha\wedge\gamma}} \sum_{t=1}^{T} \tilde{z}_{t-1} x'_{t-1} \right)^{-2} \\
& \cdot \left\| \frac{1}{T^{1+\alpha\wedge\gamma}} \left(\tilde{Z}'_{k,-1} \tilde{Z}_{k,-1} \right)^{-\frac{1}{2}} \left(\sum_{t=1}^{T} \tilde{z}_{k,t-1} u_{0t} \right) \right\|^2,
\end{aligned}
$$

where $\underline{c}_{\Phi}$ and $\bar{c}_{\Phi}$ are defined in Assumption 3.4 (i), $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimun and maximun eigenvalues respectively.

By Phillips and Magdalinos (2009), we have

$$
\lambda_{\max} \left(\frac{1}{T^{1+\alpha\wedge\gamma}} \sum_{t=1}^{T} \tilde{z}_{t-1} \tilde{z}'_{t-1} \right) = O_p(1) \quad \text{and} \quad \lambda_{\min} \left(\frac{1}{T^{1+\alpha\wedge\gamma}} \sum_{t=1}^{T} \tilde{z}_{t-1} x'_{t-1} \right) = O_p(1).
$$

Further, we apply the properties of the GARCH process and show that

$$
\begin{aligned}
& \mathbb{E}\left\|\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}\tilde{z}_{k,t-1}'\right)^{-\frac{1}{2}}\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}u_{0t}\right)\right\|^{2} \\
= & \mathbb{E}\left[\operatorname{tr}\left(\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}\tilde{z}_{k,t-1}'\mathbb{E}_{t-1}\left[u_{0t}^{2}\right]\right)\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}\tilde{z}_{k,t-1}'\right)^{-1}\right)\right] \\
= & \Sigma_{00}\mathbb{E}\left[\operatorname{tr}\left(\left(\tilde{Z}_{k,-1}'\tilde{Z}_{k,-1}\right)\left(\tilde{Z}_{k,-1}'\tilde{Z}_{k,-1}\right)^{-1}\right)\right] \\
& +\mathbb{E}\left[\operatorname{tr}\left(\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}\tilde{z}_{k,t-1}'\left(H_t-\Sigma_{00}\right)\right)\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}\tilde{z}_{k,t-1}'\right)^{-1}\right)\right] \\
\leq & \Sigma_{00}\operatorname{tr}\left(I_n\otimes I_k\right)+\mathbb{E}\left[\sup_{1\leq t\leq T}\left|H_t-\Sigma_{00}\right|\operatorname{tr}\left(\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}\tilde{z}_{k,t-1}'\right)\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}\tilde{z}_{k,t-1}'\right)^{-1}\right)\right] \\
= & \Sigma_{00}\operatorname{tr}\left(I_n\otimes I_k\right)+\mathbb{E}\left[\sup_{1\leq t\leq T}\left|H_t-\Sigma_{00}\right|\cdot\operatorname{tr}\left(I_n\otimes I_k\right)\right] \\
= & O_p\left(k\right)+o_p\left(k\right).
\end{aligned}
$$

By Markov inequality, we have,

$$
\left\|\frac{1}{T^{1+\alpha\wedge\gamma}}\left(\tilde{Z}_{k,-1}'\tilde{Z}_{k,-1}\right)^{-\frac{1}{2}}\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}u_{0t}\right)\right\|^{2}=O_p\left(k\right).
$$

Combining the above results, we have $\|A_1\| = O_p\left(\sqrt{\frac{k}{T^{1+\alpha\wedge\gamma}}}\right)$.

