4-2020

Essays on nonstationary econometrics

Yanbo LIU
Singapore Management University, yanbo.liu.2015@phdecons.smu.edu.sg

ESSAYS ON NONSTATIONARY ECONOMETRICS

YANBO LIU

SINGAPORE MANAGEMENT UNIVERSITY

2020

ESSAYS ON NONSTATIONARY ECONOMETRICS

YANBO LIU

A DISSERTATION

In

ECONOMICS

Presented to the Singapore Management University in Partial Fulfilment

of the Requirements for the Degree of PhD in Economics

2020

____

Supervisor of Dissertation

____

PhD in Economics, Programme Director

Essays on Nonstationary Econometrics

by
Yanbo Liu

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

Dissertation Committee:

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

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

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

Junye Li
Professor of Finance
ESSEC Business School, Paris-Singapore

Singapore Management University
2020

Abstract

My dissertation consists of three essays that contribute new theoretical results to robust inference procedures and machine learning algorithms in nonstationary models.

Chapter 2 compares OLS and GLS in autoregressions with integrated noise terms. Grenander and Rosenblatt (2008) gave sufficient conditions for the asymptotic equivalence of GLS and OLS in deterministic trend extraction. However when extending to univariate autoregression model $y_t = \rho_n y_{t-1} + u_t$, $\rho_n = 1 + \frac{c}{n^\alpha}$, $u_t = u_{t-1} + \epsilon_t$, and ϵ_t is one iid disturbance term with zero expectation and σ^2 variance, the asymptotic equivalence no longer holds. Under the mildly explosive ($c > 0$, $\alpha \in (0, 1)$) and pure explosive ($c > 0$, $\alpha = 0$) cases, the limiting distributions of OLS and GLS estimates are identical as standard Cauchy distribution, and the OLS estimate has a slower convergence rate. Under the mildly stationary ($c < 0$, $\alpha \in (0, 1)$) case, the limiting distribution of OLS is degenerate centered at $-c$, while the GLS estimate is Gaussian distributed. Under the local to unity ($\alpha = 1$) case, when $c \geq c^*$, the mean and variance of the asymptotic distribution of the OLS estimate are smaller than the GLS estimate, showing the efficiency gains in OLS.

Chapter 3 proposes novel mechanisms for identifying explosive bubbles in panel autoregressions with a latent group structure. Two post-classification panel data approaches are employed to test the explosiveness in time-series data. The first approach applies a recursive k-means clustering algorithm to explosive panel autoregressions. The second approach uses a modified k-means clustering algorithm for mixed-root panel autoregressions. We establish the uniform consistency of both

clustering algorithms. The abovementioned k-means procedures achieve the oracle properties so that the post-classification estimators are asymptotically equivalent to the infeasible estimators that use the true group identities. Two right-tailed t-statistics, based on post-classification estimators, are introduced to detect explosiveness. A panel recursive procedure is proposed to estimate the origination date of explosiveness. The asymptotic theory is available for concentration inequalities, clustering algorithms, and right-tailed t-tests based on mixed-root panels. Extensive Monte Carlo simulations provide strong evidence that the proposed panel approaches lead to substantial power gains compared with the time-series approach.

Chapter 4 explores predictive regression models with stochastic unit root (STUR) components and robust inference procedures that encompass a wide class of persistent and time-varying stochastically nonstationary regressors. The paper extends the mechanism of endogenously generated instrumentation known as IVX, showing that these methods remain valid for short- and long-horizon predictive regressions in which the predictors have STUR and local STUR (LSTUR) generating mechanisms. Both mean regression and quantile regression methods are considered. The asymptotic distributions of the IVX estimators are new compared to previous work but again lead to pivotal limit distributions for Wald testing procedures that remain robust for both single and multiple regressors with various degrees of persistence and stochastic and fixed local departures from unity. Numerical experiments corroborate the asymptotic theory, and IVX testing shows good power and size control. The new methods are illustrated in an empirical application to evaluate the predictive capability of economic fundamentals in forecasting excess returns in the Dow Jones industrial average index.

Table of Contents

Acknowledgments

Most of all, I would like to express my deepest gratitude to my co-supervisor, Professor Peter C.B. Phillips. When I first approached Professor Phillips in the spring semester of 2016 and shyly attempted a conversation, I could never know that this was the starting point of a blessed journey where I found my role model and supporter. By working with Professor Phillips, I acquired the skill of refining the idea, clarifying the ambiguities in theoretical justifications, and writing a paper of quality.

I cannot adequately express my gratitude to my co-supervisor, Professor Jun Yu, for his confidence in me. On the one hand, Professor Yu bestowed me with all the freedom I wanted. On the other hand, whenever I needed support, he was there to share his knowledge and to guide me. During the weekly workshop, Professor Yu was always glad to share his insightful comments, which resulted in significant improvements in the developments of my dissertation.

Special thanks go to Assistant Professor Yichong Zhang for his encouragement and insightful comments. Without his challenging questions, nothing would be the same. I would not come to the weekly workshop with the same desire. I would miss the chance to equip my job market paper with a powerful tool developed by him.

I thank Professor Junye Li for reviewing this dissertation and providing valuable comments. I also thank Professor Tassos Magdalinos, Professor Liangjun Su, Professor Qiying Wang, Professor Jia Li, Professor Kunpeng Li, Professor Shuping Shi and Assistant Professor Wayne Gao Yuan for their suggestive comments on the chapters.

My time at the SMU research group of financial econometrics has been extraordinary, and I am expressing my deepest gratitude to all the friends of this big community. I want to thank Ms. Yajie Zhang for working with me on our joint works. I am grateful to Dr. Zoe Chen Ye, Dr. Ming Zeng, and Dr. Yubo Tao, for their generous feedback on my research papers. I appreciate the support from Associate Professor Weilin Xiao for his helpful discussions on financial econometrics. I also owe thanks to Dr. Xiaohu Wang, Dr. Liang Jiang, Dr. Tao Zeng, Dr. Xiaobin Liu, Dr. Yingjie Dong, Dr. Allen Lui, and Mr. Yaohan Chen, for their suggestions and knowledge. In addition to our research group, I also would like to thank Dr. Jiangyuan Li, Mr. Xin Zheng for their friendship. Special thanks also go to my co-author and friend, Dr. Nan Liu, for helping debug technical details, typos, and writings of all chapters with enormous warmness and patience.

Most importantly, I would like to thank my parents for their love. They always stood by my side and supported me unconditionally. My dissertation is dedicated to them.

Chapter 1 Introduction

Nonstationary phenomena are commonly observed in return predictions and bubble detections. This dissertation comprises three papers that solve several significant problems in nonstationary time series models.

In the first essay, we focus on the asymptotic efficiency of estimates for time series data with roots in the vicinity of unity. In Chapter 2, co-authored with Professor Peter C.B. Phillips and Professor Jun Yu, we compare the asymptotic efficiency between the OLS and the detrending GLS estimates in the model with integrated errors. We derive the limiting distributions for pure explosive, mildly explosive, mildly stationary, and local-to-unity time series models. We find that the asymptotic equivalence between the OLS and GLS estimates does not exist. For pure explosive and mildly explosive cases, the detrending GLS estimate has a faster convergence rate. For mildly stationary cases, the OLS estimate has a degenerate distribution with an asymptotic bias, while the detrending GLS estimate follows an asymptotically normal distribution located at the origin. For the local-to-unity case, when the distance parameter is higher than a cut-off value, both the asymptotic bias and variance of the OLS estimate are smaller than the GLS counterpart. The observations under the local-to-unity case show the efficiency of OLS under a specific situation and complement the econometric theory.

In the second essay, we apply the machine learning algorithm to bubble detections. In the literature, the bubble detector implements the testing procedure based on a single time series and suffers from a severe problem of lack of powers. In Chapter 3, co-authored with Professor Jun Yu, we apply the panel approach to improve the performance of bubble detections. We impose the latent group structure on the panel model where we assume the roots within the same group are identical, and the slopes across groups are heterogeneous. We believe there are three types of groups: stationary groups, unity groups, and explosive groups. The explosive roots represent bubble phenomena. We develop a two-stage algorithm with a k-means clustering algorithm in the first stage and inferences in the second stage. We show that the clustering algorithm consistently recovers group identities, and asymptotically the estimated membership is the true membership. Besides, under the joint convergence framework, we show that the limiting distribution of our estimator is normal, and the panel t-statistic is pivotally distributed under the null hypothesis of a unit root. For the structural break case switching from non-explosive regions to explosive regions, we also propose a real-time detector to estimate the bubble origination date. In this case, the two-stage procedure still works. The test procedure can consistently estimate the bubble origination dates at a faster rate than its time series counterpart in the literature.

In the third essay, my focus turns to analyze the predictive powers of economic fundamentals on asset returns. In Chapter 4, co-authored with Professor Phillips, we extend the self-generated instrumentation to predictive regressions with unstable parameters. Due to the persistence of economic fundamentals, the spurious correlation exists and contributes to the endogeneity problem of OLS estimates on predictive regressions. This essay considers the STUR and local STUR models and employs the robust inference procedure (IVX) based on self-generated instruments. We show that with STUR and local STUR regressors, the IVX estimator follows an asymptotic normal distribution. The Wald test based on the IVX estimator follows an asymptotically chi-square distribution under the null hypothesis of no predictive phenomena. The IVX estimators in quantile regressions and long-horizon predictive regressions are also considered. Under both cases, the variant estimators of IVX follow asymptotically normal distributions and contribute to the pivotal test statistics. We also extend the above results to mixed-root cases. In the empirical analysis, we apply the IVX inference procedure to the index return of the S&P 500 stocks and find the significant predictive powers of economic fundamentals.

Chapter 2 Asymptotic Comparisons of OLS and GLS in Autoregressions with Integrated Disturbance

2.1 Introduction

Since it was first introduced by Aitken (1936) as an alternative to the ordinary least squares (OLS) method, the generalized least squares (GLS) estimate has been widely used in practice, especially when the error term in a regression model is serially correlated. In general, it is found that GLS and its detrending version are more efficient than OLS, at least asymptotically.

There exist models, where OLS and GLS are equivalent asymptotically. For example, Grenander and Rosenblatt (2008) found the sufficient conditions for this asymptotic equivalence in the context of time series regressions when the regressor has deterministic trends, and the error process is stationary. Hannan (2009) extended the Grenander-Rosenblatt theorem to the multivariate time series regression. Park and Phillips (1988) extended the Grenander-Rosenblatt theorem to the multivariate time series regression model, where the regressors have stochastic trends. Krämer and Hassler (1998) extended the Grenander-Rosenblatt theorem to the univariate time series regression model where the regressor is integrated of order d with $d \in (0.5, 1.5)$. The sufficient conditions developed in these studies are satisfied in many practically relevant time series. Therefore, empirical researchers can employ OLS in these time series models without sacrificing asymptotic efficiency loss relative to

GLS.

The Grenander-Rosenblatt theorem relies on the continuity of the spectrum of the error process at the origin. This condition is satisfied when the error process is stationary. However, it is violated when the error process has a unit root. Phillips and Lee (1996) showed that when the error process is integrated of order d with $0.5 \leq d < 1$ or has a root which is local to unity, the asymptotic equivalence of GLS and OLS breaks down and GLS is more efficient than OLS asymptotically. Xiao and Phillips (2002) extended the result of Phillips and Lee (1996) to the model where the error process is integrated of order d with $0.5 \leq d < 1.5$.

This paper compares the asymptotic efficiency of OLS and GLS in the context of autoregression (AR) when the error process has a unit root and the true slope is near unit root.

The paper shows that for the explosive side of the model, OLS is inferior to GLS only in the convergence rate. Still, for the stationary side of the model, OLS badly behaves with an asymptotic bias and degenerate distribution. The most exciting phenomenon occurs for the case of the local-to-unity model since the domain of distance parameter determines the comparisons of asymptotic behaviors between OLS and GLS. It is revealed that the convergence rates of OLS and GLS are identical. However, one cut-off point c^* divides the domain of distance parameter into two parts: when smaller than the cut-off point c^*, OLS has a larger bias in the sense of absolute value. For the domain on the right-hand side of the cut-off value, OLS has both smaller bias and smaller variance. This observation proves the efficiency gains of OLS under special cases.

The outline of the paper is as follows. In Section 2.2, the paper introduces the model. In Section 2.2.1 and 2.2.2, the asymptotic theory for explosive root is given. In Section 2.2.3, the asymptotic theory for stationary root is provided. In Section 2.2.4, the asymptotic theory for local-to-unity case is given. Section 2.3 contains a brief concluding comment. All the proofs are collected in the appendix.

2.2 Model Specification and Asymptotic Theory

Suppose a time series y_t is generated from the following model

$$y_t = \rho_n y_{t-1} + u_t, t = 0, 1, \ldots n, \, u_t = u_{t-1} + \epsilon_t, \tag{2.2.1}$$

where $\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$. Assume the true value of ρ_n takes one of the following cases:

$$\begin{cases} \text{Case 1} & \rho_n = a > 1, \\ \text{Case 2} & \rho_n = 1 + \frac{c}{k_n}, \; c > 0, \\ \text{Case 3} & \rho_n = 1 + \frac{c}{k_n}, \; c < 0, \\ \text{Case 4} & \rho_n = 1 + \frac{c}{n}, \end{cases}$$

where k_n satisfies the condition $\frac{1}{k_n} + \frac{k_n}{n} \to 0$. One widely used function that satifies this condition is $k_n = n^\alpha$ with $\alpha \in (0, 1)$.

The OLS estimator of ρ_n is given by

$$\widehat{\rho}_{nols} = \sum_{t=1}^{n} y_{t-1} y_t / \sum_{t=1}^{n} y_{t-1}^2. \tag{2.2.2}$$

The detrending GLS regression performs OLS estimation on the first-order differenced model:

$$\Delta y_t = \rho_n \Delta y_{t-1} + \Delta u_t = \rho_n \Delta y_{t-1} + \epsilon_t, t = 1, 2, \ldots, n, \tag{2.2.3}$$

where $\Delta y_t = y_t - y_{t-1}$. The detrending GLS estimator of ρ_n is

$$\widehat{\rho}_{ngls} = \sum_{t=2}^{n} \Delta y_{t-1} \Delta y_t / \sum_{t=2}^{n} \Delta y_{t-1}^2. \tag{2.2.4}$$

The GLS estimate clearly makes use of the covariance structure of u_t.

2.2.1 Case 1: pure explosiveness

In case 1, the model is

$$\begin{aligned} y_t &= \rho_n y_{t-1} + u_t, a > 1, \\ u_t &= u_{t-1} + \epsilon_t, \\ y_0 &= 0, u_0 = 0, \epsilon_1 = 0. \end{aligned} \tag{2.2.5}$$

Denote $\widehat{\rho}_{ols} := \widehat{\rho}_{n_{ols}}$ and $\widehat{\rho}_{gls} := \widehat{\rho}_{n_{gls}}$. Denote ρ_n as ρ, so that $\rho = a > 1$. The following theorem reports the limiting distribution of two important variables and GLS estimate.

Theorem 2.2.1 *Assume $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ for $t = 2, ..., n$ in Model (2.2.5). Denote*

$$X_{n,gls} := \sum_{t=1}^{n} a^{-(n-t)-1}\epsilon_t, Y_{n,gls} := \sum_{j=1}^{n} a^{-j}\epsilon_j. \tag{2.2.6}$$

As $n \to \infty$,

$$\begin{pmatrix} X_{n,gls} \\ Y_{n,gls} \end{pmatrix} \overset{a.s.}{\to} \mathcal{N}\left(0_{2\times 1}, \frac{\sigma^2}{1-a^2} I_2\right), \tag{2.2.7}$$

where I_2 is a 2-dimensional identity matrix. The limiting distribution of $\widehat{\rho}_{gls}$ is

$$\frac{a^n}{a^2 - 1}\left(\widehat{\rho}_{gls} - a\right) \overset{a.s.}{\to} \mathcal{C}, \tag{2.2.8}$$

where $\mathcal{C}$ is a standard Cauchy variate.

Remark 2.2.1 *The results in (2.2.8) are the same as those in White (1958) and Anderson (1959). The above results are not surprising as in the first-order differenced model $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ and $\Delta y_1 = 0$. Note that the almost sure convergence applies due to the martingale convergence theorem. The normality assumption cannot be relaxed, and the invariance principle is not available in this case.*

Similarly, in OLS estimates, the asymptotic normality of another two variables is derived in the following lemma.

Lemma 2.2.1 *Assume $\{\epsilon_t\}_{t=1}^n$ follow joint normality and $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0,\sigma^2)$ for $t = 2,...,n$ in Model (2.2.5). Denote*

$$X_{n,ols} := \frac{1}{\sqrt{n}}\sum_{t=1}^{n} a^{-(n-t)-1}u_t, Y_{n,ols} := \sum_{j=1}^{n} a^{-j}u_j. \tag{2.2.9}$$

As $n \to \infty$,

$$\begin{pmatrix} X_{n,ols} \\ Y_{n,ols} \end{pmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\sigma^2}{(a-1)^2} & 0 \\ 0 & \frac{a^2\sigma^2}{(a-1)^3(a+1)} \end{bmatrix}\right). \tag{2.2.10}$$

Remark 2.2.2 *The asymptotic covariance matrix of $(X_{n,ols}, Y_{n,ols})$ is diagonal, suggesting that $X_{n,ols}$ and $Y_{n,ols}$ are asymptotically independent. However, it is not proportional to the identity matrix, unlike the asymptotic covariance matrix of $(X_{n,gls}, Y_{n,gls})$.*

Theorem 2.2.2 *Assume $\{\epsilon_t\}_{t=1}^T$ follow joint normality and $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0,\sigma^2)$ for $t = 2,...,n$ in Model (2.2.5). The limiting distribution of $\widehat{\rho}_{ols}$ is*

$$\frac{a^{n+1}}{\sqrt{n}\,(a^2-1)^{\frac{3}{2}}}\left(\widehat{\rho}_{ols} - a\right) \xrightarrow{d} \mathcal{C}, \tag{2.2.11}$$

where $\mathcal{C}$ is a standard Cauchy distribution.

Remark 2.2.3 *Both OLS and GLS estimates follow standard Cauchy distributions with different rates of convergence. The rate for OLS estimate is $a^n/\sqrt{n}$ while the rate for GLS estimate is a^n. Obviously, OLS estimate converges more slowly than the GLS counterpart. This phenomenon suggests that GLS is more efficient than OLS.*

2.2.2 Case 2: mildly explosiveness

In Case 2, the model is

$$y_t = \rho_n y_{t-1} + u_t, \tag{2.2.12}$$
$$\rho_n = 1 + \frac{c}{k_n}, \frac{1}{k_n} + \frac{k_n}{n} \to 0, c > 0,$$

$$
\begin{aligned}
&u_t = u_{t-1} + \epsilon_t,\ \epsilon_t \overset{i.i.d}{\sim} (0, \sigma^2), \\
&y_0 = o_p\left(\sqrt{k_n}\right), u_0 = o_p\left(\sqrt{k_n}\right), \epsilon_1 = o_p\left(\sqrt{k_n}\right).
\end{aligned}
$$

Note that the normality assumption is not imposed here, different from the case of pure explosive root.

The detrending GLS regression takes the same form as in Equation (2.2.3) with ρ being replaced by $\rho_n(> 1)$. Since $y_1 = \rho_n y_0 + u_0 + \varepsilon_1 = o_p\left(\sqrt{k_n}\right)$, the initial condition of the detrending GLS regression model is $\Delta y_1 = o_p\left(\sqrt{k_n}\right)$.

Theorem 2.2.3 *For Model (2.2.12), denote*

$$
X_{n,gls} := \frac{1}{\sqrt{k_n}} \sum_{t=1}^{n} \rho_n^{-(n-t)-1} \epsilon_t,\ Y_{n,gls} := \frac{1}{\sqrt{k_n}} \sum_{j=1}^{n} \rho_n^{-j} \epsilon_j. \tag{2.2.13}
$$

As $n \to \infty$,

$$
\begin{pmatrix} X_{n,ols} \\ Y_{n,ols} \end{pmatrix} \overset{d}{\to} \mathcal{N}\left(0_{2\times1}, \frac{\sigma^2}{2c} I_2\right).
$$

Under Model (2.2.12), the limiting distribution of $\widehat{\rho}_{ngls}$ is

$$
\frac{\rho_n^n k_n}{2c} (\widehat{\rho}_{ngls} - \rho_n) \overset{d}{\to} \mathcal{C}, \tag{2.2.14}
$$

where $\mathcal{C}$ is a standard Cauchy distribution.