(ii) For the term A_2,

$$
\begin{aligned}
\|A_2\| = & \left\|\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}x_{k,t-1}'\right)^{-1}\sum_{t=1}^{T}\tilde{z}_{k,t-1}x_{t-1}'\left(B-B^{(k)}\right)\left(\frac{t}{T}\right)\right\| \\
\leq & \left\|\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}x_{k,t-1}'\right)^{-1}\right\|\left\|\sum_{t=1}^{T}\tilde{z}_{k,t-1}x_{t-1}'\left(B-B^{(k)}\right)\left(\frac{t}{T}\right)\right\| \\
\leq & \underline{c}_{\Phi}^{-1}\left[\lambda_{\min}\left(\frac{1}{T^{1+\alpha\wedge\gamma}}\sum_{t=1}^{T}I_k\otimes\left(\tilde{z}_{t-1}x_{t-1}'\right)\right)\right]^{-1}\cdot\bar{c}_{\Phi}^{1/2}\lambda_{\max}\left(\frac{1}{T^{1+\alpha\wedge\gamma}}\sum_{t=1}^{T}I_k\otimes\left(\tilde{z}_{t-1}x_{t-1}'\right)\right) \\
& \sup_{0\leq r\leq 1}\left\|\left(B-B^{(k)}\right)(r)\right\|,
\end{aligned}
$$

where $\underline{c}_{\Phi}$ and $\bar{c}_{\Phi}$ are defined in Assumption 3.4(i), $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues respectively.

By Phillips and Magdalinos (2009), we have

$$\lambda_{\min}\left(\frac{1}{T^{1+\alpha\wedge\gamma}}\sum_{t=1}^{T} I_k \otimes \left(\tilde{z}_{t-1}x'_{t-1}\right)\right) = \lambda_{\min}\left(\frac{1}{T^{1+\alpha\wedge\gamma}}\sum_{t=1}^{T}\tilde{z}_{t-1}x'_{t-1}\right) = O_p(1) \quad \text{and,}$$

$$\lambda_{\max}\left(\frac{1}{T^{1+\alpha\wedge\gamma}}\sum_{t=1}^{T} I_k \otimes \left(\tilde{z}_{t-1}x'_{t-1}\right)\right) = \lambda_{\max}\left(\frac{1}{T^{1+\alpha\wedge\gamma}}\sum_{t=1}^{T}\tilde{z}_{t-1}x'_{t-1}\right) = O_p(1).$$

By Assumption 3.4(ii), $\sup_{0\le r\le 1}\left\|\left(B-B^{(k)}\right)(r)\right\| = O_p\left(k^{-q}\right)$ under the smoothness condition. Combining the above results, we have $\|A_2\| = O_p\left(k^{-q}\right)$.

(iii) By combining results in Proposition 3.1 (i) and (ii), we have

$$\left\|\widehat{\beta}_k - \beta_k\right\| \le \|A_1\| + \|A_2\| = O_p\left(\sqrt{\frac{k}{T^{1+\alpha\wedge\gamma}}} + k^{-q}\right).$$

Proof of Theorem 3.1.

(i) By Proposition 3.1,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left\|\widehat{B}^{k}\left(\frac{t}{T}\right) - B\left(\frac{t}{T}\right)\right\|^2 \\
=\ & \frac{1}{T}\sum_{t=1}^{T}\left\|\left(f'_k\left(\frac{t}{T}\right)\otimes I_n\right)\left(\widehat{\beta}_k-\beta_k\right) + \left(B^{(k)}\left(\frac{t}{T}\right) - B\left(\frac{t}{T}\right)\right)\right\|^2 \\
\le\ & \frac{2}{T}\sum_{t=1}^{T}\left\|\left(\widehat{\beta}_k-\beta_k\right)'\left[I_n\otimes f_k\left(\frac{t}{T}\right)\right]\left[f'_k\left(\frac{t}{T}\right)\otimes I_n\right]\left(\widehat{\beta}_k-\beta_k\right)\right\| \\
& + \frac{2}{T}\sum_{t=1}^{T}\left\|B^{(k)}\left(\frac{t}{T}\right) - B\left(\frac{t}{T}\right)\right\|^2 \\
\le\ & 2\left\|\widehat{\beta}_k-\beta_k\right\|^2\cdot\sup_{1\le t\le T}\lambda_{\max}\left(f_k\left(\frac{t}{T}\right)f'_k\left(\frac{t}{T}\right)\right) + 2\sup_{1\le t\le T}\left\|\left(B-B^{(k)}\right)\left(\frac{t}{T}\right)\right\|^2 \\
=\ & O_p\left(\frac{k}{T^{1+\alpha\wedge\gamma}} + k^{-2q}\right),
\end{aligned}$$

provided that $0 < \sup_{0\le r\le 1}\lambda_{\max}\left(f_k\left(\frac{t}{T}\right)f'_k\left(\frac{t}{T}\right)\right) < \infty$ under Assumption 3.4(i).