Remark 2.2.4 *The limiting distribution in (2.2.14) is identical to the case in Phillip-s and Magdalinos (2007). Similar to Phillips and Magdalinos (2007), an invariance principle applies so that we do not need to assume the normally distributed errors.*

Lemma 2.2.2 *For Model (2.2.12), denote*

$$
X_{n,ols} := \frac{1}{k_n \sqrt{n}} \sum_{t=1}^{n} \rho_n^{-(n-t)-1} u_t, Y_{n,ols} := k_n^{-\frac{3}{2}} \sum_{j=1}^{n} \rho_n^{-j} u_j.
$$

As $n \to \infty$,

$$\begin{pmatrix} X_{n,ols} \\ Y_{n,ols} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\sigma^2}{c^2} & 0 \\ 0 & \frac{\sigma^2}{2c^3} \end{bmatrix} \right).$$

Remark 2.2.5 *The asymptotic covariance matrix for $X_{n,ols}$ and $Y_{n,ols}$ is diagonal, but not proportional to the identity matrix. This covariance matrix indicates the asymptotic independence between $X_{n,ols}$ and $Y_{n,ols}$. The reason is that $X_{n,ols}$ and $Y_{n,ols}$ converge to normal distributions involving different normalizations, i.e., $k_n\sqrt{n}$ and $k_n^{\frac{3}{2}}$, respectively.*

Theorem 2.2.4 *For Model (2.2.12), the limiting distribution of $\widehat{\rho}_{n ols}$ is*

$$\frac{k_n^{\frac{3}{2}}\rho_n^n}{\sqrt{n}\,(2c)^{\frac{3}{2}}}(\widehat{\rho}_{n ols} - \rho_n) \xrightarrow{d} \mathcal{C}, \tag{2.2.15}$$

where $\mathcal{C}$ is a standard Cauchy variate.

Remark 2.2.6 *Both estimators follow standard Cauchy distributions asymptotically with different rates of convergence. The convergence rate for OLS is $\frac{k_n^{\frac{3}{2}}\rho_n^n}{\sqrt{n}}$, while the rate for GLS is $k_n\rho_n^n$. Similarly, the OLS estimate converges more slowly than the GLS estimate. Again, we demonstrate the efficiency of GLS over OLS.*

2.2.3 Case 3: mildly stationarity

In Case 3, the model is

$$\begin{aligned} & y_t = \rho_n y_{t-1} + u_t,\ t = 1, 2, \dots, n, \\ & \rho_n = 1 + \frac{c}{k_n}, \frac{1}{k_n} + \frac{k_n}{n} \to 0,\ c < 0, \\ & u_t = u_{t-1} + \epsilon_t,\ \epsilon_t \overset{iid}{\sim} (0, \sigma^2), \\ & y_0 = o_p(\sqrt{k_n}), u_0 = o_p(\sqrt{k_n}), \epsilon_1 = o_p(\sqrt{k_n}). \end{aligned} \tag{2.2.16}$$

Different from Phillips and Magdalinos (2007) where a high-order moment condition is needed ($\mathbb{E}|\epsilon_1|^{2+\delta} < \infty$ for some $\delta > 0$), in our discussions, only the second-moment condition of ϵ_t has to be finite. Following the approach of Phillips and Magdalinos (2007) we develop the asymptotic theory of OLS and GLS estimates on Equation (2.2.16).

The detrending GLS regression takes the same form as in Equation (2.2.3) with ρ being replaced by $\rho_n(:= 1 + \frac{c}{k_n} < 1)$. Since $y_1 = \rho_n y_0 + u_0 + \varepsilon_1 = o_p(\sqrt{k_n})$, the initial condition in the detrending GLS regression model is $\Delta y_1 = o_p\left(\sqrt{k_n}\right)$. As $n \to \infty$, we have the following limit theory.

Theorem 2.2.5 *For Model (2.2.16), as* $n \to \infty$*,*

(a) $\frac{1}{nk_n} \sum_{t=1}^{n} (\Delta y_t)^2 \xrightarrow{p} \frac{\sigma^2}{-2c}$.

(b) $\frac{1}{\sqrt{nk_n}} \sum_{t=1}^{n} \epsilon_t \Delta y_{t-1} \xrightarrow{d} \mathcal{N}(0, \frac{\sigma^4}{-2c})$.

(c) $\sqrt{nk_n}(\widehat{\rho}_{n gls} - \rho_n) \xrightarrow{d} \mathcal{N}(0, -2c)$.

When OLS is conducted for Model (2.2.16), we have the following theorem.

Theorem 2.2.6 *For Model (2.2.12), as* $n \to \infty$*,*

(a) $\frac{1}{k_n^2 n^2} \sum_{t=1}^{n} y_{t-1}^2 = \frac{1}{-2c} \left\{ \frac{2\rho_n \sum_{t=1}^{n} y_{t-1} u_t}{n^2 k_n} \right\} + o_p(1)$;

(b) $\frac{\rho_n}{n^2 k_n} \sum_{t=1}^{n} y_{t-1} u_t = O_p(1)$;

(c) $k_n(\widehat{\rho}_{n ols} - \rho_n) \xrightarrow{p} -c$*, where* c *is the distance parameter.*

Remark 2.2.7 *Since* $\sqrt{nk_n}/\left(k_n\right) \to \infty$*, GLS has a faster rate of convergence and hence is asymptotically more efficient. Similar efficiency gains of the detrending GLS estimates are demonstrated in Phillips and Lee (1996) and Xiao and Phillips (2002). Moreover, the OLS estimate follows a degenerate distribution. One possible explanation is that the unit root has spectra with a singularity (a pole) at the origin.*

2.2.4 Case 4: local-to-unity Autoregression

The local-to-unity model initiated in Phillips (1987b) characterizes the near-unity behaviors. The model follows,

$$y_t = \rho_n y_{t-1} + u_t, \; t = 1, 2, \ldots, n, \tag{2.2.17}$$

$$
\begin{aligned}
\rho_n &= 1+\frac{c}{n},\ -\infty < c < \infty, \\
u_t &= u_{t-1}+\epsilon_t,\ \epsilon_t \overset{i.i.d}{\sim} (0,\sigma^2), \\
y_0 &= o_p(1),\ u_0 = o_p(1),\ \epsilon_1 = o_p(1).
\end{aligned}
$$

If $c = 0$, y_t has two unit-roots as differencing y_t twice gives ϵ_t. In this case, differencing y_t and running OLS on the first differenced equation yield the following well-known asymptotic theory

$$
n(\widehat{\rho}_{ngls} - 1) \Rightarrow \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W^2(r)dr}, \tag{2.2.18}
$$

where $W(r)$ is a standard Brownian motion. The discussion of I(2) co-integration has also provided in Phillips and Chang (1994), and Harris (1996), showing the $O_p(n^2)$ convergence rate.

When $c \neq 0$, the detrending GLS regression takes the same form as in Equation (2.2.3) with ρ being replaced by $\rho_n = 1 + \frac{c}{n}$. Let $J_c(r) := \int_0^r e^{c(r-s)}dW(s) \sim \mathcal{N}\left(0, \frac{e^{rc}-1}{2c}\right)$. The limiting distribution of GLS is summarized in the following theorem.

Theorem 2.2.7 *For Model (2.2.17), as $n \to \infty$,*

$$
n\left(\widehat{\rho}_{ngls} - \rho\right) \Rightarrow \frac{\int_0^1 J_c(r)dW(r)}{\int_0^1 J_c^2(r)dr}.
$$

OLS is conducted over model (2.2.17) directly. Denote $I_c(r) := \int_0^r e^{c(r-s)}W(s)ds$ satisfying $dI_c(r) = (cI_c(r) + W(r))dr$ with $I_c(r) = 0$. The asymptotic results of the OLS estimate are summarized in the following theorem.

Theorem 2.2.8 *For Model (2.2.17), as $n \to \infty$ and $t = [Tr]$,*

(a) $n^{-\frac{3}{2}}y_{[Tr]} \Rightarrow \sigma I_c(r)$.

(b) $n^{-\frac{5}{2}}\sum_{t=1}^n y_t \Rightarrow \sigma\int_0^1 I_c(r)dr$.

(c) $n^{-4}\sum_{t=1}^n y_t^2 \Rightarrow \sigma^2\int_0^1 I_c^2(r)dr$.

(d) $n^{-3}\sum_{t=1}^n y_{t-1}u_t \Rightarrow \frac{\sigma^2}{2}\left\{I_c(1)\right\}^2 - c\sigma^2\int_0^1 I_c^2(r)dr$.

(e) $n\left(\widehat{\rho}_{n_{ols}} - \rho\right) \Rightarrow \frac{\frac{1}{2}\{I_c(1)\}^2 - c\int_0^1 I_c^2(r)dr}{\int_0^1 I_c^2(r)dr}$.

Remark 2.2.8 *Both $\widehat{\rho}_{n_{gls}}$ and $\widehat{\rho}_{n_{ols}}$ are consistent estimates. The rates of convergence of $\widehat{\rho}_{n_{gls}}$ and $\widehat{\rho}_{n_{ols}}$ are the same (n), although the limit distributions are different. The expectations of the two ratios are not zero, representing the existence of asymptotic bias for both $\widehat{\rho}_{n_{gls}}$ and $\widehat{\rho}_{n_{ols}}$. Unfortunately, it is almost infeasible to compute the moments of $\frac{\frac{1}{2}\{I_c(1)\}^2 - c\int_0^1 I_c^2(r)dr}{\int_0^1 I_c^2(r)dr}$ in the closed forms.*

One exciting fact found here is that the detrending GLS is inferior to OLS in the sense of bias and variance for some region of distance parameter in the local-to-unity model. This observation contradicts the common sense in econometric theory.

By checking Table 2.1, on the stationary and unit root side, the OLS limiting distribution has a larger bias than the detrending GLS case. However, when $c \geq 1$, the asymptotic distribution of OLS has both the smaller bias and smaller variance. Therefore, in contrast to the standard results of GLS and OLS, there are efficiency gains for OLS when $c \geq 1$. Table 2.2 illustrates the abovementioned results. Moreover, from Table 2.3 and Table 2.4, one structure change occurs at $c^* \in (0,\ 1)$. When $0 \leq c \leq c^*$, OLS has a larger bias than detrending GLS. Again, the interesting fact occurs when $c^* \leq c \leq 1$. In this case, the OLS distribution has both a smaller bias and smaller variance, similar to the discussions of $c \geq 1$.

Table 2.1: Local-to-unity on stationary side(c≤0), iteration=2,000

Distance Parameter		c=-3		c=-2		c=-1		c=0	
Estimation Method		OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS
n=100	mean	4.4044	-1.9615	3.5369	-1.9530	2.7164	-1.9114	1.9642	-1.8110
	variance	1.2743	14.9964	1.3089	13.1734	1.3419	11.4686	1.3575	10.0237
n=500	mean	4.5026	-1.8516	3.6150	-1.8473	2.7705	-1.8219	1.9925	-1.7487
	variance	1.3588	15.8125	1.3652	14.0048	1.3591	12.2900	1.3485	10.7483
n=1,000	mean	4.4963	-1.9243	3.6093	-1.9024	2.7683	-1.8565	1.9944	-1.7526
	variance	1.2817	14.9628	1.3010	13.0747	1.3115	11.2435	1.3032	9.4810
n=2,000	mean	4.5038	-2.0183	3.6163	-1.9934	2.7740	-1.9371	1.9976	-1.8144
	variance	1.2605	15.9271	1.2819	13.9369	1.2932	11.9208	1.2816	9.9006

Results for the local-to-unity case are collected in the following corollary.

Corollary 2.2.9 *For Model (2.2.17) with $c^* \in (0, 1)$,*

(a) when $c < c^*$, the detrending GLS estimate has a smaller asymptotic bias;

Table 2.2: Local-to-unity on explosive side(c≥1), iteration=2,000

Distance Parameter		c=1		c=2		c=3		c=4	
Estimation Method		OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS
n=100	mean	1.3060	-1.6087	0.7666	-1.2768	0.3924	-0.8980	0.1784	-0.5477
	variance	1.3187	8.7617	1.2436	7.1120	1.1301	5.4769	0.8629	3.7867
n=500	mean	1.3190	-1.5777	0.7903	-1.2729	0.4094	-0.8884	0.1593	-0.5555
	variance	1.3263	9.3550	1.2262	7.7584	0.9670	6.6490	0.7831	3.7412
n=1,000	mean	1.3153	-1.5481	0.7694	-1.2745	0.3853	-0.9230	0.1578	-0.5655
	variance	1.2691	7.9567	1.1964	7.6006	1.0739	6.0589	0.8404	4.5792
n=2,000	mean	1.3123	-1.5852	0.7538	-1.2633	0.3836	-0.8562	0.1511	-0.5314
	variance	1.2371	8.0033	1.2236	6.3402	0.9360	4.6719	0.7540	3.6020

Table 2.3: Local-to-unit on the explosive side(0≤c≤0.4), iteration=2,000

Distance Parameter		c=0.1		c=0.2		c=0.3		c=0.4	
Estimation Method		OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS
n=100	mean	1.8937	-1.7963	1.8242	-1.7805	1.7905	-1.7195	1.6882	-1.7454
	variance	1.3568	9.8951	1.3556	9.7685	1.4095	10.2341	1.3509	9.5196
n=500	mean	1.9196	-1.7370	1.8479	-1.724	1.7773	-1.7104	1.7151	-1.7191
	variance	1.3474	10.6057	1.3464	10.4649	1.3452	10.3257	1.3537	10.4739
n=1,000	mean	1.9217	-1.7372	1.8500	-1.7207	1.7793	-1.7030	1.6744	-1.7417
	variance	1.3009	9.3116	1.2984	9.1443	1.2955	8.9796	1.2892	9.0974
n=2,000	mean	1.9246	-1.7968	1.8525	-1.7779	1.7813	-1.7579	1.7112	-1.7367
	variance	1.2784	9.7025	1.2748	9.5058	1.2708	9.3106	1.2664	9.1170

Table 2.4: Local-to-unity explosive model(0.4<c<1), iteration=2,000

Distance Parameter		c=0.5		c=0.6		c=0.7		c=0.8	
Estimation Method		OLS	GLS	OLS	GLS	OLS	GLS	OLS	GLS
n=100	mean	1.6218	-1.7260	1.5564	-1.7053	1.4921	-1.6832	1.4290	-1.6598
	variance	1.3475	9.3963	1.3432	9.2728	1.3382	9.1485	1.3323	9.0224
n=500	mean	1.6397	-1.6791	1.5728	-1.6616	1.5073	-1.6427	1.4431	-1.6225
	variance	1.3423	10.0503	1.3403	9.9132	1.3379	9.7758	1.3348	9.6374
n=1,000	mean	1.6410	-1.6642	1.5735	-1.6431	1.5071	-1.6209	1.4419	-1.5976
	variance	1.2890	8.6598	1.2854	8.5061	1.2816	8.3577	1.2776	8.2157
n=2,000	mean	1.6420	-1.7143	1.5739	-1.6907	1.5069	-1.6660	1.4409	-1.6401
	variance	1.2616	8.9253	1.2566	8.7356	1.2515	8.5482	1.2464	8.3635

(b) when $c \geq c^*$, the OLS estimate has both the smaller asymptotic bias and variance.

2.3 Conclusion

The main results discussed above are classified as follows: For autoregressive model with I(1) error as in the (2.2.1), there is no asymptotic equivalence between detrending GLS and OLS,

(1) For pure explosiveness ($\rho_n = a > 1$), the limiting distributions of both the OLS and the detrending GLS estimates are standard Cauchy distributions. However, the OLS estimate has a slower convergence rate than the GLS estimate.

(2) For mildly explosiveness ($\rho_n = 1 + \frac{c}{k_n}$, where $c > 0$), the limiting distributions of both the OLS and the detrending GLS estimates are standard Cauchy distributions. The OLS estimate has a slower convergence rate.

(3) For mildly stationary models ($\rho_n = 1 + \frac{c}{k_n}$, where $c < 0$) and local-to-unity models ($\rho_n = 1 + \frac{c}{n}$ with $c \leq c^*$), there are efficiency gains in the detrending GLS estimate.

(4) For local-to-unity models ($\rho_n = 1 + \frac{c}{n}$ with $c > c^*$), there are efficiency gains in the OLS estimate.

Chapter 3 Panel Approaches to Econometric Analysis of Bubble Behaviour

3.1 Introduction

Financial bubbles, such as the dot-com bubble, are well recognized as explosive deviations of asset prices from their fundamental values. According to the present value model,

$$P_t = \sum_{i=0}^{\infty} \left(\frac{1}{1+r_f} \right)^i \mathbb{E}_t \left(D_{t+i} \right) + B_t, \tag{3.1.1}$$

where at time t, P_t is the price of an asset, D_t is the payoff of the asset, r_f is the risk-free interest rate, and B_t represents the bubble component which satisfies the submartingale property:

$$\mathbb{E}_t \left(B_{t+1} \right) = \left(1 + r_f \right) B_t.$$

When there are no bubbles (i.e., $B_t = 0$), the asset price is completely determined by dividends and unobserved fundamentals. If $\{D_{t+i}\}$ contains a unit root (i.e., I(1)), then the asset prices $\{P_t\}$ cannot be explosive. However, if there is a bubble (i.e., $B_t \neq 0$), $\{B_t\}$ and hence, $\{P_t\}$ must be explosive. This outcome is the economic reason why econometric analysis of bubble behaviour has been focused on doing right-tailed unit root tests on asset prices adjusted by fundamentals; see, for example, Phillips, Shi, and Yu (2015a) (PSY hereafter).

Conventional econometric methods for bubble detection, including the Dickey-Fuller (DF) test and the augmented DF (ADF) test of Diba and Grossman (1987), the SADF test of Phillips, Wu and Yu (2011) (PWY hereafter) and Phillips and

Yu (2009, 2011), and the GSADF test of PSY (2015a, b), are always based on single time-series. After a bubble is found, the time series method is then used to estimate the bubble origination and termination dates. For example, PWY (2011) used the ADF test and the first-crossing principle to timestamp a bubble whereas, PSY (2015a) used the maximum of ADF test and the first-crossing principle to timestamp each bubble when there are multiple bubbles in the sample.

Unfortunately, the time-series methods may not have good powers, especially when a bubble is short-lived or when a bubble grows slowly. To demonstrate the low-power problem of the DF test when a bubble is short-lived or when a bubble grows slowly, we design two experiments. In both experiments, we simulate data from the following explosive AR(1) model,

$$y_t = \rho y_{t-1} + u_t,\ y_0 = 0, u_t \sim N(0,1), t = 1, 2, ..., T, \tag{3.1.2}$$

and use the DF statistic $(\widehat{\rho} - 1)/s.e.(\widehat{\rho})$ to test $\mathcal{H}_0 : \rho = 1$ against $\mathcal{H}_1 : \rho > 1$, where $\widehat{\rho} = \sum_{t=1}^{T}\left(y_t - \frac{1}{T}\sum_{t=1}^{T} y_t\right)\left(y_{t-1} - \frac{1}{T}\sum_{t=1}^{T} y_{t-1}\right) / \sum_{t=1}^{T}\left(y_{t-1} - \frac{1}{T}\sum_{t=1}^{T} y_{t-1}\right)^2$ is the least-squares (LS) estimator of ρ and $s.e.(\widehat{\rho})$ is the standard error of $\hat{\rho}$. In the first experiment, we take the empirical estimate of ρ as found in PWY as the true value of ρ (i.e., $\rho = 1.033$ and $\rho = 1.040$) but set $T = 10, 20, 30$. In this experiment, the bubble is short-lived but empirically realistic, judged by the empirical results reported in PSY (2015a). Table 3.1 reports the powers (i.e., relative frequency out of 10,000 replications) of the right-tailed DF test rejecting the null hypothesis. Essentially, when the bubble is short-lived, the power of the right-tailed DF test is deficient, ranging from 0.0916 to 0.2334. In the second experiment, we take the empirical estimates of ρ as found in empirical evidence as the true value of ρ (i.e., $\rho = 1.0017$ and $\rho = 1.0068$). In this experiment, the bubble grows very slowly, but these growth rates are empirically reasonable judged by the empirical results that will be reported later. Once again, when the bubble grows slowly, the power of the right-tailed DF test is very low, ranging from 0.0598 to 0.2892.

In practice, prices of multiple (say n) assets are often available over the same

Table 3.1: Power of the right-tailed DF test when a bubble is short-lived.

T	10	20	30	10	20	30
ρ		1.033			1.040	
Power	0.0916	0.1202	0.1732	0.0975	0.1403	0.2334

Table 3.2: Power of the right-tailed DF test when the bubble grows slowly

T	50	100	200	50	100	200
ρ		1.0017			1.0068	
Power	0.0598	0.0604	0.0766	0.0791	0.1095	0.2892

time, leading to the availability of panel data. In this paper, we propose to use panel data models to improve the power in bubble detection. Intuitively, as long as there is some homogeneity over cross-sectional units within groups and the group structure is known, panel data models based on pooling cross-sectional data within the same group should sharpen the statistical inferences on the common explosive root and hence, deliver better power performance than the method based on single time series.

Unfortunately, in almost all practically relevant cases, the true group structure is latent and has to be estimated from the panel data. In this paper, to identify the latent membership, we consider several grouping strategies. To determine the number of groups, we use the Bayesian information criterion (BIC).

Two different specifications for the panel data model are considered in this paper. The first specification is the explosive panel autoregressive model. We use the recursive k-means algorithm of Bonhomme and Manresa (2015) to identify the group structure. The second specification is the mixed-root panel autoregressive model in which explosive, stationary, and unit root time series are mixed together. For bubble detection, it is perhaps too restrictive to impose homogeneous explosive roots. Thus, it is more realistic to argue that explosive roots exist in a proportion of individuals. For the case of mixed roots, we apply the modified k-means algorithm of Lin and Ng (2012) to identify the group structure. We show the uniform consistency of both classification algorithms.

We derive the oracle property of two post-classification estimators under the joint asymptotic scheme, that is, $n \to \infty$ and $T \to \infty$. The oracle property reveals that the distance between the post-classification estimators and the oracle-within estimators is diminishing. The diminishing distance verifies the optimality of two the k-means classification procedures from the estimation perspective.

After the classification of groups is made, we provide two right-tailed t-statistics for the detection of explosiveness. Under the null hypothesis of a unit root in a specific group, the proposed statistics converge to standard normal distributions. They diverge under the alternative hypothesis of explosive roots. Our panel t-statistics are superior to the ADF test in two aspects. First, our panel data based t-tests are more powerful than the time series based DF test. Our asymptotic theory shows that the panel t-statistics diverge at a faster rate than the ADF statistic. Extensive Monte Carlo simulations demonstrate that the empirical power of the panel t-statistics is much higher than the ADF statistic. Second, unlike the ADF statistic whose limiting distribution is non-standard, the panel t-statistics have the standard asymptotically normal distribution under the null hypothesis. Hence, it is easier to implement the proposed tests than the ADF test.

Based on the panel t-statistics, we then propose a real-time estimate for the bubble origination date and develop the asymptotic theory of the estimator. In particular, based on the k-means classifications, we employ the forward recursive right-tailed panel t-test to estimate the starting date of the bubble.

Our paper makes several contributions. First, it contributes to the literature on bubble detection. Based on the uniformly consistent classification of individuals, the proposed panel procedure greatly enhances the power of methods based on a single time series.

Second, our paper contributes to the literature on data-driven classification. Although there are several existing classification algorithms, most of the methods are developed for the stationary case only. An exception is the LASSO algorithm in Huang et al. (2019), which, however, does not directly apply to the mixed-root panel autoregression. To the best of our knowledge, our study is the first attempt to extend classification algorithms to mixed-root panels.

We use the following notations throughout the study. The notations $\xrightarrow{p}$, $\xrightarrow{d}$, and $\Rightarrow$ denote convergence in probability, convergence in distribution, and convergence in functional space, respectively. Correspondingly, $(n,T)\to\infty$ denotes the joint limit. The notation $A \succ B$ implies $B/A = o(1)$ as $(n,T)\to\infty$. The notation $A \succ_p B$ implies $B/A = o_p(1)$ as $(n,T)\to\infty$. The notation $A \sim B$ represents $B/A = O_p(1)$ as $(n,T)\to\infty$

3.2 Model Setup

The generic panel autoregressive model we consider is

$$\begin{aligned} y_{it} &= \mu_i + \rho_i y_{i,t-1} + u_{it}, \\ y_{i0} &= o_p\left(T^{\frac{\gamma}{2}}\right), \\ \rho_i &= 1 + \frac{c_i^0}{T^\gamma}, \\ i &= 1,2,...,n, \\ t &= 1,2,...,T, \end{aligned} \tag{3.2.1}$$

where $\{u_{it}\}$ is a martingale difference sequence with a conditional second moment σ^2 (i.e., $\mathbb{E}(u_{it}^2|\mathcal{F}_{i,t-1}) = \sigma^2$ for any i and t, where $\mathcal{F}_{i,t-1} := \sigma\{u_{i,t-1}, u_{i,t-2}, ...\}$) and finite q^{th} moments with $q \geq 4$ for all i and t. We assume $\gamma \in (0,1)$. In this model, μ_i is an individual fixed effect for each i with $\mu_i = O(T^{-1})$ whose magnitude depends on the sample size and is diminishing asymptotically.

For the AR coefficient, we assume there exist positive values c_l and c_u such that $c_i^0 \in [-c_u, -c_l] \cup \{0\} \cup [c_l, c_u]$. The boundedness imposes an identification condition for parameters. Otherwise, there exists an identification problem. Generally, we assume a known and homogeneous scaling parameter γ and unknown heterogeneous distance parameters, $\{c_i^0\}_{i=1}^n$. As we only care about the signs of distance parameters, the value of γ is of no interest.

In this study, we adopt a setup that lies between the homogeneous panel ($c_i^0 = c^0$) and the heterogeneous panel ($c_i^0 \neq c_j^0$ for any $i \neq j$). In particular, we assume the following group structure as

$$c_i^0 = \sum_{g=1}^{K^0} \alpha_g^0 \mathbf{1}\left\{i \in G_g^0\right\}, \tag{3.2.2}$$

where $\alpha_g^0 \neq \alpha_l^0$ for any $g \neq l$, $\bigcup_{g=1}^{K_0} G_g^0 = \{1, 2, ..., n\}$, and $G_g^0 \cap G_l^0 = \varnothing$ for any $j \neq g$. Let $n_g := \# G_g^0$ represents the cardinality of the true group G_g^0. Let $\mathcal{A}$ be a set of arbitrary $K^0 \times 1$ vectors $\alpha\,(:= (\alpha_1, \alpha_2, ..., \alpha_{K^0}))$, and $\mathcal{C}$ be a set of group-specific distance parameters, so that $\overline{c}\,(:= (\overline{c}_1, \overline{c}_2, ..., \overline{c}_{K^0})) \in \mathcal{C}$. Within the same market sector or convergence club, all cross-sectional units share the identical distance parameter α_g. To obtain the asymptotic properties of classification and inference, we first assume that the true group number, K^0, is known while the true memberships are latent and unknown. We then propose to use the BIC to estimate the number of groups.

The true group-specific parameters are defined as $\overline{c}^0 := \left(\overline{c}_1^0, \overline{c}_2^0, ..., \overline{c}_{K^0}^0\right) \in \mathcal{C}$, $\alpha^0 := (\alpha_1^0, ..., \alpha_{K^0}^0) \in \mathcal{A}$ and $c^0 := (c_1^0, ..., c_n^0) \in \Phi^n$, where $\Phi := [-c_u, -c_l] \cup \{0\} \cup [c_l, c_u]$. The true group membership variable $\{g_i^0\}_{i=1}^n$ maps individual units into groups. For each $i = 1, 2, ..., n$, and $g = 1, 2, ..., K^0$, the event '$g_i^0 = g$' is equivalent to '$i \in G_g^0$'. With any estimator $\{\widehat{g}_i\}_{i=1}^n$, the event '$\widehat{g}_i = g$' is equivalent to '$i \in \widehat{G}_g$' for each $i = 1, 2, ..., n$, and $g = 1, 2, ..., K^0$. We denote $\delta := (g_1, g_2, ..., g_n) \in \Delta_{K^0}$ as a particular grouping of n, where Δ_{K^0} unit is the set of all groupings of $\{1, 2, ..., n\}$ into at most K^0 groups. For the g^{th} group, we define the AR coefficients and their estimators as $\overline{\rho}_g^0\left(:= \exp\left(\alpha_g^0/T^\gamma\right)\right)$, $\overline{\rho}_{\widehat{g}}^0\left(:= \exp\left(\alpha_{\widehat{g}}^0/T^\gamma\right)\right)$, $\widehat{\overline{\rho}}_{\widehat{g}}\,(:= \exp\left(\widehat{\alpha}_{\widehat{g}}/T^\gamma\right))$. For simplicity, we write $\rho_i^0\,(:= \exp\left(c_i^0/T^\gamma\right))$ as ρ_i.

Two kinds of panel autoregressive models are considered in this paper. The first one is the pure explosive panel, while the second one is the mixed-root panel. For the explosive panel, its group structure follows,

$$\begin{cases} \text{Group } 1: & \alpha_1^0 > 0 \\ \text{Group } 2: & \alpha_2^0 > 0 \\ \vdots & \vdots \\ \text{Group } K^0: & \alpha_{K^0}^0 > 0 \end{cases}.$$

For the mixed-root panel autoregressive model, three potential classes of groups are considered: (1) explosive roots ($\alpha_g^0 > 0$); (2) unity roots ($\alpha_g^0 = 0$); (3) stationary roots ($\alpha_g^0 < 0$). For the mixed-root panel, the group structure of the mixed-root panels follows as

$$\begin{cases} \text{Explosive Groups}: & \begin{cases} \text{Group } 1: & \alpha_1^0 > 0 \\ \text{Group } 2: & \alpha_2^0 > 0 \\ \vdots & \vdots \\ \text{Group } k: & \alpha_k^0 > 0 \end{cases} \\ \text{Unit Root Group}: & \text{Group } : (k+1) \quad \alpha_{(k+1)}^0 = 0 \\ \text{Stationary Groups}: & \begin{cases} \text{Group } (k+2): & \alpha_{(k+2)}^0 < 0 \\ \text{Group } (k+3): & \alpha_{(k+3)}^0 < 0 \\ \vdots & \vdots \\ \text{Group } K^0: & \alpha_{K^0}^0 < 0 \end{cases} \end{cases}.$$

As it is highly restrictive to assume all individual assets have bubbles, the mixed-root panels accommodate explosive roots in a proportion of cross-sectional units.

3.3 A Two-stage Approach

For explosive analysis, we apply classification methods in the first stage. Based on the estimated group structures in the first stage, we build post-classification estimators and testing statistics in the second stage. We consider two inference procedures for explosive analysis. The first is to detect the existence of explosive root using the right-tailed t-test. The second is to estimate the bubble origination dates using a recursive algorithm. Both inference approaches rely on estimated group identities.

In the first panel model, we employ the recursive k-means classification proposed in Bonhomme and Manresa (2015). In the second panel model, we consider the modified k-means classification proposed in Lin and Ng (2012).

3.3.1 Stage 1: classification

Recursive k-means algorithm for explosive panels

In this subsection, we consider using the recursive k-means classification algorithm to identify the group structure in the explosive panel autoregressive model. When memberships are unobserved, two types of parameters are considered: the parameter vector $\{c_i^0\}_{i=1}^n \subseteq [c_l, c_u]$, and the group membership variable $\{g_i^0\}_{i=1}^n$, which maps cross-sectional units into groups. Note that group-specific distance parameters, $\{c_i^0\}_{i=1}^n$, are well separated with minimum distance $c^* > 0$; otherwise, we cannot correctly allocate individuals into the true groups.

The grouped estimators of $\{c_i^0\}_{i=1}^n$, $\{g_i^0\}_{i=1}^n$ in (3.2.1) are defined as the solution to the following optimization problem:

$$(\widehat{\overline{c}}, \widehat{\delta}) = \arg \min_{(\overline{c},\delta)\in \mathcal{C}\times\Delta_{K^0}} \frac{1}{nT^{2\gamma}} \sum_{i=1}^{n} \frac{1}{\breve{\rho}_i^{2T}} \sum_{t=1}^{T} \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \exp\left(\frac{\overline{c}_{g_i}}{T^{\gamma}} \right) \right)^2, \quad (3.3.1)$$

where $\delta = \{g_1, g_2, ..., g_n\}$ groups n units into K^0 groups. We employ $\{\breve{\rho}_i\}_{i=1}^n$ as the collection of least squares (LS) estimates for each individual time series. To eliminate fixed effects, we employ a demeaned process as $\widetilde{y}_{it} := y_{it} - \overline{y}_i$ and $\widetilde{y}_{i,t-1} := y_{it} - \overline{y}_{i,-1}$. For given values of $\{\overline{c}_g\}_{g=1}^{K^0}$, the optimal group classification for each $i = 1, 2, ..., n$ is

$$\widehat{g}_i(\overline{c}) = \arg \min_{g\in\{1,2,...,K^0\}} \frac{1}{T^{2\gamma}\breve{\rho}_i^{2T}} \sum_{t=1}^{T} \left[\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\exp\left(\frac{\overline{c}_g}{T^{\gamma}} \right) \right]^2, \quad (3.3.2)$$

where the minimum g optimizes a k-means classification problem. The estimator

$\left\{\widehat{\overline{c}}_g\right\}_{g=1}^{K^0}$ of (3.3.1) optimizes the following objective function:

$$\widehat{\overline{c}} = \arg\min_{\overline{c}\in\mathcal{C}} \frac{1}{nT^{2\gamma}} \sum_{i=1}^{n} \frac{1}{\breve{\rho}_i^{2T}} \sum_{t=1}^{T} \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\exp\left(\frac{\overline{c}_{\widehat{g}_i(\overline{c})}}{T^{\gamma}} \right) \right)^2, \qquad (3.3.3)$$

where $\widehat{g}_i(\overline{c})$ is derived by (3.3.2). The classification estimates of $\{g_i^0\}_{i=1}^n$ are simply $\widehat{g}_i(\widehat{\overline{c}})$.

The following algorithm summarizes the recursive k-means procedure to minimize (3.3.1) in the following steps.

Step 1: Let the initial values $\left\{\overline{c}_g^{(0)}\right\}_{g=1}^{K^0}$ be the collection of individual LS estimates for all $i \in \{1, 2, ..., n\}$;

Step 2: For any $i = 1, 2, ..., n$, compute

$$g_i^{(s+1)} = \arg \min_{g\in\{1,2,...,K^0\}} \frac{1}{T^{2\gamma}\breve{\rho}_i^{2T}} \sum_{t=1}^{T} \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \exp\left(\frac{\overline{c}_g^{(s)}}{T^{\gamma}} \right) \right)^2; \qquad (3.3.4)$$

Step 3: Compute

$$\widehat{\overline{c}}_g^{(s+1)} = \arg\min_{\overline{c}\in\mathcal{C}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T^{2\gamma}\breve{\rho}_i^{2T}} \sum_{t=1}^{T} \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \exp\left(\frac{\overline{c}_{g_i^{(s+1)}}}{T^{\gamma}} \right) \right)^2; \qquad (3.3.5)$$

Step 4: Set $s = s + 1$ and go to Step 2 (until numerical convergence).

This computation algorithm consists of two iterated steps, 'assignment' as in Step 2 and 'update' as in Step 3. In the 'assignment' step, each cross-section unit i is assigned to the nearest group g_i based on the distance defined in (3.3.4). In the 'assignment' step, to re-allocate centres of groups $\{g_i\}_{i=1}^n$, we compute $\{\widehat{c}_i\}_{i=1}^n$ (or $\left\{\widehat{\overline{c}}_{g_i}\right\}_{i=1}^n$ equivalently) by minimizing (3.3.5).

When conducting the numerical simulations, we assume the value of γ is known, and report the performance of the two-stage algorithm based on this prior information. As γ is assumed to be homogeneous across individuals, the specific value of γ will not disturb the performance of the classifier.

Modified k-means algorithm for mixed-root panels

If (3.2.1) incorporates explosive, stationary, and unit roots, the recursive k-means algorithm fails. Because of heterogeneity in adjustment rates, the sample moments of stationary individuals are asymptotically unstable. To accommodate the mixed roots, we follow the clustering approach in Lin and Ng (2012). We summarize the algorithm in the following steps.

Step 1: We derive LS estimates $\{\widehat{c}_i\}_{i=1}^n$ as

$$\widehat{c}_i - c_i^0 = T^\gamma \frac{\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}}{\sum_{t=1}^T \widetilde{y}_{i,t-1}^2} = T^\gamma \frac{\sum_{t=1}^T \left(y_{i,t-1} - \overline{y}_{i,-1}\right)\left(u_{it} - \overline{u}_i\right)}{\sum_{t=1}^T \left(y_{i,t-1} - \overline{y}_{i,-1}\right)^2}; \qquad (3.3.6)$$

Step 2: To recover latent memberships, we apply the k-means cluster algorithm for $\{\widehat{c}_i\}_{i=1}^n$. Specifically, with $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{K^0}) \in \mathcal{A}$ being any arbitrary $K^0 \times 1$ vector for $\alpha_1, \alpha_2, ..., \alpha_{K^0}$, we define

$$\widehat{Q}_n(\alpha) = \frac{1}{n}\sum_{i=1}^n \min_{1\le l\le k} \left(\widehat{c}_i - \alpha_l\right)^2,$$

and $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2, ..., \widehat{\alpha}_{K^0})$ with $\widehat{\alpha} := \arg\min_{\alpha\in\mathcal{A}} \widehat{Q}_n(\alpha)$. Therefore, we further compute the estimated cluster identity as

$$\widehat{g}_i = \arg\min_{1\le l\le K^0} \left|\widehat{c}_i - \widehat{\alpha}_l\right|,$$

where if there are multiple l's that achieve the minimum, $\widehat{g}_i$ takes the value of the smallest one.

When $c_i^0 > 0$, we have $\widehat{c}_i - c_i^0 = O_p\left(\frac{1}{\rho_i^T}\right)$. When $c_i^0 = 0$, we have $\widehat{c}_i - c_i^0 = O_p\left(\frac{1}{T}\right)$. When $c_i^0 < 0$, we have $\widehat{c}_i - c_i^0 = O_p\left(\frac{1}{T^{\frac{1-\gamma}{2}}}\right)$. However, the pointwise convergence rate of each i is insufficient for showing classification consistency. Instead, we need to verify the uniform convergence rate for $\widehat{c}_i$ over $i = 1, 2, ..., n$.

3.3.2 Stage 2: post-classification estimation, bubble detection and bubble timestamping

Based on classifications, we consider two pooled LS estimators for α^0 ($\overline{c}^0$ or c^0, equivalently), namely, the oracle estimator and the post-classification estimator. The oracle within estimator for the AR coefficient in the g^{th} group is

$$\widehat{\overline{\rho}}_g - \overline{\rho}_g^0 = \frac{\sum_{i \in G_g^0} \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{i,t}}{\sum_{i \in G_g^0} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2}. \tag{3.3.7}$$

Similarly, the post-classification within estimator for the g^{th} group is

$$\widehat{\overline{\rho}}_{\widehat{g}} - \overline{\rho}_g^0 = \frac{\sum_{i \in \widehat{G}_g} \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{i,t}}{\sum_{i \in \widehat{G}_g} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2}, \tag{3.3.8}$$

where we define $\left\{\widehat{G}_g\right\}_{g=1}^{K^0}$ as any consistent classification estimates on $\left\{G_g^0\right\}_{g=1}^{K^0}$.

Under the model (3.2.1) with latent memberships, we can estimate σ_g^2 consistently for each $g = 1, 2, ...K^0$. Define

$$\widetilde{\sigma}_{\widehat{g}}^2 = \frac{1}{2n_{\widehat{g}}T} \sum_{i \in \widehat{G}_g} \sum_{t=1}^T \widehat{\widetilde{u}}_{it}^2, \tag{3.3.9}$$

where $\widehat{\widetilde{u}}_{it} = \widetilde{y}_{it} - \widehat{\overline{\rho}}_{\widehat{g}} \widetilde{y}_{i,t-1}$, $n_{\widehat{g}} = \# \widehat{G}_g$, and $\widehat{\overline{\rho}}_g$ is defined by (3.3.8). Since we assume homoskedasticity over $i = 1, 2, ..., n$, we have

$$\widetilde{\sigma}^2 = \frac{1}{2n_{\widehat{g}}T} \sum_{i \in \widehat{G}_g} \sum_{t=1}^T \widehat{\widetilde{u}}_{it}^2, \tag{3.3.10}$$

for any $\widehat{g} = 1, 2, ...K^0$.

Based on (3.3.10), we can establish the inference procedure to justify the accuracy of our post-classification estimator. To test the null hypothesis $\mathcal{H}_0 : \overline{c}_g^0 = c_r$, we propose the following t-statistic:

$$t_{\widehat{g}} = \frac{(\widehat{\overline{\rho}}_{\widehat{g}} - \rho_r)\sqrt{D_{\widehat{g},nT}}}{\widetilde{\sigma}_{\widehat{g}}}, \tag{3.3.11}$$

where $D_{\widehat{g},nT} = \sum_{i \in \widehat{G}_g} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2$. If we let $c_r = 0$, the statistic (3.3.11) can be employed to test the existence of bubbles. To test the null hypothesis $\mathcal{H}_0 : \overline{c}_g^0 = 0$, we choose the following the panel statistic as

$$\widetilde{t}_{\widehat{g}} = \frac{\left(\widehat{\overline{\rho}}_{\widehat{g}} - 1\right)\sqrt{D_{\widehat{g},nT}}}{\widetilde{\sigma}_{\widehat{g}}}. \tag{3.3.12}$$

Under the alternative $\mathcal{H}_1 : \overline{c}_g^0 > 0$, the fact that the statistics (3.3.12) diverge faster than the time-series statistic of Diba and Grossman (1988) illustrates the superiority of our panel approach. Obviously, the statistic $\widetilde{t}_{\widehat{g}}$ of (3.3.12) corresponds to the full-sample statistic and can detect the signal of bubbles. However, the full-sample statistics cannot date the origination of bubbles.

To consistently estimate the origination of explosive subperiod, we propose the following subsample statistic:

$$\widetilde{t}_{\widehat{g}}(r) = \frac{\left(\widehat{\overline{\rho}}_{\widehat{g}}(r) - 1\right)\sqrt{D_{\widehat{g},nT}(r)}}{\widetilde{\sigma}_{\widehat{g}}(r)}, \tag{3.3.13}$$

where $\widehat{\overline{\rho}}_{\widehat{g}}(r)$ is the post-classification within estimator of $\overline{\rho}_g^0$ based on the first $\tau = [Tr]$ observations in the g^{th} estimated group, and $\widetilde{\sigma}_{\widehat{g}}^2(r)$ is the estimator of σ_g^2 based on the first $\tau = [Tr]$ observations in the g^{th} estimated group. The notation $D_{\widehat{g},nT}(r)$ is the sample moment on the first $\tau = [Tr]$ observations in the g^{th} estimated group. The sub-sample statistic (3.3.13) is the foundation for the estimate of the bubble origination date.