(ii) We intend to bound that

$$\begin{aligned}
&\sup_{1\le t\le T}\left\|\widehat{B}^{k}\left(\frac{t}{T}\right)-B\left(\frac{t}{T}\right)\right\| \\
=&\sup_{1\le t\le T}\left\|\left(f_k'\left(\frac{t}{T}\right)\otimes I_n\right)\left(\widehat{\beta}_k-\beta_k\right)+\left(B^{(k)}\left(\frac{t}{T}\right)-B\left(\frac{t}{T}\right)\right)\right\| \\
=&\sup_{1\le t\le T}\left\|f_k\left(\frac{t}{T}\right)\right\|\cdot\left\|\widehat{\beta}_k-\beta_k\right\|+\sup_{1\le t\le T}\left\|\left(B-B^{(k)}\right)\left(\frac{t}{T}\right)\right\| \\
=&\zeta(k)\,O_p\left(\sqrt{\frac{k}{T^{1+\alpha\wedge\gamma}}}+k^{-q}\right)+O_p\left(k^{-q}\right) \\
=&\zeta(k)\,O_p\left(\sqrt{\frac{k}{T^{1+\alpha\wedge\gamma}}}+k^{-q}\right).
\end{aligned}$$

As the B-spline is considered, $\zeta(k)\lesssim\sqrt{k}$. Then the proof is complete. ■

Proof of Corollary 3.1. The proof follows Theorem 3.1 naturally. ■

Proof of Theorem 3.2.

Without losing generality, we consider one grid point, $d=1$. The proof of $d\ge 2$ follows a similar proving strategy. For the chosen $t^*=\lfloor Tr^*\rfloor$, note the fact that

$$\begin{aligned}
&\widehat{B}^{(k)}\left(\frac{t^*}{T}\right)-B\left(\frac{t^*}{T}\right) \\
=&f_k'\left(\frac{t^*}{T}\right)\otimes I_n\left(\widehat{\beta}_k-\beta_k\right)+\left(B^{(k)}\left(\frac{t^*}{T}\right)-B\left(\frac{t^*}{T}\right)\right),
\end{aligned}$$

where $\left(\widehat{\beta}_k-\beta_k\right)$ can be expressed as formula (C.1). Define

$$\begin{aligned}
A_3:&=M_k^{-\frac{1}{2}}\left[f_k'\left(\frac{t^*}{T}\right)\otimes I_n\right]\left(\tilde{Z}_{k,-1}'X_{k,-1}\right)^{-1}\tilde{Z}_{k,-1}'U_0, \\
A_4:&=M_k^{-\frac{1}{2}}\left[f_k'\left(\frac{t^*}{T}\right)\otimes I_n\right]\left(\tilde{Z}_{k,-1}'X_{k,-1}\right)^{-1}\left(\sum_{t=1}^{T}\tilde{z}_{k,t-1}x_{t-1}'\left(B-B^{(k)}\right)\left(\frac{t}{T}\right)\right), \\
A_5:&=M_k^{-\frac{1}{2}}\left(B^{(k)}\left(\frac{t^*}{T}\right)-B\left(\frac{t^*}{T}\right)\right).
\end{aligned}$$

Suppose Assumptions 3.1 to 3.5 hold, it is sufficient to show that (a) $A_3\rightsquigarrow\mathcal{N}(0,\Sigma_{00})$; (b) $A_4=o_p(1)$; (c) $A_5=o_p(1)$.

For part (a), let $V_t:=M_k^{-\frac{1}{2}}\left[f_k'\left(\frac{t^*}{T}\right)\otimes I_n\right]\left(\tilde{Z}_{k,-1}'X_{k,-1}\right)^{-1}\tilde{z}_{k,t-1}u_{0t}$, so that

$A_3 = \sum_{t=1}^T V_t$. For each t, V_t is a martingale difference sequence. Therefore the martingale central limit theorem can applied to show the limiting distribution.