Our inference procedure extends the framework of PWY and PSY. In PSY, the recursive approach is proposed to detect the explosive behaviour and date-stamp the origination of bubbles. The regression model used in PSY is

$$\widetilde{y}_t = \alpha + \beta \widetilde{y}_{t-1} + u_t,$$

where u_t is the equation residual. The parameter $\beta = 0$ under the null hypothesis of no bubble and $\beta > 0$ under the alternative hypothesis of bubbles. Therefore, the standard DF test under the null hypothesis $\mathcal{H}_0 : \beta = 0$ is

$$\text{DF} = \frac{\widehat{\beta}}{se\left(\widehat{\beta}\right)},$$

where

$$\begin{aligned}
\widehat{\beta} &= \frac{T \sum_{t=1}^{T} \Delta\widetilde{y}_t \widetilde{y}_{t-1} - \sum_{t=1}^{T} \Delta\widetilde{y}_t \sum_{t=1}^{T} \widetilde{y}_{t-1}}{T \sum_{t=1}^{T} \widetilde{y}_{t-1}^2 - \left(\sum_{t=1}^{T} \widetilde{y}_{t-1}\right)^2}, \\
se\left(\widehat{\beta}\right) &= \widehat{\sigma}_{\widetilde{y}} \left[\sum_{t=1}^{T} \widetilde{y}_{t-1}^2 - \frac{1}{T} \left(\sum_{t=1}^{T} \widetilde{y}_{t-1} \right)^2 \right]^{-\frac{1}{2}}, \\
\widehat{\sigma}_{\widetilde{y}}^2 &= \frac{1}{T} \sum_{t=1}^{T} \left(\Delta\widetilde{y}_t - \widehat{\alpha} - \widehat{\beta}\widetilde{y}_{t-1} \right)^2, \widehat{\alpha} = \frac{1}{T} \left(\Delta\widetilde{y}_t - \widehat{\beta}\widetilde{y}_{t-1} \right).
\end{aligned}$$

The DF statistics obtained from these subsample (starting from r_1 and ending at r_2) regressions are represented in the sequence $\{\text{DF}_{r_1,r_2}\}$. The detection for the existence of bubbles relies on the supreme statistics as,

$$\text{PSY}_r = \sup_{r_1 \in [0, r - r_{\min}], r_2 = r} \{\text{DF}_{r_1,r_2}\},$$

where $\tau_{\min} = [T r_{\min}]$ is the minimum sample size required to initiate the procedure. The origination date of a bubble is defined to be the first observation where the supreme statistic exceeds the diverging critical value as,

$$\widehat{r} = \inf_{s \geq r_{\min}} \{s : \text{PSY}_s > cv_{\beta_{Tn}}\},$$

where $cv_{\beta_{Tn}}$ is the critical value with significance level $\beta_{Tn} \to 0$.

Following PWY, we propose a recursive algorithm to date the bubble origination. Since we only consider a single bubble case, we employ the t-statistic rather than the supreme t-test. Based on the recursive procedure of the panel t-test, we use the classified groups of panels to data-stamp the origination. We date the origination of an explosive episode as

$$\widehat{r}_g^e = \inf_{s \geq r_0} \left\{ s : \widetilde{t}_g(s) > cv_{\beta_{Tn}} \right\}, \tag{3.3.14}$$

and

$$\widehat{r}_{\widehat{g}}^e = \inf_{s \geq r_0} \left\{ s : \widetilde{t}_{\widehat{g}}(s) > cv_{\beta_{Tn}} \right\}, \tag{3.3.15}$$

where $cv_{\beta_{Tn}}$ is the right-side $100\beta_{Tn}\%$ critical value of the limiting distribution of $\widetilde{t}_{\widehat{g}}$ and $\widetilde{t}_g$ statistics based on $\tau_s = [Ts]$ time horizon, and β_{Tn} is the size of the one-sided statistics. The parameter r_0 is the minimum sample size required to initiate the regression. We allow $\beta_{Tn} \to 0$ as $(n, T) \to \infty$ because, in this event, $cv_{\beta_{Tn}} \to \infty$. This recursive method can apply in the same way to the PWY procedure based on the ADF statistic.

3.3.3 Estimation of K^0

To estimate the true number of groups, we rely on the Bayesian information criterion (BIC) which is defined as:

$$\text{BIC}(K) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \widehat{\widehat{\rho}}_{\widehat{g}_i}^{(K)} \right)^2 + \frac{K+n}{nT} \log(nT),$$

where $\left(\widehat{\widehat{\rho}}_{\widehat{g}_i}^{(K)} \right)_{i=1}^{n}$ is the post-classification estimator based on K groups and $\widetilde{\sigma}^2$ is the variance estimator (3.3.10). Post-classification estimators based on both the recursive k-means and the modified k-means are applicable. The estimation of the group number is achieved by choosing the optimal K which minimize the BIC, that is,

$$\widehat{K} = \arg \min_{K=1,2,\ldots,K_{\max}} \text{BIC}(K), \tag{3.3.16}$$

where $K_{\max}$ is a generic upper bound of K. The BIC represents a balance between model fitness and penalty of over-fitness. In addition to the BIC function considered in Bai and Ng (2002), Bai (2003, 2009), the Deviance information criterion (DIC) in Spiegelhalter et al. (2002) and Li et al. (2019) also can be employed to select the true number of groups.

3.4 Asymptotic Theory

In this section, we study the asymptotic properties of the k-means algorithms under the respective panel models. We will show that the classification algorithms can recover the latent group structure consistently. Given this consistency, we can demonstrate that the feasible estimators of the AR coefficients are asymptotically equivalent to the oracle estimators that are derived as if the true group structure was known. We also provide the asymptotic distributions of the estimators of the AR coefficients. In this section, we justify the consistency of explosiveness statistics under the alternative hypothesis of explosive roots. We also show our recursive method can date the origination dates of explosive episodes with better accuracy than PWY. At last, we demonstrate that the BIC estimator on K^0 is consistent.

To demonstrate the asymptotic theory of our two-stage procedure, we impose two assumptions.

Assumption 1 *(i) For each i, the individual fixed effect $\mu_i = O\left(T^{-1}\right)$ or $\mu_i = 0$.*

(ii) The error process $\{u_{it}\}$ is a martingale difference sequence with a homogeneous conditional second moment σ^2 ($\mathbb{E}\left(u_{it}^2|\mathcal{F}_{i,t-1}\right) = \sigma^2$ for all i and t, where $\mathcal{F}_{i,t-1} := \sigma\left\{u_{i,t-1}, u_{i,t-2}, ...\right\}$) and finite q^{th} moments with $q \geq 4$ for all i and t.

(iii) Initial conditions: Assume that $y_{i0} = 0$ almost surely for all i, y_{i0} is independent of u_{it} for all i and t, and $u_{is} = 0$ for all $s \leq 0$.

(iv) There exist c_u and c_l such that for each $i \in \{1, 2, ..., n\}$ so that we have $0 < c_l \leq |c_i^0| \leq c_u < \infty$ or $c_i^0 = 0$.

(v) There exists a constant $c^* \in (0, \infty)$ such that $\inf_{1 \leq g \leq g' \leq K^0} \left|\alpha_g^0 - \alpha_{g'}^0\right| \geq c^*$.

Assumption 1(i) provides restrictions on individual fixed effects. Assumption 1(ii) assumes the martingale property for innovations. Assumption 1(iii) imposes restrictions on the initial conditions. Assumption 1(iv) imposes an identification condition for distance parameter. Assumption 1(v) gives another identification condition where the group-specific parameters are well separated from each other.

Define $\widetilde{\sigma}^2_{G(K)} := \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t} - \widehat{\overline{\rho}}^{(K)}_{\widehat{g}_i^{(K)}}\widetilde{y}_{i,t-1}\right)^2$ where $\widehat{g}_i^{(K)}$ and $\widehat{\overline{g}}^{(K)}_{\widehat{g}_i^{(K)}}$ are membership and slope estimates assuming the number of group is K.

Assumption 2 *(i) Let $\{n_g\}_{g=1}^{K^0}$ denote the cardinality of latent groups. For each $g \in \{1, 2, ..., K^0\}$, $\frac{n_g}{n} \to \pi_g < \infty$. Moreover, $\inf_{1\leq g\leq K^0}\pi_g \geq \underline{M}$ for some $\underline{M} > 0$.*

(ii) The following rate restrictions hold: $T^{1-3\gamma}n(\log n)^2 \to 0$, $T^{2\gamma-2}n(\log n)^2(\log_2 T)^4 \to 0$, and $T^{\frac{5\gamma-3}{4}}n(\log n)^2 \to 0$, where $\log_2(\cdot) := \log\log(\cdot)$.

(iii) The relationship $\delta_{nT} < \frac{M(c^*)^2}{15c_u}$ holds for each n and T, where δ_{nT} is defined in Lemma A.2.11.

(iv) As $(n, T) \to \infty$, $\min_{1\leq K<K^0}\inf_{\delta_K\in\Delta_K}\widetilde{\sigma}^2_{G(K)} \xrightarrow{p} \underline{\sigma}^2 > 2\sigma^2$, while $\widetilde{\sigma}^2_{G(K)}$ is defined above.

Assumption 2(i) implies that each group size increases proportionally to the dimension of individuals. Assumption 2(ii) imposes rate restrictions so that the concentration inequality is applicable. Assumption 2(iii) provides necessary conditions for the uniform consistency of classifications. Assumption 2(iv) imposes necessary conditions for the consistency of BIC procedure.

We find that incidental parameters do not show up in the asymptotics of sample moments. The reason why this happens is that for each $i = 1, 2, ..., n$, the fixed effect is asymptotically diminishing and dominated by noise. To illustrate this point clearly, note that,

$$y_{it} = \frac{\rho_i^t - 1}{\rho_i - 1}\mu_i + \sum_{s=1}^{t}\rho_i^{t-s}u_{is}. \tag{3.4.1}$$

Since $\frac{\rho_i^t-1}{\rho_i-1}\mu_i = O\left(\rho_i^t T^{\gamma-1}\right)$ and $\sum_{s=1}^{t}\rho_i^{t-s}u_{is} = O_p\left(\rho_i^t T^{\gamma}\right)$, y_{it} is asymptotically equivalent to $\sum_{s=1}^{t}\rho_i^{t-s}u_{is}$ for each $i = 1, 2, ..., n$. Moreover, noise, not fixed effects, determines the asymptotics uniformly for all $i = 1, 2, ..., n$. Therefore, to determine uniform bounds of sample moments, we only need to consider innovations. Without losing generality, we assume $\mu_i = 0$ when deriving asymptotics for the two-stage algorithm.

3.4.1 Recursive k-means algorithm

We establish the consistency of the coefficient estimates in the explosive panel autoregressions.

Theorem 3.4.1 *(Individual Consistency of Classification) Let Assumption 1 and 2 hold. Assume $c_i^0 > 0$ for each $i = 1, 2, ..., n$. With joint convergence $(n, T) \to \infty$,*

$$\Pr\left(\max_{1\leq i\leq n} \left|\widehat{g}_i - g_i^0\right| > 0\right) = o(1).$$

Theorem 3.4.1 justifies the consistency of recursive k-means classification algorithm. It is similar to Theorem 2 of Bonhonmme and Manresa (2015). This theorem states that under the joint convergence framework $(n, T) \to \infty$, we can correctly recover the true group structure. From the Theorem 3.4.1, we observe that correct classifications strongly rely on Assumption 1(ii). In our discussion, as long as the distance parameters $\left\{\overline{c}_g^0\right\}_{g=1}^{K^0}$ are separated across groups, the misclassification errors are asymptotically negligible.

The following theorem indicates that the post-classification estimator $\widehat{\alpha}_{\widehat{g}}$ is asymptotically equivalent to the oracle estimator $\widehat{\alpha}_g$ for each $g \in \{1, 2, ...K^0\}$. With the classification consistency of the recursive k-means algorithm, we can verify the asymptotic equivalence between our parameter estimate and the true value:

$$\sqrt{n_g}\left(\overline{\rho}_g^0\right)^T\left(\widehat{\alpha}_{\widehat{g}} - \alpha_g^0\right) = \sqrt{n_g}\left(\overline{\rho}_g^0\right)^T\left(\widehat{\alpha}_g - \alpha_g^0\right) + o_p(1).$$

Therefore, the post-classification estimator $\widehat{\alpha}_{\widehat{g}}$ shares the identical limiting distribution as the oracle estimator. The following theorem shows the limiting distribution of $\widehat{\alpha}_{\widehat{g}}$.

Theorem 3.4.2 *Suppose Assumption 1 and 2 hold. Assume $c_i^0 > 0$ for $i = 1, 2, ..., n$. Under joint convergence $(n, T) \to \infty$ and rate restriction $\frac{n}{T^{2-2\gamma}} = o(1)$,*

$$\sqrt{n_g} \left(\bar{\rho}_g^0\right)^T \left(\widehat{\alpha}_{\widehat{g}} - \alpha_g^0\right) \xrightarrow{d} \mathcal{N}(0, 4\left(\alpha_g^0\right)^2).$$

Remark 3.4.1 *To test for explosive roots, we need to accommodate both the grouped explosive roots and the unit root. The new data generating mechanism follows (3.2.1) and the group structure as*

$$\begin{cases} \textit{Explosive Groups:} & \begin{cases} \textit{Group } 1: & \alpha_1^0 > 0, \\ \textit{Group } 2: & \alpha_2^0 > 0, \\ \vdots & \vdots \\ \textit{Group } (K^0 - 1): & \alpha_{K^0-1}^0 > 0, \end{cases} \\ \textit{Unit Root Group:} & \textit{Group } K^0: \quad \alpha_{K^0}^0 = 0. \end{cases} .$$

To apply the recursive k-means approach, we can modify the objective function (3.3.1) with the self-normalized adjustment rate as

$$(\widehat{\bar{c}}, \widehat{\delta}) = \arg \min_{(\bar{c},\delta)\in\mathcal{C}\times\Delta_{K^0}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\Upsilon_{iT}} \sum_{t=1}^{T} \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \exp\left(\frac{\bar{c}_{g_i}}{T^\gamma}\right)\right)^2, \qquad (3.4.2)$$

where $\Upsilon_{iT} := \sum_{t=1}^{T} y_{i,t-1}^2$. Basically, the optimization of (3.4.2) follows the identical algorithm in (3.3.2) and (3.3.3) except that we just need to replace $\breve{\rho}_i^{2T} T^{2\gamma}$ by Υ_{iT}. Due to the assumption of latent group structures, self-normalization rates do not disturb our theoretical derivations and codings. Therefore, the consistency of classifications and the oracle properties of parameter estimates still hold. For the technical details, kindly check the proof of Theorem 4.1.

3.4.2 Modified k-means algorithm

We now establish the classification consistency of the modified k-means algorithm in the mixed-root panel model.

Theorem 3.4.3 *(Individual Consistency of Classification) Suppose Assumptions 1 and 2 hold. When* $(n, T) \to \infty$*,*

$$\sup_{1 \leq i \leq n} \mathbf{1} \left\{ \widehat{g}_i \neq g_i^0 \right\} = o_p (1) .$$

Theorem 3.4.3 shows that the modified k-means algorithm consistently recovers latent memberships for the mixed-root panel model. This machinery incorporates more general cases than the recursive classification algorithm. The following theorem reports the asymptotic distribution of $\widehat{\alpha}_{\widehat{g}}$, the parameter estimate based on the estimated membership.

Theorem 3.4.4 *Suppose Assumptions 1 and 2 hold. Assume* $\frac{n}{T^{2-2\gamma}} = o(1)$*. When* $(n, T) \to \infty$*,*

$$\sqrt{n_g} \left(\overline{\rho}_g^0 \right)^T \left(\widehat{\alpha}_{\widehat{g}} - \alpha_g^0 \right) \xrightarrow{d} \mathcal{N}(0, 4 \left(\alpha_g^0 \right)^2).$$

Based on the modified k-means algorithm, we also obtain the oracle property of the post-classification estimator. Therefore, we can develop reliable tests for explosiveness in the context of mixed-root panels. The significant advantage of the post-classification estimates is that they employ both time-series and cross-sectional asymptotics.

Remark 3.4.2 *The modified k-means algorithm can also consistently identify the* *latent membership with structural breaks in the following model.*

$$\begin{cases} \textit{Stationary-to-Explosive Groups:} & \begin{cases} \textit{Group } 1: & \alpha^0_{1,1} < 0 & \alpha^0_{1,2} > 0 \\ \textit{Group } 2: & \alpha^0_{1,2} < 0 & \alpha^0_{2,2} > 0 \\ \vdots & \vdots & \vdots \\ \textit{Group } k_1: & \alpha^0_{1,k_1} < 0 & \alpha^0_{2,k_1} > 0 \end{cases} \\ \textit{Unity-to-Explosive Groups:} & \begin{cases} \textit{Group } (k_1+1): & \alpha^0_{1,k_1+1} = 0 & \alpha^0_{1,k_1+1} > 0 \\ \textit{Group } (k_1+2): & \alpha^0_{1,k_1+2} = 0 & \alpha^0_{2,k_1+2} > 0 \\ \vdots & \vdots & \vdots \\ \textit{Group } k_2: & \alpha^0_{1,k_2} = 0 & \alpha^0_{2,k_2} > 0 \end{cases} \\ \textit{Unit Root Group:} & \textit{Group } (k_2+1): \ \alpha^0_{k_2+1} = 0 \\ \textit{Stationary Groups:} & \begin{cases} \textit{Group } (k_2+2): & \alpha^0_{k_2+2} < 0 \\ \textit{Group } (k_2+3): & \alpha^0_{k_2+3} < 0 \\ \vdots & \vdots \\ \textit{Group } K^0: & \alpha^0_{K^0} < 0 \end{cases} \end{cases} . \tag{3.4.3}$$

There are four types of latent groups in the panel: Case (1), Stationary-to-Explosive Groups (Before the time point τ^e_g*,* $\alpha^0_{1,g} < 0$*; After the time point* τ^e_g*,* $\alpha^0_{2,g} > 0$*); Case (2), Unity-to-Explosive Groups (Before the time point* τ^e_g*,* $\alpha^0_{1,g} = 0$*; After the time point* τ^e_g*,* $\alpha^0_{2,g} > 0$*); Case (3), Unit Root Group (*$\alpha^0_g = 0$*); Case (4), Stationary Groups (*$\alpha^0_g < 0$*). For both Case (1) and (2), when* $i = 1, 2, ..., n$ *and* $r^e_{g^0_i} < 1$*, the time-series LS estimate* $\widehat{c}_i$ *is convergent to* α^0_{2,g^0_i}*. This fact shows that the signal of the explosive subperiods dominates those of unit roots and stationarity. For technical details, kindly check the proof of the Theorem 3.4.7. As the validity of the modified* k*-means algorithm relies on the consistency of LS estimates, we can still recover the latent membership with structural breaks in slopes. The classification consistency for the parameter instability lays down the foundation for bubble analysis in the forthcoming discussions.*

3.4.3 Test statistic for bubbles

We provide the consistency of both (3.3.9) and (3.3.10) in the following lemma.

Lemma 3.4.1 *Suppose Assumptions 1 and 2 hold. When $(n,T) \to \infty$,*

$$\widetilde{\sigma}^2_{\widehat{g}} \xrightarrow{p} \sigma^2,$$

for any $g = 1, 2, ..., K^0$.

Consistency of $\widetilde{\sigma}^2_{\widehat{g}}$ is essential for valid inference procedures as we can properly scale the statistic. We collect the details of (3.3.11) in the following theorem.

Theorem 3.4.5 *Suppose Assumptions 1 and 2 hold. Under $\mathcal{H}_0 : \overline{c}^0_g = 0$,*

$$\widetilde{t}_{\widehat{g}} \xrightarrow{d} \mathcal{N}(0,1).$$

Under $\mathcal{H}_1 : \overline{c}^0_g > 0$,

$$\widetilde{t}_{\widehat{g}} = O_p\left(\left(\overline{\rho}^0_g\right)^T \sqrt{n}\right),$$

where $(n,T) \to \infty$.

By comparison, the DF statistic under the alternative hypothesis diverges at the rate $O_p\left((\overline{\rho}^0_g)^T\right)$. The divergence rate is slower than that in (3.3.12), leading to a lower power than the panel-based test.

3.4.4 Estimate of bubble origination date

When $\tau = [Tr] \to \infty$ for all $r \in [r_0, 1]$, we have

$$\widetilde{t}_{\widehat{g}}\,(r) \xrightarrow{d} \mathcal{N}\,(0,1)\,,$$

under the null hypothesis of unit root in the g^{th} group. For a bubble to have a meaningful origination date, we assume the following DGP for at least one group (say G_g^0) and all $i \in G_g^0$,

$$y_{it} = \mu_i + \rho_{it} y_{i,t-1} + u_{it},\ t = 1, 2, ..., T,\ \text{and } i = 1, 2, ..., n, \tag{3.4.4}$$

where u_{it} is a martingale difference sequence with a conditional 2^{nd} moment σ^2 and a finite q^{th} moment with $q \geq 4$. The initialization of the process is $y_{i0} = 0$ for $i = 1, 2, ..., n$. The fixed effect is $\mu_i = O_p\left(\frac{1}{T}\right)$. The AR coefficient $\rho_{it} = 1 + \frac{c_{it}}{T^\gamma}$ with time-varying c_{it}. Following PWY and Phillips and Yu (2009), we assume

$$c_{it} = c_{1i}^0 \mathbf{1}\left\{t < \tau_{g_i^0}^e\right\} + c_{2i}^0 \mathbf{1}\left\{t \geq \tau_{g_i^0}^e\right\}, \text{ with } c_{1i}^0 \leq 0 \text{ and } c_{2i}^0 > 0. \tag{3.4.5}$$

Hence, model (3.4.4) allows for two regimes: a unit root or stationary regime and an explosive regime with a bubble originating at $\tau_{g_i^0}^e$. If individual i does not contain any explosive root, then $r_{g_i^0}^e = 1$.

By Remark 3.4.2, under the case of parameter instability, we can still apply the modified k-means algorithm to consistently recover the true membership, showing that the estimated groups are equivalent to the true ones. Therefore we can directly employ the feasible estimator on the origination date $\widehat{r}_{\widehat{g}}^e$ rather than $\widehat{r}_g^e$. Based on $\widehat{r}_g^e$ and $\widehat{r}_{\widehat{g}}^e$, we establish a limit theory for dating the origination of an explosive root under the case of no bubbles.

Theorem 3.4.6 *Suppose Assumptions 1 and 2 hold. Assume $n/T^{2-2\gamma} = o(1)$. Under the null hypothesis of no episode of explosiveness ($\alpha_g^0 \leq 0$ for each $g = 1, 2, ..., K^0$), and provided that $cv_{\beta_{Tn}} \to \infty$, the probability of detecting the origination of a bubble using $\widetilde{t}_g$ and $\widehat{t}_{\widehat{g}}$ is zero: As $(n, T) \to \infty$,*

$$\Pr\left(\widehat{r}_g^e \in [0, 1]\right) \to 0, \textit{ and } \Pr\left(\widehat{r}_{\widehat{g}}^e \in [0, 1]\right) \to 0,$$

where $\widehat{r}_g^e$ is defined in (3.3.14) and $\widehat{r}_{\widehat{g}}^e$ in (3.3.15).

Next, we show the consistency of the estimator under the case of a single bubble for $g = 1, 2, ..., K^0$.

Theorem 3.4.7 *Suppose Assumptions 1 and 2 hold. Assume* $n/T^{2-2\gamma} = o(1)$*. If* $\frac{1}{cv_{\beta_{Tn}}} + \frac{cv_{\beta_{Tn}}}{P_{Tn}} \to 0$ *with* $P_{Tn} = \sqrt{n}T^{\frac{2-\gamma}{2}}$*, under the model (3.4.4) and (3.4.5),*

$$\widehat{r}_g^e \xrightarrow{p} r_g^e, \text{ and } \widehat{r}_{\widehat{g}}^e \xrightarrow{p} r_g^e.$$

When we include observations from the explosive regimes, the signals from explosive roots dominate those from non-explosive regimes (unit root and stationary regimes). In this case, we can show that the statistics diverge at the rate of $O_p(P_{Tn})$. When the critical value $cv_{\beta_{Tn}}$ increases at a slower speed than P_{Tn}, we can obtain the consistency of the origination estimate.

Remark 3.4.3 *From practical implementation, we set the critical value sequences* $v_{\beta_{Tn}}$ *according to a rule of thumb such as* $cv_{\beta_{Tn}} = \frac{2}{3}\log\log(nT)$*. The critical value diverges at a slower speed than* P_{Tn}*.*

3.4.5 Estimate of K^0

We show the consistency of the estimator on K^0 using BIC.

Theorem 3.4.8 *Suppose Assumptions 1 and 2 hold. When* $(n, T) \to \infty$*,*

$$\widehat{K} \xrightarrow{p} K^0$$

where $\widehat{K}$ *is defined in (3.3.16).*

Remark 3.4.4 *In the above analysis, we derive all the results, assuming that the number of groups* K^0 *is known. In practice, the researcher has to determine* K^0 *from data before conducting the two-stage procedure.*

3.5 Monte Carlo Studies

We design several Monte Carlo experiments to check the finite sample performance of the proposed two-stage algorithm.[1] First, we demonstrate the behaviours of the recursive k-means algorithm and the modified k-means classification algorithm. The number of replications is always set to 1,000.

3.5.1 Recursive k-means algorithm

To verify the classification accuracy of the recursive k-means classification, we consider only a DGP on explosive roots. We simulate $\mu_i \sim \frac{\mathcal{N}(0,1)}{T}$ and $u_{it} \overset{iid}{\sim} \mathcal{N}(0,2)$ over i and t. Data are simulated from a DGP with three groups ($K^0 = 3$) with $n_1 : n_2 : n_3 = \frac{1}{3} : \frac{1}{3} : \frac{1}{3}$. The following settings are considered (DGP 1): $n =$ 30, 60, $T =$ 25, 50, 100, $\bar{c} = (0.2, 0.9, 1.6)$, $\gamma = 0.5$.

With correctly selected group numbers, we show the classification consistency of the recursive k-means procedure and provide the numerical results of its post-classification estimator. Tables 3.3 shows the classification errors, RMSE, asymptotic bias, and probability coverage of the post-classification estimator and its oracle estimator, the infeasible within estimator when informed of the true group membership. As a comparison, we show results for the oracle within estimator. We only replace $\widehat{G}_g$ with G_g^0, since we can observe the true membership.

Tables 3.3 shows extensive discussions on explosive roots. As shown here, the classification error approaches zero as the time horizon increases, and the RMSE and bias of the oracle estimator are smaller than the post-classification estimator. For post-classification estimators, the RMSE and bias generally decrease when $T \to \infty$. When $T \geq 100$, the asymptotic difference between the oracle estimator and the post-classification estimator is almost negligible. The diminishing distance verifies the asymptotic equivalence between these two estimators. The diminishing distance

[1]The experiment of chapter uses the codes provided by Professor Stéphane Bonhomme on his website: https://sites.google.com/site/stephanebonhommeresearch/. This declaration is to honor the contribution of Professor Stéphane Bonhomme to this chapter and the originality of his invention on the classification algorithm.

Table 3.3: Classification and Estimation by Recursive k-means Algorithm

	DGP 1			Post-Classification			Oracle		
	n	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	30	25	0.0436	0.176	-0.0081	0.874	0.1031	-0.0363	0.8663
	30	50	0.0081	0.0886	-0.0151	0.938	0.0458	-0.0149	0.9338
	30	100	0.0018	0.1162	-0.0157	0.93	0.024	-0.0069	0.9299
	60	25	0.0451	0.1616	0.0065	0.886	0.0697	-0.024	0.8743
	60	50	0.0073	0.0574	-0.006	0.954	0.0269	-0.009	0.9485
	60	100	0.0006	0.0665	-0.0073	0.97	0.0143	-0.0044	0.968
Group 2	30	25	0.0436	0.1515	-0.0286	0.934	0.0095	-0.0022	0.9326
	30	50	0.0081	0.0505	-0.0033	0.94	0.0011	-0.0001	0.9439
	30	100	0.0018	0.0001	0	0.94	0.0001	0	0.94
	60	25	0.0451	0.1528	-0.0298	0.914	0.0068	-0.0021	0.9121
	60	50	0.0073	0.0516	-0.0034	0.936	0.0008	-0.0001	0.9349
	60	100	0.0006	0	0	0.944	0	0	0.944
Group 3	30	25	0.0436	0.0005	-0.0001	0.954	0.0005	-0.0001	0.954
	30	50	0.0081	0	0	0.954	0	0	0.954
	30	100	0.0018	0	0	0.952	0	0	0.952
	60	25	0.0451	0.0004	-0.0001	0.924	0.0004	-0.0001	0.924
	60	50	0.0073	0	0	0.944	0	0	0.944
	60	100	0.0006	0	0	0.954	0	0	0.954

is due to the uniform consistency of our recursive k-means algorithm. The recursive k-means induces much better finite sample performance than the modified k-means algorithm. The main explanation for this phenomenon is that once we have larger $n,$ the uniform convergence rate of individual least squares estimators is reduced.

Last, we evaluate the performance of explosive detection statistic. The nominal level is set to 5%. We evaluate these tests with the correct number of groups of $K^0 = 3$. To demonstrate the superiority over the right-tailed DF test, we choose $n = 30$, 60, 90, and $T = 25$, 50, 100. With three groups, we assume $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$. We present the simulated results in Tables 3.4.

The sizes of panel explosiveness tests are well controlled around nominal levels. The power of the test is greatly improved by the additional degree of cross-sectional asymptotics, n, when compared with the DF statistic. The power improvement corresponds to our aim to create power-enhanced statistic for explosiveness detection. Based on the panel t-statistics, we can construct a recursive real-time algorithm, like PWY.

Table 3.4: Sizes and Powers of DF and Panel Tests on Explosive Roots Detection (Recursive k-means Algorithm)

n	Statistics		Size	
	$\bar{c}_1$=0($\bar{c}_2$=1.5, $\bar{c}_3$=3),γ=0.5	T=25	T=50	T=100
1	DF	0.07	0.054	0.068
10(30)	Post-Classification	0.048	0.044	0.054
20(60)	Post-Classification	0.068	0.52	0.048
30(90)	Post-Classification	0.066	0.056	0.05
n	Statistics		Power	
	$\bar{c}_1$=0.3($\bar{c}_2 = 1.5$, $\bar{c}_3 = 3$),γ=0.5	T=25	T=50	T=100
1	DF	0.236	0.384	0.514
10(30)	Post-Classification	0.84	0.978	0.998
20(60)	Post-Classification	0.984	1	1
30(90)	Post-Classification	0.994	1	1
n	Statistics		Power	
	$\bar{c}_1$=0.5($\bar{c}_2$=1.5, $\bar{c}_3$=3),γ=0.5	T=25	T=50	T=100
1	DF	0.492	0.63	0.816
10(30)	Post-Classification	0.998	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1

3.5.2 Modified k-means algorithm

To verify the classification accuracy of the modified k-means algorithm, we consider a DGP of both explosive and stationary roots. We simulate $\mu_i \sim \frac{\mathcal{N}(0,1)}{T}$ and $u_i \overset{iid}{\sim} \mathcal{N}(0,2)$ over i and t. Data are simulated from a DGP with three groups ($K^0 = 3$) of $n_1 : n_2 : n_3 = \frac{1}{3} : \frac{1}{3} : \frac{1}{3}$. The following settings are considered (DGP 2): $n = 30$, 60, $T = 25$, 50, 100, $\bar{c} = (-1, 0.5, 1.5)$, $\gamma = 0.5$.

With correctly selected group numbers, we show the classification consistency of the modified k-means procedure and provide the numerical results of the post-classification estimator based on the modified k-means procedure. Tables 3.5 shows the classification errors, RMSE, asymptotic bias, and probability coverage of the post-classification estimator and its oracle estimator.

Tables 3.5 provides results on mixed-root cases, in which both explosive and stationary roots appear. As illustrated in the numerical simulations, the classification errors approach zero as the time horizon increases, and the RMSE and bias of the oracle estimator are much smaller than those of the post-classification estimator based on the modified k-means algorithm. For post-classification estimators, the RMSE and bias generally decrease and get closer to the RMSE and bias of the ora-

Table 3.5: Classification and Estimation by Modified k-means Algorithm

	DGP 2			Post-Classification			Oracle		
	n	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	30	25	0.113	0.5056	-0.0372	0.838	0.2378	-0.0312	0.8091
	30	50	0.0944	0.3712	-0.028	0.878	0.1991	-0.0284	0.8692
	30	100	0.0841	0.2867	-0.0197	0.874	0.1769	-0.0026	0.8797
	60	25	0.1288	0,435	-0.0059	0.766	0.16	-0.0049	0.7557
	60	50	0.1183	0.3493	-0.021	0.806	0.1382	-0.0173	0.8099
	60	100	0.0935	0.2653	-0.0204	0.834	0.1284	-0.0048	0.8523
Group 2	30	25	0.113	1.3086	0.1195	0.764	0.0138	-0.0021	0.7342
	30	50	0.0944	1.3991	0.1288	0.768	0.0029	-0.0002	0.7814
	30	100	0.0841	1.0227	0.213	0.792	0.001	-0.0001	0.8109
	60	25	0,1288	1.1999	0.2093	0.723	0.0094	-0.0012	0.7117
	60	50	0.1183	1.3589	0.2134	0.742	0.0019	-0.0001	0.7343
	60	100	0.0935	0.87	0.2523	0.784	0.0006	0	0.7952
Group 3	30	25	0.113	4.0159	-1.7779	0.942	0	0	0.942
	30	50	0.0944	5.1474	-2.208	0.95	0	0	0.95
	30	100	0.0841	5.9771	-2.5076	0.952	0	0	0.952
	60	25	0.1288	4.3123	-2.0501	0.948	0	0	0.948
	60	50	0.1183	5.755	-2.76	0.976	0	0	0.976
	60	100	0.0935	6.1775	-2.6785	0.964	0	0	0.964

cle estimators as $T \to \infty$. The decreasing RMSE and bias demonstrate that as the time horizon diverges, the asymptotic difference between the post-classification and oracle estimators is asymptotically diminishing. The diminishing distance is due to the uniform consistency of the modified k-means classification technology. Unfortunately, since classification errors of the modified k-means algorithm diminish at a slower rate than the recursive k-means algorithm, the RMSE and bias of post-classification estimators are reduced more slowly. Therefore the misclassification errors contribute to the low probability coverages in Group 1 and 2.

Next, we investigate the performance of the panel t-statistics to detect the existence of explosive roots. The significance level is set to 5%. We evaluate these tests with the correct number of groups $K^0 = 3$. We examine the performance of the proposed tests under a mixture case of both explosive and stationary roots. The sample sizes over the cross-sectional dimension and time horizon are chosen as $n = 30$, 60, 90, and $T = 25$, 50, 100, respectively. With three groups, we assume $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$. We present the detailed results in Tables 3.6.

The sizes of the proposed panel t-test are well controlled under mixed roots. As illustrated here, empirical rejection frequency under the null hypothesis is very

Table 3.6: Sizes and Powers of DF and Panel Tests on Explosive Roots Detection (Modified k-means Algorithm)

n	Statistics		Size	
$\bar{c}_1 = 0(\bar{c}_2 = -1, \bar{c}_3 = 1.5)$, γ=0.5		T=25	T=50	T=100
1	DF	0.066	0.054	0.044
10(30)	Post-Classification	0.074	0.064	0.068
20(60)	Post-Classification	0.68	0.058	0.06
30(90)	Post-Classification	0.61	0.05	0.056
n	Statistics		Power	
$\bar{c}_1 = 0.3(\bar{c}_2 = -1, \bar{c}_3 = 1.5)$, γ=0.5		T=25	T=50	T=100
1	DF	0.33	0.566	0.774
10(30)	Post-Classification	0.966	0.996	1
20(60)	Post-Classification	0.998	1	1
30(90)	Post-Classification	1	1	1
n	Statistics		Power	
$\bar{c}_1 = 0.5(\bar{c}_2 = -1, \bar{c}_3 = 1.5)$, γ=0.5		T=25	T=50	T=100
1	DF	0.64	0.828	0.918
10(30)	Post-Classification	0.998	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1

close to the nominal level in the finite sample. The most interesting finding is the significant power improvement brought by an additional degree of cross-sectional asymptotics. For example, in Tables 3.6, under the alternative hypothesis of the distance parameter as small as 0.3, the rejection rate of the panel t-test is almost unity when $n \geq 50$, $T \geq 25$. However, the counterpart, the DF test, can detect explosive roots only by a 50% chance. The loss of powers shows the high priority of the panel t-test on explosiveness detections.

3.5.3 Detection of bubble origination

This subsection reports some brief simulations examining the performance of the dating procedure and the accuracy of its asymptotic theory. We employ the panel recursive dating estimator $\widehat{r}^e_{\widehat{g}}$ based on $\widetilde{t}_{\widehat{g}}$ for $\widehat{g} = 1, 2, ..., K^0$. We also conduct the time-series recursive estimate $\widehat{r}^e$ proposed in PWY. We simulate $\mu_i \sim \frac{\mathcal{N}(0,1)}{T}$ and $u_i \overset{iid}{\sim} \mathcal{N}(0, 2)$ over i and t. Data are simulated from a DGP with three groups ($K^0 = 3$) of $n_1 : n_2 : n_3 = \frac{1}{3} : \frac{1}{3} : \frac{1}{3}$. We set the sample size as $n = 30$ and $T = 50$, 100, 150. The following settings are considered: $\bar{c}_1 = 0$, $\bar{c}_2 = -1$, $\bar{c}_3 = -2.5$, $\gamma = 0.5$ when $r \leq r^e_g\,(:= 0.5)$; $\bar{c}_1 = 0.2, 0.4, 0.6, 0.8$, $\bar{c}_2 = -1$, $\bar{c}_3 = -2.5$, $\gamma = 0.5$ when $r > r^e_g\,(:= 0.5)$. We use critical values as $\log(nT)$ for the panel recursive algorithm and $\log(T)$ for PWY procedure.

Table 3.7: Results on Estimations of Bubble Origination Date Based on Panel Recursive Algorithm and Time-Series Recursive Algorithm

$(\bar{c}_2 = -1, \bar{c}_3 = -2.5)$	$\bar{c}_1 = 0.2$		$\bar{c}_1 = 0.4$		$\bar{c}_1 = 0.6$		$\bar{c}_1 = 0.8$	
N=30, T=50								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5385	0.5929	0.5315	0.5746	0.5261	0.5644	0.5229	0.5565
Std	0.0088	0.0567	0.0071	0.0419	0.0066	0.0372	0.006	0.034
RMSE	0.6834	1.8841	0.5594	1.4822	0.4667	1.2879	0.4096	1.1414
$(\bar{c}_2 = -1, \bar{c}_3 = -2.5)$	$\bar{c}_1 = 0.2$		$\bar{c}_1 = 0.4$		$\bar{c}_1 = 0.6$		$\bar{c}_1 = 0.8$	
N=30, T=100								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5311	0.5745	0.5251	0.5601	0.521	0.551	0.5185	0.5444
Std	0.0065	0.0406	0.0055	0.032	0.0047	0.0264	0.0046	0.0231
RMSE	0.5504	1.4684	0.4445	1.1786	0.3725	0.9945	0.3305	0.8676
$(\bar{c}_2 = -1, \bar{c}_3 = -2.5)$	$\bar{c}_1 = 0.2$		$\bar{c}_1 = 0.4$		$\bar{c}_1 = 0.6$		$\bar{c}_1 = 0.8$	
N=30, T=150								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5247	0.5599	0.52	0.5479	0.5168	0.5403	0.5146	0.5352
Std	0.0045	0.037	0.0038	0.0283	0.0033	0.0244	0.0029	0.0202
RMSE	0.4348	1.2184	0.3521	0.9629	0.2968	0.8158	0.2577	0.7022

Tables 3.7 report results for both the panel recursive algorithm and the time-series recursive algorithm, giving means, standard errors, and RMSE for $\widehat{r}^e_{\widehat{g}}$. We can observe the following four patterns. First, the panel recursive estimate $\widehat{r}^e_{\widehat{g}}$ can estimate the true bubble origination date with high accuracy, reflected by a small bias and standard error. When $T = 150$, for cases, the mean of $\widehat{r}^e_{\widehat{g}}$ is very close to 0.5, the true value, with small standard errors. Second, the panel recursive estimate $\widehat{r}^e_{\widehat{g}}$ converges to the true value faster than the time-series estimate $\widehat{r}^e$. For almost all cases, the means of $\widehat{r}^e_{\widehat{g}}$ are closer to 0.5 with smaller standard errors and RMSE. Especially when the bubble signal is small ($\bar{c}_1 = 0.5$) or the bubble period is short ($T = 50$), the bias of $\widehat{r}^e_{\widehat{g}}$ is much smaller than 0.04 while the bias of $\widehat{r}^e$ can be bigger than 0.09. Third, when the explosive signal ($\bar{c}_1$) is stronger, it is easier to estimate the true origination date for both algorithms. In this case, bias, standard error, and RMSE of both recursive estimates are smaller. Fourth, when the sample size increases, it is easier to estimate $\left\{r^e_g\right\}_{g=1}^{K^0}$. Both the bias and standard errors become smaller, corroborating the consistency results. Specifically, the consistency of the panel recursive estimate benefits from the increase in both n and T. However, the time-series recursive estimate benefits only from the increase in T.

3.6 Conclusions

Explosiveness in time series is related to asset price bubbles in economics. That is why right-tailed unit root tests have been widely used to detect asset price bubbles in the literature. For example, PWY develop the sup ADF statistic for the presence of bubbles and the sequential algorithm for real-time dating of the origination of a bubble. This procedure relies on the recursive right-tailed unit root tests. PSY generalize the methods of PWY to deal with multiple bubble episodes. Like PWY, PSY only use a single time series. We show that when the bubble duration is short or when the bubble grows slowly, these tests have low power to identify bubbles.

In practice, panel data may be available where multiple time series have explosive behaviour. In this paper, we propose to use panel data models to improve the power in bubble detection. When there is homogeneity over cross-sectional units within groups, pooling cross-sectional data within the same group should sharpen the statistical inferences on the common explosive root.

However, it may be too strong to assume all of the time series in the panel data to have the same explosive root. In many applications, it may be too strong to consider all of the time series in the panel data to be explosive. That is why, in this paper, we introduce two panel models with a latent group structure. In the first model, we consider all of the time series are explosive, allowing individual heterogeneity through latent group structures. In the second model, we assume time series have mixed roots, some with stationary roots, some unit roots, and some explosive roots, again allowing individual heterogeneity through latent group structures. We propose a two-stage algorithm to detect the presence of the explosive behaviour and to estimate the origination date of the explosive period. In the first stage, we apply the k-means algorithms to recover group structures. In the second stage, we establish the post-classification estimates and tests based on the estimated groups.