We first show the conditional stability condition holds. For any n dimensional $\delta \in \mathbb{R}^n_+$,

$$\sum_{t=1}^T \mathbb{E}\left[(\delta' V_t)^2 | \mathcal{F}_{t-1}\right] = \sum_{t=1}^T \mathbb{E}\left[tr\left(\delta\delta' V_t V_t'\right) | \mathcal{F}_{t-1}\right] = tr\left(\delta\delta' M_k^{-\frac{1}{2}} M_k M_k^{-\frac{1}{2}} \Sigma_{00}\right) = \Sigma_{00}\delta'\delta.$$

In addition, the Lindeberg condition can be proved in the following way. For any $\varepsilon > 0$,

$$\begin{aligned}
& \sum_{t=1}^T \mathbb{E}_{t-1}\left[(\delta' V_t)^2 \mathbf{1}\left(|\delta' V_t| > \varepsilon\right)\right] \\
= \; & \varepsilon^2 \sum_{t=1}^T \mathbb{E}_{t-1}\left[\left(\frac{\delta' V_t}{\varepsilon}\right)^2 \mathbf{1}\left(\left|\frac{\delta' V_t}{\varepsilon}\right| > 1\right)\right] \\
\leq \; & \varepsilon^2 \sum_{t=1}^T \mathbb{E}_{t-1}\left[\left(\frac{\delta' V_t}{\varepsilon}\right)^4\right] \\
\leq \; & \frac{\delta'\delta}{\varepsilon^2}\left[\lambda_{\min}\left(\frac{1}{T^{1+\alpha\wedge\gamma}}\sum_{t=1}^T \tilde{z}_{k,t-1}\tilde{z}_{k,t-1}'\right)\right]^{-2} \sum_{t=1}^T \|\tilde{z}_{k,t-1}\|^4 \mathbb{E}_{t-1} u_{0t}^4 \cdot \frac{\zeta^2(k)}{T^{2+2(\alpha\wedge\gamma)}} \\
\lesssim \; & O_p\left(\zeta^2(k)\frac{k}{T^{1+\alpha\wedge\gamma}}\right) \\
= \; & o_p(1).
\end{aligned}$$

By combining the above results and applying the martingale central limit theorem (Proposition 3.1, Hall & Heyde, 2014), we can show that

$$A_3 \rightsquigarrow \mathcal{N}(0, \Sigma_{00}).$$

For part (b), we consider A_4:

$$\begin{aligned}
\|A_4\| &= \left\| M_k^{-\frac{1}{2}} \left[f_k'\left(\frac{t^*}{T}\right) \otimes I_n\right] \left(\tilde{Z}_{k,-1}' X_{k,-1}\right)^{-1} \left[\sum_{t=1}^T \tilde{z}_{k,t-1} x_{t-1}' \left(B - B^{(k)}\right)\left(\frac{t}{T}\right)\right]\right\| \\
&\leq \underline{c}_\Phi^{-\frac{1}{2}} \bar{c}_\Phi^{\frac{1}{2}} \left[\lambda_{\min}\left(\frac{1}{T^{1+\alpha\wedge\gamma}}\sum_{t=1}^T \tilde{z}_{t-1}\tilde{z}_{t-1}'\right)\right]^{-\frac{1}{2}} \lambda_{\max}\left(\frac{1}{T^{1+\alpha\wedge\gamma}}\sum_{t=1}^T \tilde{z}_{t-1} x_{t-1}'\right) \\
&\quad T^{\frac{1+\alpha\wedge\gamma}{2}} \sup_{1\leq t\leq T} \left\|\left(B - B^{(k)}\right)\left(\frac{t}{T}\right)\right\| \\
&= O_p\left(T^{\frac{1+\alpha\wedge\gamma}{2}} k^{-q}\right) = o_p(1).
\end{aligned}$$

Similarly as part (b), an identical proving strategy applies and (c). We can show that,

$$\|A_5\| = O_p\left(T^{\frac{1+\alpha\wedge\gamma}{2}} k^{-q}\right) = o_p\left(1\right).$$

The proof of Theorem 3.2 is complete. ■

Proofs of Theorems 3.3 and 3.4. The derivations follow the limiting normality of Theorem 3.2 and the asymptotic independence between the numerator and the denominator as shown in Phillips & Magdalinos (2009). ■