Both model specifications are in the form of panel autoregressions, where we model individual heterogeneity through latent group structures. The group patterns represent the homogeneous slope coefficient within the identical group and heterogeneous autoregression coefficients across different groups. Under explosive panel autoregressions with latent group structures, we apply the recursive k-means algorithm and illustrate the consistency of the group clustering algorithm. Similarly, within the mixed-root panel autoregressions, a modified k-means clustering algorithm is implemented with consistency. Therefore, in the first stage of the computing algorithm, we successfully recover the group identities.

With estimated group structures, we can furthermore build post-classification estimates and inferences in the second stage. The post-classification estimators based on both k-means algorithms are asymptotically equivalent to the oracle estimates. Based on post-classification estimates, we provide two consistent t-tests on explosiveness detections. By applying the recursive t-statistics, we demonstrate a consistent estimate for bubble origination dates. Compared to PWY, the additional asymptotics can provide better limit behaviours.

It is possible to extend our model specifications. For example, an empirically restrictive assumption for the error process is martingale difference sequence with cross-sectional independence. For another example, allowing cross-sectionally and intertemporally dependent noise is more realistic in empirical analysis. These extensions will be considered in future work.

Chapter 4 Robust Inference with Stochastic Local Unit Root Regressors in Predictive Regressions

4.1 Introduction

Testing the predictability of financial asset returns has generated a vast literature in empirical finance and remains a major focus of ongoing research. Continuing concerns in the use of testing procedures are the validity, accuracy, and robustness of the econometric methods to the properties of the regressors that are used as economic fundamentals in the regressions. Several methods have been suggested to address these concerns and are now used in empirical work. Much attention in the development of these methods has been given to achieving robustness and size control in testing to the (typically unknown) degree of persistence in the regressors.

One well-established approach employs a local to unity (LUR) formulation to model potential persistence and simulation-based Bonferroni bounds for inference, as developed in Campbell and Yogo (2006). This approach is successful in the scalar LUR case but fails for stationary regressors (Phillips, 2014a), is challenging to extend to multiple regressors and is not designed for regressors with varying degrees of persistence or potentially stochastic departures from unit-roots (Phillips and Lee, 2013). Since these characteristics commonly arise in economic fundamentals, and their precise nature remains unknown, methods that are capable of robust inference under these conditions and which apply in multiple regressions are needed to support the growing body of empirical research on financial predictability.

A second approach that seeks to address these needs involves the use of endogenous instrumentation for the predictive regressors. This method, which is known as IVX instrumentation, was developed in Phillips and Magdalinos (2009) and utilized in an extensive empirical application by Kostakis, Magdalinos, and Stamatogiannis (2015; KMS, hereafter) to reveal its potential in applied research. The method has several significant advantages compared with Bonferroni-type simulation-based approaches. First, standard significance testing procedures such as Wald tests are applicable with convenient pivotal chi-square limit distributions that hold for a wide range of persistence characteristics in the regressors, including LUR and mildly integrated root (MIR) regressors (Phillips and Magdalinos, 2007). No simulation methods are therefore needed for implementation.

Second, no prior knowledge regarding the degree of persistence or the presence of stochastic departures from unit root conditions is required for IVX instrumentation. In contrast, the Bonferroni bound approach applies only to LUR regressors, and supplemental methods (like switching) are needed to support testing in the presence of MIR or stationary regressors (Elliott, Müller and Watson, 2015). Third, in contrast to other methods, IVX conveniently accommodates multiple regressors, as illustrated in KMS (2015), and allows an extension to locally explosive and mildly explosive regressor cases as well as mixed-root cases (Phillips and Lee, 2016). Fourth, the method applies in both short-horizon and long-horizon predictive regressions, again with pivotal chi-square limit theory for Wald tests. Fifth, the IVX methodology may be used in quantile regressions, as well as mean predictive regressions, as shown by Lee (2016) and subsequently, Fan and Lee (2019) in models with heteroskedasticity. This QR-IVX approach is particularly useful in checking for predictability under tail conditions where more extreme return behavior occurs.

A feature of the empirical finance literature that is particularly important for the present study is the recognition that parameter instability can be critical in both asset price determination as well as economic fundamentals. Coefficients can vary over time for many reasons, such as changes in regulatory conditions, shifts in market sentiment, or adjustments in monetary policy and targeting, as well as evolution in financial institutions and the impact of technology on transmission mechanisms and information dissemination. These influences can, in turn, affect the generating mechanism of predictive regressors, including the degree of persistence, thereby supporting formulations such as stochastic departures from unit roots and time-varying coefficient formulations. In the stock price application discussed later in the paper, a detection method for stochastic deviations from unity in the autoregressive coefficient is used to show empirical evidence of STUR behavior in the predictive regressor. In past research, there is ample support for time variation of the parameters in much financial econometric and macroeconomic modeling work. Bossaerts and Hillion (1999), for instance, cited poor performance in many prediction models and indications that the parameters of even the best models change over time. Amongst many other studies, Bekaert et al. (2007) showed patterns of time variation in model coefficients across sub-periods, and random walk specifications are frequently employed to capture parameter randomness and time variation in dynamic macroeconomic models (e.g., Cogley and Sargent, 2001, 2005).

The econometric literature has a long history of modeling with time-varying parameters, including stochastic process formulations. Granger and Swanson (1997) introduced the stochastic unit root model in which a stationary process is embedded into the autoregression coefficient, proposing a unit root test that accommodates such departures. Other early contributions of stochastic deviations from unit-roots appear in Leybourne et al. (1996), McCabe and Smith (1998), and Yoon (2006). More recently, Lieberman and Phillips (2017; hereafter LP) developed asymptotic theory and inferential procedures for nonlinear least-squares estimation of the STUR model and extended the Black-Scholes asset pricing formula to this more general model setup. LP (2020) further extended the analysis to a stochastic version of the local unit root model called the LSTUR model. Tao et al. (2019) studied a continuous-time variant of the same LSTUR model, which has been used to model derivative pricing in mathematical finance, employed infill asymptotics to establish asymptotic properties and analyzed evidence of instability and bubble behavior in financial data.

An essential feature of LSTUR models that is relevant for the use of persistent regressors in predictive regressions is that LSTUR processes have means and variances the same as an elementary random walk process but with kurtosis exceeding 3. This property is consistent with the well-known behavior of much financial data, thereby offering the prospect of improved modeling representations for both asset prices and economic fundamentals in predictive regressions. Whereas these features of STUR, LSTUR, and more general time-varying stochastic models are recognized as useful in capturing relevant financial data characteristics, there is as yet no treatment of the properties of predictive regression in the presence of such regressors.

The present paper seeks to respond to this need by studying short, and long-horizon mean predictive regressions and quantile predictive regressions with L-STUR regressors. The analysis reveals size distortions in predictability testing for both short horizon and long-horizon regressions with standard methods. A version of the endogenous instrumentation technique IVX is developed to address this problem in conventional predictive regression test procedures, together with the asymptotic properties of the IVX estimators and associated asymptotically pivotal tests. Both mean predictive regression (IVX) and quantile predictive regression (QR-IVX) approaches are considered. The IVX methods are shown to have excellent finite sample performance in simulations, not only with mixed (stationary and explosive) roots but also with random departures. In sum, the attractive features of IVX methodology are the availability of standard asymptotic chi-square inference in models with multiple predictive regressors and the allowance for random departures from unity in the autoregressive roots of the predictors.

Throughout the paper we use the following notation. For some arbitrary matrix A, we use $\|A\| \left(:= \max_i \left\{ \lambda_i^{\frac{1}{2}} : \lambda_i = \text{an eigenvalue of } A'A \right\} \right)$ to denote the spectral norm. The L_2 (Frobenius) and L_1 norms are specified as $\|\cdot\|_F$ and $\|\cdot\|_1$. 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, $\int$ represents $\int_0^1$ unless otherwise stated, and $\rightsquigarrow$ signifies weak convergence in both Euclidean space and function space according to context.

4.2 Model Setup and Size Distortions

This section presents three models that are commonly used in predictive regression: the short-horizon mean predictive regression model, the short-horizon quantile predictive regression model, and the long-horizon mean prediction model. Problems of size control in such predictive regressions when the regressors have persistent characteristics are well known. The present section discusses these issues in the context of conventional estimation and inferential methods. The framework provides a unified setting for inferences that allows for both stochastic and deterministic local departures from unity in the generating mechanisms of the persistent regressors.

4.2.1 Model and assumptions

The standard predictive mean regression model has the form

$$y_t = \beta_0 + \beta_1' x_{t-1} + u_{0t}, \text{ with } \mathbb{E}(u_{0t}|\mathcal{F}_{t-1}) = 0, \tag{4.2.1}$$

where β_1 is an n-vector of regression coefficients and $\mathcal{F}_t$ is a suitable filtration, defined later, for which u_{0t} is a martingale difference sequence (mds). The n-vector of predictors x_{t-1} in (4.2.1) is assumed throughout this paper to have a stochastic unit root generating mechanism that belongs to the STUR or LSTUR family, implying the following persistent autoregressive form,

$$x_t = R_{Tt} x_{t-1} + u_{xt}, \; R_{Tt} = \begin{cases} I_n + \frac{C}{T} + \frac{\check{D}_{at}}{\sqrt{T}} + \frac{\check{D}_{at}^2}{T} & \text{under LSTUR,} \\ I_n + \frac{\check{D}_{at}}{\sqrt{T}} + \frac{\check{D}_{at}^2}{T} & \text{under STUR.} \end{cases} \tag{4.2.2}$$

In (4.2.2), T is the sample size, $C := diag\{c_1, c_2, ..., c_n\}$ with $\{c_i\}_{i=1}^n$ being a set of scalar localizing coefficients, and $\breve{D}_{at} := \text{diag}\{a_1' u_{at}, a_2' u_{at}, ..., a_n' u_{at}\}$ with u_{at} a p-vector of random variables that influence stochastic departures from unit roots in the coefficient matrix R_{Tt}. We assume there is one generic bound $\bar{c}$, for all i, $|c_i| < \bar{c}$. The set $\{a_i\}_{i=1}^n$ collect the p-vector coefficients that appear in the STUR formulation and $A' := [a_1, ..., a_n]$. For simplicity, we assume $\{a_i\}_{i=1}^n$ have the same value, a. Therefore $\breve{D}_{at} = \left(a' u_{at}\right) \cdot I_n$. Under the asymptotic scheme, n is fixed while $T \to \infty$ as only finite regressors are considered.

In practical work, the degree of persistence and presence of stochastic departures from unit roots in economic time series are not observable. In the endogenous case where there is correlation between u_{at} and u_{xt} in the regression model (4.2.1) a further complication is that the STUR coefficients a_i are not consistently estimable, as shown in LP (2017). Moreover, standard unit root tests have little discriminatory power in distinguishing between unit root and STUR (or LSTUR) processes in models such as (4.2.2). It is therefore desirable to have methods of estimation and inference that are robust to different formulations within this general class of near $I(1)$ regressors.

The following generating structure for the innovations in (4.2.1) and (4.2.2) allows for weak intertemporal dependences among the u_{xt} and u_{0t}. Within this structure, a conventional conditionally homoskedastic martingale difference sequence assumption is made for u_{0t}.

$$
\begin{aligned}
u_t &= \begin{bmatrix} u_{0t} & (1 \times 1) \\ u_{xt} & (n \times 1) \\ u_{at} & (p \times 1) \end{bmatrix} = \sum_{j=0}^{\infty} F_j \epsilon_{t-j},\ \epsilon_t \sim mds(0, \Sigma),\ \mathbb{E}\left\|\epsilon_1\right\|^4 < \infty, \sum_{j=0}^{\infty} j \left\|F_j\right\| < \infty, \\
F(1) &= \sum_{j=0}^{\infty} F_j \neq 0,\ F_j = \begin{bmatrix} F_{0j} \\ F_{xj} \\ F_{aj} \end{bmatrix},\ F_{0j} = \begin{cases} \left[1 : O_{1\times(n+p)}\right] & \text{for } j = 0 \\ O_{1\times(1+n+p)} & \text{for } j \geq 1 \end{cases},\ u_t = F(1)\epsilon_t - \Delta\widetilde{\epsilon}_t, \\
\widetilde{\epsilon}_t &= \sum_{j=0}^{\infty} \widetilde{F}_j \epsilon_{t-j},\ \widetilde{F}_j = \sum_{s=j+1}^{\infty} F_s,\ F(z) = \sum_{j=0}^{\infty} F_j z^j, \Omega_{00} = \mathbb{E}(u_{0t}^2),
\end{aligned}
$$

$$\begin{aligned}
\Omega_{0x} &= \sum_{j=-\infty}^{\infty} \mathbb{E}(u_{0t}u'_{x,t-h}) = \Sigma F'_x(1),\ \Omega_{0a} = \sum_{j=-\infty}^{\infty} \mathbb{E}(u_{0t}u'_{a,t-h}) = \Sigma F'_a(1), \qquad (4.2.3)\\
\Omega_{xx} &= \sum_{j=-\infty}^{\infty} \mathbb{E}(u_{xt}u'_{x,t-h}) = F_x(1)\Sigma F'_x(1),\ \Omega_{aa} = \sum_{j=-\infty}^{\infty} \mathbb{E}(u_{at}u'_{a,t-h}) = F_a(1)\Sigma F'_a(1),\\
\overline{\Delta}_{ax} &= \sum_{h=0}^{\infty} \mathbb{E}(\breve{D}_{at}u_{x,t+h}),\ \overline{\Omega}_{aa} = \sum_{h=-\infty}^{\infty} \mathbb{E}\left(\breve{D}_{at}\breve{D}_{a,t-h}\right).\\
\overline{\Sigma}_{aa} &= \mathbb{E}\left(\breve{D}_{at}^2\right), \overline{\Delta}_{aa} = \sum_{h=0}^{\infty} \mathbb{E}\left(\breve{D}_{at}\breve{D}_{a,t+h}\right). \qquad (4.2.4)
\end{aligned}$$

Under these conditions, the usual functional limit theory holds (c.f., Phillips and Solo, 1992)

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{[Ts]} u_j := \frac{1}{\sqrt{T}}\sum_{j=1}^{[ns]} \begin{bmatrix} u_{0j} \\ u_{xj} \\ u_{aj} \end{bmatrix} \rightsquigarrow \begin{bmatrix} B_0(s) \\ B_x(s) \\ B_a(s) \end{bmatrix} = BM \begin{bmatrix} \Omega_{00} & \Omega_{0x} & \Omega_{0a} \\ \Omega_{x0} & \Omega_{xx} & \Omega_{xa} \\ \Omega_{a0} & \Omega_{ax} & \Omega_{aa} \end{bmatrix}.$$

The notation BM denotes a Brownian motion. The limit processes for the STUR and LSTUR cases are given by the following limits:

$$\frac{x_{\lfloor Tr \rfloor}}{\sqrt{T}} \rightsquigarrow G_a(r) := e^{\breve{D}_{B_a}(r)} \left\{ \int_0^r e^{-\breve{D}_{B_a}(p)} dB_x(p) - \left(\int_0^r e^{-\breve{D}_{B_a}(p)} dp \right) \overline{\Delta}_{ax} \right\}, \qquad (4.2.5)$$

and

$$\frac{x_{\lfloor Tr \rfloor}}{\sqrt{T}} \rightsquigarrow G_{a,c}(r) := e^{rC+\breve{D}_{B_a}(r)} \left\{ \int_0^r e^{-pC-\breve{D}_{B_a}(p)} dB_x(p) - \left(\int_0^r e^{-pC-\breve{D}_{B_a}(p)} dp \right) \overline{\Delta}_{ax} \right\}, \qquad (4.2.6)$$

where $\breve{D}_{B_a}(r) := \text{diag}\left\{a'_1 B_a(r), a'_2 B_a(r), ..., a'_n B_a(r)\right\}$. When u_{at} is exogenous, the covariance $\overline{\Delta}_{ax} = 0$.

The limits (4.2.5) and (4.2.6) were obtained by LP (2017, 2020) in the scalar case and similar derivations (not repeated here) lead to (4.2.5) and (4.2.6). LP (2020) derived the stochastic differential equation satisfied by these stochastic processes and showed that they have instantaneous means and variances that resemble those of Brownian motion but with instantaneous kurtosis exceeding 3, a feature that is more coherent with the properties of much financial data for which heavy tail behavior is commonly present. For this reason, persistent processes such as STUR and LSTUR with limit processes of the type $G_a(r)$ and $G_{a,c}(r)$ have the potential to achieve better predictive performance in practical work with financial data. In what follows, we examine this potential in the context of both short-horizon predictive regression (both mean and quantile regressions), and long-horizon mean predictive regression.

Case I: Short-horizon Mean and Quantile Predictive Regressions

For the short-horizon mean predictive regression case, the model setup and innovation structures are given in (4.2.1), (4.2.2), and (4.2.3).

In addition to the mean predictive regression, the following quantile predictive regression model is considered. Following Xiao (2009), our model is based on the linear quantile representation,

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \beta_{0,\tau} + \beta_{1,\tau}^{'} x_{t-1}, \tag{4.2.7}$$

where $Q_{y_t}(\tau|\mathcal{F}_{t-1})$ is the conditional quantile of y_t so that

$$\Pr(y_t \leq Q_{y_t}(\tau|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}) = \tau \in (0,1).$$

When $\tau = \frac{1}{2}$, the quantile predictive regression reduces to a median predictive regression, whereas (4.2.7) allows for potential predictability for other quantiles besides the median. The loss function of quantile regression is defined by $\rho_\tau(u) := u \cdot \Psi_\tau(u)$ with $\Psi_\tau := \tau - 1(u < 0)$. The quantile regression innovations follow $\Psi_\tau(u_{0t\tau}) \sim mds(0, \tau(1-\tau))$ where $u_{0t\tau} := u_{0t} - P_{u_0}^{-1}(\tau)$ and $P_{u_0}^{-1}(\tau)$ is the unconditional $\tau-$quantitle of u_{0t}. By Theorem 2.1 of Phillips and Durlauf (1986), a functional law for relevant components in the quantile regression involving the innovations $\Psi_\tau(u_{0t\tau})$ is given by

$$\frac{1}{\sqrt{T}}\sum_{j=1}^{\lfloor Ts \rfloor}\begin{bmatrix}\Psi_\tau(u_{0j\tau})\\ u_{xj}\\ u_{aj}\end{bmatrix} \rightsquigarrow \begin{bmatrix}B_{\Psi_\tau}(s)\\ B_x(s)\\ B_a(s)\end{bmatrix} = BM\begin{bmatrix}\tau(1-\tau) & \Omega_{\Psi_\tau x} & \Omega_{\Psi_\tau a}\\ \Omega_{x\Psi_\tau} & \Omega_{xx} & 0\\ \Omega_{a\Psi_\tau} & 0 & \Omega_{aa}\end{bmatrix}. \quad (4.2.8)$$

In addition, to facilitate the asymptotic development, several regularity conditions on the conditional density of $u_{0t\tau}$ are imposed.

Assumption 3 *(i) The sequence of stationary conditional densities pdf* $\{p_{u0t\tau,t-1}(\cdot)\}$ *evaluated at zero satisfies a functional central limit theorem with a nondegenerate mean* $p_{u0\tau}(0) = \mathbb{E}[p_{u0t\tau,t-1}(0)]$*, where* $p_{u0t\tau,t-1}(\cdot)$ *is the conditional density of* $u_{0t\tau}$ *given* $\mathcal{F}_{t-1}$*,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr \rfloor}(p_{u0t\tau,t-1}(0) - p_{u0\tau}(0)) \rightsquigarrow B_{pu0\tau}(r).$$

(ii) For each t and $\tau \in (0,1)$*, the conditional density* $p_{u0t\tau,t-1}(0)$ *is bounded above with probability one, i.e.* $p_{u0t\tau,t-1}(x) < \infty$ *in probability for all* $|x| < M$ *and some* $M > 0$*.*

Case II: Long-horizon Mean Predictive Regression

In short-horizon predictive regressions, the time horizon in prediction accommodates a single period. It is common to use longer horizons in empirical research. For long horizon predictive models, the time horizon is mildly increasing with the sample size. In particular, we set the horizon as $k = T^\nu$ for some $\nu \in (0,1)$ and have the implied rate restriction, $\frac{1}{k} + \frac{k}{T} \to 0$. As discussed in Phillips and Lee (2013), the long horizon prediction model with a mildly increasing time window has formulations as follows

$$\begin{aligned} y_{t+k} &= B_0^* + (B_1^*)x_t^k + u_{0,t+k},\ x_t^k = \sum_{j=1}^{k} x_{t+j-1}, \\ \mathcal{H}_0'(k) &: B_1^* = 0, \end{aligned} \quad (4.2.9)$$

under the null hypothesis. The alternative hypothesis is $\mathcal{H}_1'(k) : B_1^* \neq 0$. Inference concerning long-horizon predictability under STUR and LSTUR regressors can be conducted by empirically fitting equation (4.2.9) allowing for stationary disturbances in the long-horizon model.

4.2.2 Size distortions in traditional regressions

Allowing for STUR and LSTUR regressors, we provide below the limit theory for standard short-horizon quantile and mean predictive regressions. These results facilitate the study of the size distortions that arise in the use of standard testing methods.

The ordinary least squares (OLS) estimator $\widehat{\beta}$ has the usual form in mean regressions, and standard quantile regression estimators optimize the following objective function

$$\widehat{\beta}_\tau^{QR} := \arg\min_\beta \sum_{t=1}^T \rho_\tau(y_t - \beta' X_{t-1}), \tag{4.2.10}$$

where $\rho_\tau(u) := u(\tau - 1(u < 0))$, $\tau \in (0, 1)$. The notation $X_{t-1} := (1, x_{t-1}')'$ includes both the intercept and the persistent regressor x_{t-1}. By the standardization matrix $D_T := diag\left\{\sqrt{T}, T \times I_n\right\}$, we have the limit theory of OLS and standard quantile estimators.

Proposition 4.2.1 *Under LSTUR regressors, as $T \to \infty$,*

$$\begin{aligned}
D_T(\widehat{\beta} - \beta) &\rightsquigarrow \begin{bmatrix} 1 & \int_0^1 G_{a,c}'(r) \\ \int_0^1 G_{a,c}(r) & \int_0^1 G_{a,c}(r) G_{a,c}'(r) \end{bmatrix}^{-1} \begin{bmatrix} B_0(1) \\ \int_0^1 G_{a,c}(r) dB_0(r) \end{bmatrix}, \\
D_T(\widehat{\beta}_\tau^{QR} - \beta_\tau) &\rightsquigarrow p_{u0\tau}^{-1}(0) \begin{bmatrix} 1 & \int_0^1 G_{a,c}'(r) \\ \int_0^1 G_{a,c}(r) & \int_0^1 G_{a,c}(r) G_{a,c}'(r) \end{bmatrix}^{-1} \begin{bmatrix} B_{\Psi_\tau}(1) \\ \int_0^1 G_{a,c}(r) dB_{\Psi_\tau}(r) \end{bmatrix}.
\end{aligned}$$

In the scalar regressor case where $n = 1$, $u_{xt} \sim mds(0, \Sigma_{xx})$, and $C = 0$, the LSTUR model reduces to the scalar STUR. By the orthogonal decomposition of Brownian motion (Phillips, 1989) $dB_{\Psi_\tau} = dB_{\Psi_\tau|x} + \Sigma_{\Psi_\tau x} \Sigma_{xx}^{-1} dB_x$, and setting $\overline{G}_{a,c} := G_{a,c}(r) - \int_0^1 G_{a,c}(r) dr$, we have the following limit theory for the QR regression coefficient with given τ,

$$
\begin{aligned}
T(\widehat{\beta}^{QR}_{1,\tau}-\beta_{1,\tau}) &\rightsquigarrow p_{u0\tau}(0)^{-1}\frac{\int \overline{G}_{a,c}dB_{\Psi_\tau}}{\int(\overline{G}_{a,c})^2} \\
&= p_{u0\tau}(0)^{-1}\left[\frac{\int \overline{G}_{a,c}dB_{\Psi_\tau|x}}{\int(\overline{G}_{a,c})^2}+\Sigma_{\Psi_\tau x}\Sigma_{xx}^{-1}\frac{\int \overline{G}_{a,c}dB_x}{\int(\overline{G}_{a,c})^2}\right],
\end{aligned}
$$

under the null hypothesis of no predictability. The notation $\int$ is $\int_0^1$ for short. Using the estimated standard error $s.e.(\widehat{\beta}^{QR}_{1,\tau}) = \tau(1-\tau)\widehat{p}_{u0\tau}(0)^{-1}\left\{\sum_{t=1}^T(x^\mu_{t-1})^2\right\}^{-\frac{1}{2}}$ where $x^\mu_{t-1}=x_{t-1}-\frac{1}{T}\sum_{t=1}^T x_{t-1}$, and a consistent nonparametric estimate $\widehat{p}_{u0\tau}(0)$ for $p_{u0\tau}(0)$, the t-ratio test statistic for a given τ under the null hypothesis $\mathcal{H}_0$: $\beta_{1,\tau}=0$ satisfies,

$$
\begin{aligned}
t_{\widehat{\beta}^{QR}_{1,\tau}-\beta_{1,\tau}} = \frac{(\widehat{\beta}^{QR}_{1,\tau}-\beta_{1,\tau})}{s.e.(\widehat{\beta}^{QR}_{1,\tau})} &\rightsquigarrow \left[1-\frac{\Sigma^2_{\Psi_\tau|x}}{\Sigma_{xx}\tau(1-\tau)}\right]^{\frac{1}{2}}\mathcal{N}(0,1)+\frac{\Sigma_{\Psi_\tau|x}}{(\Sigma_{xx}\tau(1-\tau))^{\frac{1}{2}}}\frac{\int \overline{G}_{a,c}dB_x}{\left[\int(\overline{G}_{a,c})^2\right]^{\frac{1}{2}}} \\
&= \left[1-\lambda(\tau)^2\right]^{\frac{1}{2}}Z+\lambda(\tau)\eta_{LP}(a,c), \qquad (4.2.11)
\end{aligned}
$$

where $\Sigma_{\Psi_\tau x}$ is defined as the instantaneous covariance of $u_{0t\tau}$ and u_{xt}, and $\Sigma_{\Psi_\tau|x}$ is define as $\tau(1-\tau)-\Sigma_{xx}^{-1}\Sigma^2_{\Psi_\tau x}$. The random variables $Z :=_d \mathcal{N}(0,1)$, and $\eta_{LP}(a,c) := \int \overline{G}_{a,c}dB_x/\left(\int(\overline{G}_{a,c})^2\right)^{1/2}$. The non-zero factor $\lambda(\tau)$ in the limit expression (4.2.11) reveals the presence of size distortions in the usual t-ratio statistic. The notation Σ_{xx}, and $\Sigma_{\Psi_\tau x}$ denote the instantaneous variance and covariance of u_{xt} and Ψ_τ $(u_{0t\tau})$.

Similar size distortion problems arise in inferences using OLS. Under the null hypothesis $\mathcal{H}_0 : \beta_1 = 0$, the limit theory of the usual t-ratio statistic is

$$
\begin{aligned}
t_{\widehat{\beta}_1-\beta_1} = \frac{(\widehat{\beta}_1-\beta_1)}{s.e.(\widehat{\beta}_1)} &\rightsquigarrow \left[1-\frac{\Sigma^2_{0x}}{\Sigma_{xx}\Omega_{00}}\right]^{\frac{1}{2}}\mathcal{N}(0,1)+\frac{\Sigma_{0x}}{(\Sigma_{xx}\Omega_{00})^{\frac{1}{2}}}\frac{\int \overline{G}_{a,c}dB_x}{\left[\int(\overline{G}_{a,c})^2\right]^{\frac{1}{2}}} \\
&= \left[1-\lambda^2\right]^{\frac{1}{2}}Z+\lambda\eta_{LP}(a,c),
\end{aligned}
$$

where $s.e.(\widehat{\beta}_1) := \widehat{\Omega}_{00}\left\{\sum_{t=1}^T(x^\mu_{t-1})^2\right\}^{-\frac{1}{2}}$ and $\widehat{\Omega}_{00}$ is a consistent estimator of Ω_{00}. Again the non-zero λ arising from the covariance of u_{0t} and u_{xt}, reveals the presence of size distortions in the standard hypothesis testing using mean predictive regressions.

4.3 Limit Theory for IVX Filtering

The intuition that underlies IVX instrument creation is to filter persistent data on endogenous regressors x_t to generate an instrument $\widetilde{z}_t$ that has the appealing property of asymptotic orthogonality to structural equation errors while retaining asymptotic relevance for the regressors. The generation process can be described in terms of the following equation

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

so that the time series of innovations $\{\Delta x_t\}$ in the endogenous regressor is passed through an autoregressive filter to produce $\widetilde{z}_t$ using some suitable autoregressive coefficient matrix F. When $F = 0$, then $\widetilde{z}_t = \Delta x_t$, and the first-order difference operator is used to remove distortions at the cost of substantial information loss. When $F = I_n$, then $\widetilde{z}_t = x_t$ and IVX reduces to ordinary least squares and ordinary quantile regression estimations with non-negligible bias and size distortion as discussed above. The key idea in the successful choice of the coefficient matrix F is to generate instruments $\widetilde{z}_t$, which are intermediate in forms and properties between first-difference and level data.

A simple way to accomplish this intermediate form is to select F so that by virtue of its construction $\widetilde{z}_t$ is a mildly integrated (MIR) process. In particular, for $F = R_{Tz} = I_n + C_z/T^{\gamma}$, we have

$$\begin{aligned} \widetilde{z}_t &= R_{Tz}\widetilde{z}_{t-1} + \Delta x_t,\ R_{Tz} = I_n + \frac{C_z}{T^{\gamma}}, \\ \text{where } \gamma &\in (0,1), C_z = c_z I_n,\ c_z < 0,\ \widetilde{z}_0 = 0. \end{aligned} \tag{4.3.1}$$

With this simple construction, the coefficient matrix R_{Tz} is diagonal with entries that lie between zero and unity but which are much closer to unity, especially for large sample sizes. The advantage of IVX estimation, including both QR-IVX and LH-IVX is that this technique of instrumental variable regression attains robust inference that includes stationary, unit root, and local unit root regressors, as well as regressors with mixed properties, thereby covering a broad class of cointegrating regression and predictive regression models and avoiding the size distortions that are known to arise in traditional methods of estimating and conducting inferences with such systems.

In this paper, these robustness properties of IVX estimation and inference are shown to extend to cases of STUR and LSTUR regressors, as well as short-horizon and long-horizon predictive regression models. With this methodology, inferences are valid using standard chi-square limit theory for Wald tests, thereby providing a convenient tool for empirical work in predictive regression with an extensive class of endogenous regressors.

4.3.1 IVX in short-horizon predictive regression

In short-horizon predictive regressions, IVX instruments are constructed using the observed data $\{x_t\}_{t=1}^{T}$ as

$$\widetilde{z}_t = \sum_{j=1}^{t} R_{Tz}^{t-j} \Delta x_j, \text{ where } R_{Tz} = I_n + \frac{C_z}{T^\gamma}, \ \gamma \in (0,1), \ C_z = c_z I_n, \text{ with } c_z < 0, \tag{4.3.2}$$

for some suitable choices of c_z and γ, such as $\gamma = 0.95$ and $c_z = -1$. The short-horizon IVX estimate is conducted in terms of the standard IV procedure and the short-horizon QR-IVX estimate is provided by minimizing the following objective function:

$$\widehat{\beta}_{1,\tau}^{QRIVX} := \arg\min_{\beta_1} \left(\sum_{t=1}^{T} m_t(\beta_1) \right)' \left(\sum_{t=1}^{T} m_t(\beta_1) \right), \tag{4.3.3}$$

where $m_t(\beta_1) = \widetilde{z}_{t-1}\left(\tau - 1\left(y_{t\tau} \leq \beta_1' x_{t-1}\right)\right)$. As the objective function is non-smooth, this paper employs the method of Xiao (2009) and Lee (2016) to derive the limit theory.

Since $\Delta x_j = u_{xj} + \frac{\breve{D}_{aj}}{\sqrt{T}} x_{j-1} + \frac{C}{T} x_{j-1} + \frac{\breve{D}_{aj}^2}{T} x_{j-1}$ in the LSTUR case and $\Delta x_j = u_{xj} + \frac{\breve{D}_{aj}}{\sqrt{T}} x_{j-1} + \frac{\breve{D}_{aj}^2}{T} x_{j-1}$ under STUR, the corresponding decomposition for $\widetilde{z}_{t-1}$ is

$$
\begin{aligned}
\widetilde{z}_{t-1} &= \sum_{j=1}^{t-1} R_{Tz}^{t-j} \left(u_{xj} + \frac{\breve{D}_{aj}}{\sqrt{T}} x_{j-1} + \frac{C}{T} x_{j-1} + \frac{\breve{D}_{aj}^2}{T} x_{j-1} \right) \qquad (4.3.4) \\
&= \sum_{j=1}^{t-1} R_{Tz}^{t-j} u_{xj} + \frac{1}{\sqrt{T}} \sum_{j=1}^{t-1} R_{Tz}^{t-j} \breve{D}_{aj} x_{j-1} + \frac{C}{T} \sum_{j=1}^{t-1} R_{Tz}^{t-j} x_{j-1} \\
&\quad + \frac{1}{T} \sum_{j=1}^{t-1} R_{Tz}^{t-j} \breve{D}_{aj}^2 x_{j-1} \\
&= z_{t-1} + \frac{1}{\sqrt{T}} \eta_{T,t-1}^{(2)} + \frac{C}{T} \eta_{T,t-1}^{(1)} + \frac{1}{T} \eta_{T,t-1}^{(3)}, \qquad (4.3.5)
\end{aligned}
$$

where $\eta_{T,t}^{(2)} := \sum_{j=1}^{t} R_{Tz}^{t-j} \breve{D}_{aj} x_{j-1}$, $\eta_{T,t}^{(1)} := \sum_{j=1}^{t} R_{Tz}^{t-j} x_{j-1}$, and $\eta_{T,t}^{(3)} := \sum_{j=1}^{t} R_{Tz}^{t-j} \breve{D}_{aj}^2 x_{j-1}$. Similarly, for STUR, the decomposition follows $\widetilde{z}_{t-1} = z_{t-1} + \frac{1}{\sqrt{T}} \eta_{T,t-1}^{(2)} + \frac{1}{T} \eta_{T,t-1}^{(3)}$. By Phillips and Magdalinos (2009, equation A18),

$$
\max_{1 \le t \le T} \frac{1}{T^{\gamma/2}} \mathbb{E} \left\| z_{t-1} \right\| = O_p(1), \text{and } \max_{1 \le t \le T} \frac{1}{T^{1/2+\gamma/4}} \left\| z_{t-1} \right\| = o_p(1).
$$

With LSTUR and STUR regressors in the remainder term, $\frac{1}{\sqrt{T}} \eta_{T,t-1}^{(2)}$, the asymptotic theory turns out to be more complicated than the cases studied in Phillips and Magdalinos (2009). The remainder term has a substantial effect on the limit behavior of both the numerator and denominator in short-horizon IVX estimations. Under local unit root and mildly integrated regressor cases, the following decomposition for IVX instruments applies

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

where $\alpha \in (0, 1]$, z_{t-1} and $\eta_{T,t-1}^{(1)}$ are defined as above. In the numerator, asymptotically the term associated with latent instrument z_{t-1} dominates the one containing the remainder $\frac{C}{T^{\alpha}}\eta^{(1)}_{Tt-1}$, while in the denominator the two terms demonstrate the same stochastic order. Also, when mildly explosive regressors (See, Phillips and Lee, 2013) are under consideration, the robust inference is possible using IVX and QR-IVX. But the asymptotic theory for mildly explosive cases is different from the stationary side of unity: in this case, the term related to $\frac{C}{T^{\alpha}}\eta^{(1)}_{Tt-1}$ dominates the ones with the latent instrument z_{t-1} in both the numerator and the denominator of IVX estimator.

For STUR and LSTUR regressors, the asymptotic behavior of $\widetilde{z}_{t-1}$ is quite different. In the numerator of the IVX estimator, terms containing $\frac{1}{\sqrt{T}}\eta^{(2)}_{T,t-1}$ and z_{t-1} dominate and lead to a joint mixed normality, so the terms containing the IVX remainders , $\frac{C}{T}\eta^{(1)}_{T,t-1}$ and $\frac{1}{T}\eta^{(3)}_{T,t-1}$, are dominated and asymptotically vanish. In the denominator of the IVX estimator, the term containing $\frac{1}{\sqrt{T}}\eta^{(2)}_{T,t-1}$ shares the same stochastic order as the ones composed of z_{t-1}, $\frac{C}{T}\eta^{(1)}_{T,t-1}$ and $\frac{1}{T}\eta^{(3)}_{T,t-1}$. We collect the asymptotic results in the following theorem.

Theorem 4.3.1 *As $T \to \infty$,*

$$T^{\frac{1+\gamma}{2}}(\widehat{\beta}^{QR-IVX}_{1,\tau} - \beta_{1,\tau}) \rightsquigarrow \mathcal{MN}\left(0, \frac{\tau(1-\tau)}{p^2_{u0\tau}(0)}(V_{xz})^{-1}\left[V_{zz} + V^{(2)}_{\eta\eta}\right]((V_{xz})^{-1})'\right),$$

and

$$T^{\frac{1+\gamma}{2}}(\widehat{\beta}^{IVX}_{1} - \beta_{1}) \rightsquigarrow \mathcal{MN}\left(0, \Omega_{00}(V_{xz})^{-1}\left[V_{zz} + V^{(2)}_{\eta\eta}\right]((V_{xz})^{-1})'\right),$$

where the expressions of the limiting matrices V_{zz}, $V^{(2)}_{\eta\eta}$ and V_{xz} are provided in the Appendix.

Although the limiting distributions of the IVX and QR-IVX estimates are non-pivotal, the corresponding Wald tests are asymptotically chi-square distributed.

Theorem 4.3.2 *(i) As $T \to \infty$, under the null hypothesis $\mathcal{H}_0 : \beta_{1,\tau} = \beta_{1,\tau}^0$,*

$$\frac{\widehat{p_{u0\tau}(0)}^2}{\tau(1-\tau)}(\widehat{\beta}_{1,\tau}^{QR-IVX} - \beta_{1,\tau}^0)'(X'M_{\widetilde{Z}}X)(\widehat{\beta}_{1,\tau}^{QR-IVX} - \beta_{1,\tau}^0) \rightsquigarrow \chi^2(n),$$

where $X'M_{\widetilde{Z}}X := (\sum_{t=2}^T x_{t-1}\widetilde{z}_{t-1}')(\sum_{t=2}^T \widetilde{z}_{t-1}\widetilde{z}_{t-1}')^{-1}(\sum_{t=2}^T x_{t-1}\widetilde{z}_{t-1}')'$ and $\widehat{p_{u0\tau}(0)}$ is any consistent nonparametric estimator for $p_{u0\tau}(0)$.

(ii) As $T \to \infty$, under the null hypothesis $\mathcal{H}_0 : \beta_1 = \beta_1^0$,

$$\frac{1}{\widehat{\Omega_{00}}}(\widehat{\beta}_1^{IVX} - \beta_1^0)'(X'M_{\widetilde{Z}}X)(\widehat{\beta}_{1,\tau}^{IVX} - \beta_1^0) \rightsquigarrow \chi^2(n),$$

where $X'M_{\widetilde{Z}}X$ is given in (i) and $\widehat{\Omega_{00}}$ is any consistent estimator for Ω_{00}.

These results are readily extended to test predictability under general linear restrictions such as $\mathcal{H}_0 : H\beta_{1,\tau} = h$ where H is a known $\ell \times n$ matrix of rank ℓ and h is a known ℓ-vector. The statistic for short-horizon QR-IVX testing is

$$\frac{\widehat{p_{u0\tau}(0)}^2}{\tau(1-\tau)}(H\widehat{\beta}_{1,\tau}^{QR-IVX} - h)'\left\{H\left(X'M_{\widetilde{Z}}X\right)^{-1}H'\right\}^{-1}(H\widehat{\beta}_{1,\tau}^{QR-IVX} - h) \rightsquigarrow \chi^2(\ell).$$

Similarly, to test $\mathcal{H}_0 : H\beta_1 = h$,

$$\frac{1}{\widehat{\Omega_{00}}}(H\widehat{\beta}_1^{IVX} - h)'\left\{H\left(X'M_{\widetilde{Z}}X\right)^{-1}H'\right\}^{-1}(H\widehat{\beta}_1^{IVX} - h) \rightsquigarrow \chi^2(\ell).$$

4.3.2 IVX in long-horizon predictive regression

As the time horizon in predictive regressions rises, it is convenient to use an alternative form of the mean predictive regression model, following Phillips and Lee (2013). We employ a similar formulation and apply the long-run variant of the IVX approach, called LHIVX, in the following analysis with LSTUR and STUR regressors. It is shown that LHIVX leads to a mixed normal limit distribution in estimation and a pivotal chi-square limit theory in Wald testing.

The regressor in the long-horizon fitted regression model (4.2.9) is given by the partial sum $x_t^k := \sum_{j=1}^k x_{t+j-1}$. Each component of the LHIVX instrument is constructed in the identical fashion with similar notation as $\widetilde{z}_t^k := \sum_{j=1}^k \widetilde{z}_{t+j-1}$, $z_t^k := \sum_{j=1}^k z_{t+j-1}$, $\eta_{T,t}^{1,k} := \sum_{j=1}^k \eta_{T,t+j-1}^{(1)}$, $\eta_{T,t}^{2,k} := \sum_{j=1}^k \eta_{T,t+j-1}^{(2)}$ and $\eta_{T,t}^{3,k} := \sum_{j=1}^k \eta_{T,t+j-1}^{(3)}$ where $\widetilde{z}_{t+j-1}$, z_{t+j-1}, $\eta_{T,t+j-1}^{(1)}$, $\eta_{T,t+j-1}^{(2)}$ and $\eta_{T,t+j-1}^{(3)}$ are each defined in the short-horizon IVX case of (4.3.5). Therefore, we have the following decomposition as

$$\widetilde{z}_t^k := \begin{cases} z_t^k + \frac{C}{T}\eta_{T,t}^{1,k} + \frac{1}{\sqrt{T}}\eta_{T,t}^{2,k} + \frac{1}{T}\eta_{T,t}^{3,k} & \text{under LSTUR,} \\ z_t^k + \frac{1}{\sqrt{T}}\eta_{T,t}^{2,k} + \frac{1}{T}\eta_{T,t}^{3,k} & \text{under STUR.} \end{cases} \tag{4.3.6}$$

For the development of asymptotic theory, we place further conditions on the time horizon parameter k and IVX rate parameter γ:

$$\frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0.$$

This restriction requires the horizon k to rise not as fast as T but faster than T^γ. The estimator $\widehat{B^*}_1^{LHIVX}$ of B_1^* satisfies

$$\widehat{B^*}_1^{LHIVX} - B_1^* = \left(\sum_{t=1}^{T-k} u_{0,t+k}(\widetilde{z}_t^k)'\right)\left(\sum_{t=1}^{T-k} x_t^k(\widetilde{z}_t^k)'\right)^{-1}. \tag{4.3.7}$$

The limit theory for this LHIVX estimator with LSTUR and STUR regressors differs from the LUR and MIR cases. In the LUR and MIR cases, the term $\sum_{t=1}^{T-k} u_{0,t+k}(z_t^k)'$ determines the behavior of the numerator in the matrix quotient, whereas in the mildly explosive case the term $T^{-1}\sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{1,k})'$ dominates other terms in the numerator. In this paper, with STUR and LSTUR regressors, both $\sum_{t=1}^{T-k} u_{0,t+k}(z_t^k)'$ and $T^{-1/2}\sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{2,k})'$ dominate $T^{-1}\sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{1,k})'$ and $T^{-1}\sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{3,k})'$ in the numerator. Nonetheless, joint convergence to mixed normality still holds and the numerator of the centred LHIVX estimator (4.3.7) has the following decomposition

$$\frac{1}{T^{\frac{1}{2}+\gamma}\sqrt{k}}\sum_{t=1}^{T-k}u_{0,t+k}(\widetilde{z}_t^k)' = \frac{1}{T^{\frac{1}{2}+\gamma}\sqrt{k}}\sum_{t=1}^{T-k}u_{0,t+k}(z_t^k)' + \frac{1}{T^{1+\gamma}\sqrt{k}}\sum_{t=1}^{T-k}u_{0,t+k}(\eta_{T,t}^{2,k})' + o_p(1). \tag{4.3.8}$$

When LSTUR degenerates into STUR ($C = 0_{n\times n}$), the decomposition of (4.3.8) is still applicable. Similarly, the properties of each term in the LHIVX denominator can be obtained. Therefore, the denominator of the LHIVX estimator with LSTUR regressors follows a decomposition as

$$\begin{aligned}\frac{1}{T^{1+\gamma}k^2}\sum_{t=1}^{T-k}x_t^k(\widetilde{z}_t^k)' &= \frac{1}{T^{1+\gamma}k^2}\sum_{t=1}^{T-k}x_t^k(z_t^k)' + \frac{1}{T^{\frac{3}{2}+\gamma}k^2}\sum_{t=1}^{T-k}x_t^k(\eta_{T,t}^{2,k})' \\ &+ \frac{1}{T^{2+\gamma}k^2}\sum_{t=1}^{T-k}x_t^k(\eta_{T,t}^{1,k})'C \\ &+ \frac{1}{T^{2+\gamma}k^2}\sum_{t=1}^{T-k}x_t^k(\eta_{T,t}^{3,k})',\end{aligned} \tag{4.3.9}$$

where all four terms contribute to the asymptotics of the denominator. Combining asymptotic approximations in (4.3.10) and (4.3.8), the limit distribution of the LHIVX estimator is collected in the following theorem.

Theorem 4.3.3 *If* $\frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0$,

$$\sqrt{T}k^{\frac{3}{2}}\left(\widehat{B^*_1}^{LHIVX} - B_1^*\right)' \rightsquigarrow \mathcal{MN}\left(0,\ (\Upsilon^{-1})\left(V_{zz}^{LH} + V_{\eta\eta}^{(2),LH} + V_{z\eta}^{(2),LH} + (V_{z\eta}^{(2),LH})'\right)\left(\Upsilon^{-1}\right)' \cdot \Omega_{00}\right),$$

where expressions of V_{zz}^{LH}, $V_{\eta\eta}^{(2),LH}$, $V_{z\eta}^{(2),LH}$, *and* Υ *are collected in the Appendix.*

The LHIVX estimator is consistent, asymptotically unbiased, and with a mixed normal limit distribution. The distribution is nonpivotal as both coefficients A, and C occur in the variance matrix. However, given the mixed normality and consistent estimate of variance, Wald tests have standard chi-square distributions. The feature as mentioned above again demonstrates the critical advantages of IVX type estimation in predictive regression that were emphasized by Kostakis et al. (2015) over procedures that rely on simulations and restrictions on scalar regressor formulations.

Theorem 4.3.4 *Under $\mathcal{H}_0'(k) : HB_1^* = h$, where H is a known $\ell \times n$ matrix of rank ℓ and h is a known ℓ-vector,*

$$W_T^{LH} := \left(H \left(\widehat{B^*}_1^{LHIVX} \right)' - h \right)' \left[H \left\{ \left(X' M_{\widetilde{Z}} X \right)^{-1} \widehat{\Omega_{00}} \right\} H' \right]^{-1} \left(H \left(\widehat{B^*}_1^{LHIVX} \right)' - h \right) \rightsquigarrow \chi^2(\ell),$$

where

$$(X' K_{\widetilde{Z}} X)^{-1} = \left\{ \left(\sum_{t=1}^{T-k} x_t^k (\widetilde{z}_t^k)' \right) \left(\sum_{t=1}^{T-k} (\widetilde{z}_t^k)(\widetilde{z}_t^k)' \right)^{-1} \left(\sum_{t=1}^{T-k} x_t^k (\widetilde{z}_t^k)' \right)' \right\}^{-1},$$

and $\widehat{\Omega_{00}}$ is any consistent estimator for Ω_{00}.

4.4 IVX Regression with Mixed Roots

This section applies the IVX procedures to the predictive regression models containing both mixed roots and random departures. Under this case, we prove the robustness of pivotal chi-square distributions for the corresponding test statistics. Besides, we propose a method to detect the randomness in the autoregressive coefficients.

4.4.1 Detection on randomness of autoregression coefficient

For simplicity, we consider $n = q = p = 1$ case where $C = c_1$ and $\breve{D}_{at} = a_1 \cdot u_{at}$, and the dimensions of x_t and u_{at} are both equal to 1. The difference between the LUR and LSTUR processes depends on the existence of u_{at}. To consistently select the model, we employ an exogenous instrumental variable to detect the presence of u_{at}. The reason to use IV estimate is that the OLS estimate for a_1 is inconsistent due to the endogeneity issue (i.e. u_{at} and u_{xt} are correlated). When estimating a_1 with an exogenous variable Z_t, we have the consistent estimate $\widehat{a}_1$.

In (4.2.2), the OLS estimate of distance parameter c_1 is also inconsistent. Unfortunately, the limiting behavior of $\widehat{a}_1$ relies on the true value of c_1. As an alternative, we develop a two-step IV procedure for a_1. As the IV-based test has a nonpivotal distribution, we employ Bonferroni corrections to compute the confidence intervals of the statistic.

The two-step IV procedure calls upon the following restrictions.

Assumption 4 *Assume Z_t is one $q \times 1$ vector of instruments for u_{at}. For all t,*

$$\mathbb{E}(Z_t) = 0,\ \mathbb{E}(Z_t)^2 < \infty,\ \mathbb{E}(Z_t u_{xt}^{'}) = 0,\ \Omega_{za} := \mathbb{E}(Z_t u_{at}^{'}) \text{ has full rank } p.$$

We only consider the case where $n = p = q = 1$. In the first step of our procedure, we apply the IV estimates of a_1 to the linear model,

$$x_t - x_{t-1} = c_1 \left(\frac{x_{t-1}}{T}\right) + a_1 \left(\frac{u_{at}x_{t-1}}{\sqrt{T}}\right) + (a_1)^2 \left(\frac{(u_{at})^2 x_{t-1}}{T}\right) + u_{xt},$$

and derive the estimator $\widehat{a}_1^{IV}$ as,

$$\widehat{a}_1^{IV} := \frac{\sum_{t=1}^T z_t(x_t - x_{t-1})}{\sum_{t=1}^T z_t \frac{u_{at}x_{t-1}}{\sqrt{T}}},$$

where we write Z_t as z_t for the case of $p = 1$. As $T \to \infty$, the IV estimate $\widehat{a}_1^{IV}$ follows a nonstandard distribution:

$$\sqrt{T}(\widehat{a}_1^{IV} - a_1)\Omega_{za} \rightsquigarrow \frac{B_{zx}(1)}{\int_0^1 G_{a_1,c_1}(r)dr}, \tag{4.4.1}$$

where $B_{zx}(1)$ is a Brownian motion. In the second step, we plug the consistent estimate $\widehat{a}_1^{IV}$ back into the autoregressive model,

$$\begin{aligned} & x_t - x_{t-1} - \widehat{a}_1^{IV} \left(\frac{u_{at}x_{t-1}}{\sqrt{T}}\right) - (\widehat{a}_1^{IV})^2 \left(\frac{(u_{at})^2 x_{t-1}}{T}\right) && (4.4.2) \\ = \ & c_1 \frac{x_{t-1}}{T} + (a_1 - \widehat{a}_1^{IV}) \left(\frac{u_{at}x_{t-1}}{\sqrt{T}}\right) + \left((a)_1^2 - (\widehat{a}_1^{IV})^2\right) \left(\frac{(u_{at})^2 x_{t-1}}{T}\right) + u_{xt}. \end{aligned}$$

In (4.4.2), the LHS of the equation, $x_t - x_{t-1} - \widehat{a}_1^{IV} \left(\frac{u_{at}x_{t-1}}{\sqrt{T}}\right) - (\widehat{a}_1^{IV})^2 \left(\frac{(u_{at})^2 x_{t-1}}{T}\right)$ is known to us, so we can run OLS to derive the estimate for c_1. It is trivial to check the limiting distribution of $\widehat{c}_1$ as follows,

$$\widehat{c}_1 - c_1 \rightsquigarrow \frac{\int_0^1 J_{c_1}(r)dW_x(r)}{\int_0^1 J_{c_1}^2(r)dr}, \tag{4.4.3}$$

where $T \to \infty$. Based on (4.4.1), (4.4.2) and (4.4.3), we propose the following test statistic,

$$R_{a_1} := \sqrt{T}\widehat{a}_1^{IV}\Omega_{za}.$$

Under the null hypothesis $\mathcal{H}_0:\ a_1 = 0$, the limitng distribution of R_{a_1} is $\frac{B_{zx}}{\int_0^1 J_{c_1}(r)dr}$, where $J_{c_1}(r) = c_1 J_{c_1}(r)dr + dW_x(r)$, and $W_x(r)$ is the Brownian motion induced by u_{xt}.

We cannot compute the confidence interval for $\widehat{a}_1$ directly because of the inconsistent estimate $\widehat{c}_1$. In this paper, we borrow the idea from Cavanagh et al. (1995) and construct the confidence interval of R_{a_1} using Bonferroni corrections.

To construct a Bonferroni confidence interval for R_{a_1}, we first construct a $100(1-\epsilon_1)\%$ confidence interval for c_1, denoted as $C_{c_1}(\epsilon_1)$. We then construct a $100(1-\epsilon_2)\%$ confidence interval for a_1 given c_1, denoted as $C_{a_1|c_1}(\epsilon_2)$. A confidence interval that does not depend on c_1 can be obtained by

$$C_{a_1}(\epsilon) = \bigcup_{c_1 \in C_{c_1(\epsilon_1)}} C_{a_1|c_1}(\epsilon_2).$$

By Bonferroni's inequality, this confidence interval has coverage of at least $100(1-\epsilon)\%$, where $\epsilon = \epsilon_1 + \epsilon_2$.

When the distance parameter c_1 is very small and sample size T are both large, there is negligible difference between the STUR and LSTUR model. At this case, there is no harm to replace the statistic R_{a_1} by the IV-assisted test of Lieberman and Phillips (2018), a statistic for testing the local STUR model against a simple UR null. In the empirical part, we directly employ the test of Lieberman and Phillips (2018) to detect the randomness of autoregressive roots.

4.4.2 IVX regression with mixed roots in short-horizon case

The model is discussed assuming $n = 2$ and $p = 1$. We consider y_t as one scalar and x_{t-1} as one bivariate AR(1) process with both mixed roots and random departures from unity. The simplified predictability system is given as,

$$\begin{aligned} y_t &= \beta' x_{t-1} + u_{0t},\ \beta' = [\beta_1, \beta_2],\ x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}, \\ x_t &= R_{Tt} x_{t-1} + u_{xt},\ R_{Tt} = \begin{bmatrix} \rho_{Tt} & 0 \\ 0 & \theta_T \end{bmatrix},\ u_{xt} = \begin{bmatrix} u_{x1t} \\ u_{x2t} \end{bmatrix}. \end{aligned} \tag{4.4.4}$$

The paper imposes mixed roots and stochastic departures from unit root as

$$\rho_{Tt} = 1 + \frac{c_1}{T^{\alpha_1}} + \frac{a_1 u_{at}}{\sqrt{T}} + \frac{(a_1 u_{at})^2}{T},\ \text{where} \begin{cases} c_1 \in [-\infty, \infty] & \alpha_1 = 1 \\ c_1 \in (-\infty, 0) & \alpha_1 \in (0,1) \end{cases} \text{and } a_1 \in (-\infty, +\infty), \tag{4.4.5}$$

$$\theta_T = 1 + \frac{c_2}{T^{\alpha_2}},\ \text{where } c_2 \in (0, \infty) \text{ and } \alpha_2 \in (0,1). \tag{4.4.6}$$

Accordingly, x_{1t} falls within one of the following specifications as unit root (UR), LUR, STUR, LSTUR, or MIR regressors, while x_{2t} is a mildly explosive (MER) regressor. The innovation process follows,

$$u_t = \begin{bmatrix} u_{0t} \\ u_{x1t} \\ u_{x2t} \\ u_{at} \end{bmatrix} = \begin{bmatrix} F_0(1) & 1\times 4 \\ F_{x1}(1) & 1\times 4 \\ F_{x2}(1) & 1\times 4 \\ F_a(1) & 1\times 4 \end{bmatrix} \begin{bmatrix} \epsilon_{0t} \\ \epsilon_{x1t} \\ \epsilon_{x2t} \\ \epsilon_{at} \end{bmatrix} - \Delta \begin{bmatrix} \widetilde{\epsilon}_{0t} \\ \widetilde{\epsilon}_{x1t} \\ \widetilde{\epsilon}_{x2t} \\ \widetilde{\epsilon}_{at} \end{bmatrix}, \tag{4.4.7}$$

and the long-run variance and the limit theory are the same except that the subscript $0, 1, 2$, and a now signify u_{0t}, u_{x1t}, u_{x2t} and u_{at}. The strict exogeneity holds for (4.4.7) as $\mathbb{E}(u_{at} u_{1t}) = 0$. The IVX instrument has the identical structure as $C_z := diag\{c_{z1}, c_{z2}\} < 0$ and $\gamma = 0.95$.

When $\mathcal{H}_0 : a_1 = 0$ cannot be rejected, UR or LUR regressors rather than S-

TUR or LSTUR regressors need to be considered. Such case has been discussed extensively in Phillips and Lee (2013). If the statistic R_{a_1} rejects the null hypothesis $\mathcal{H}_0 : a_1 = 0$, there is strong evidence of stochastic deviation from unit-roots. Therefore, we prefer STUR/LSTUR regressors.

We provide the pivotal chi-square distribution of the corresponding testing statistic under the null hypothesis of no predictability: We assume $\alpha_1 \in (\frac{1}{3}, 1)$, $\gamma \in ((\alpha_2 \vee \frac{2}{3}),\ 1)$, and $a_1 \neq 0$. Under the null hypothesis $\mathcal{H}_0 : \beta = \beta_0$, as $T \to \infty$,

$$(\widetilde{\beta} - \beta_0)' \left[(X' M_{\widetilde{Z}} X)^{-1} \widehat{\Omega_{00}} \right]^{-1} (\widetilde{\beta} - \beta_0) \rightsquigarrow \chi^2(2),$$

where $\widehat{\Omega_{00}}$ is any consistent estimator for Ω_{00}, and

$$(X' M_{\widetilde{Z}} X) = \left\{ \left(\sum_{t=1}^{T} x_{t-1} \widetilde{z}_{t-1}' \right) \left(\sum_{t=1}^{T} \widetilde{z}_{t-1} \widetilde{z}_{t-1}' \right)^{-1} \left(\sum_{t=1}^{T} x_{t-1} \widetilde{z}_{t-1}' \right) \right\}.$$

Similarly, the results of QR-IVX estimates are collected here. With persistent regressors satisfying both mixed roots and random coefficients, the QR-IVX estimate still follows asymptotic normality. Moreover, we show the pivotal distribution of the Wald test under the null hypothesis of no predictability.

We assume $\alpha_1 \in (\frac{1}{3}, 1)$, $\gamma \in ((\alpha_2 \vee \frac{2}{3}),\ 1)$, and $a_1 \neq 0$. For any given τ, under the null hypothesis $\mathcal{H}_0 : \beta_\tau = \beta_{\tau 0}$,

$$\frac{\widehat{p_{u0\tau}(0)}^2}{\tau(1-\tau)} (\widetilde{\beta}_\tau^{QR-IVX} - \beta_{\tau 0})' (X' M_{\widetilde{Z}} X) (\widetilde{\beta}_\tau^{QR-IVX} - \beta_{\tau 0}) \rightsquigarrow \chi^2(2),$$

where $\widehat{p_{u0\tau}(0)}$ is any consistent estimator for $p_{u0\tau}(0)$, and

$$(X' M_{\widetilde{Z}} X) = \left\{ \left(\sum_{t=1}^{T} x_{t-1} \widetilde{z}_{t-1}' \right) \left(\sum_{t=1}^{T} \widetilde{z}_{t-1} \widetilde{z}_{t-1}' \right)^{-1} \left(\sum_{t=1}^{T} x_{t-1} \widetilde{z}_{t-1}' \right) \right\}.$$

The pivotal chi-square distribution illustrates the robustness of the IVX methodology under the circumstance of mixed roots and random deviations from the unity.

4.4.3 IVX regression with mixed roots in long-horizon case

Similar to short-horizon predictive regressions with mixed roots and random departures from unity, the long-horizon model follows (4.4.4) (4.4.5) and (4.4.6).

The long-run variance and the limit theory are the same except that the subscript $0, 1, 2,$ and a now signify u_{0t}, u_{x1t}, u_{x2t} and u_{at}. The LHIVX formula follows

$$\widehat{b}^{LH} - b = \begin{bmatrix} \sum_{t=1}^{T-k} \widetilde{z}_{1t}^{k} x_{1t}^{k} & \sum_{t=1}^{T-k} \widetilde{z}_{1t}^{k} x_{2t}^{k} \\ \sum_{t=1}^{T-k} \widetilde{z}_{2t}^{k} x_{1t}^{k} & \sum_{t=1}^{T-k} \widetilde{z}_{2t}^{k} x_{2t}^{k} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T-k} \widetilde{z}_{1t} u_{0,t+k} \\ \sum_{t=1}^{T-k} \widetilde{z}_{2t} u_{0,t+k} \end{bmatrix}.$$

Under the null hypothesis of no predictability, the limit distribution of the Wald test based on LHIVX estimates is pivotal. Again, we verify the robustness of the self-generated instruments.

We assume that $\frac{\sqrt{T}}{T^{\gamma}} + \frac{T^{\gamma}}{k} + \frac{k}{T} \to 0$ with $k = T^{\upsilon}$ and

$$\min\{1 + \alpha_1 - \alpha_2, 1 + \gamma - \alpha_2\} > \upsilon > \alpha_1 + \alpha_2 - 1.$$

Under the null hypothesis $\mathcal{H}_0'(k) : b = b_0$,

$$(\widehat{b}^{LH} - b_0)' \left[(X' M_{\widetilde{Z}} X)^{-1} \widehat{\Omega_{00}} \right]^{-1} (\widehat{b}^{LH} - b_0) \rightsquigarrow \chi^2(2),$$

where

$$(X' M_{\widetilde{Z}} X) = \left\{ \left(\sum_{t=1}^{T-k} x_{t-1}^{k} (\widetilde{z}_{t-1}^{k})' \right) \left(\sum_{t=1}^{T-k} \widetilde{z}_{t-1}^{k} (\widetilde{z}_{t-1}^{k})' \right)^{-1} \left(\sum_{t=1}^{T-k} (\widetilde{z}_{t-1}^{k})' x_{t-1}^{k} \right) \right\},$$

with any consistent estimator $\widehat{\Omega_{00}}$ for Ω_{00}.

4.5 Monte Carlo Simulation

This section presents numerical performances of short-horizon QR-IVX statistics under local power and size criteria. To reduce the computational complexity, this numerical experiment employs the alternative QR algorithm based on the self-generated instrument $\widetilde{z}_t$ (Lee, 2016).[1] The DGP follows (4.4.4), (4.4.5), with

$$u_t = \begin{pmatrix} u_{0t} \\ u_{xt} \\ u_{at} \end{pmatrix} \sim \mathcal{N}(0_{(1+n+p)}, \Omega_{(1+n+p)\times(1+n+p)}).$$

We generate QR-IVX instruments for (4.4.4). we set $C_z = -5 \cdot I_n$ and vary γ to explore the size and local power performance. The sample of size is $T = 200$. The scaling parameter $\gamma \in [0.75, 1)$.

To investigate local power performances, we employ a sequence of local alternatives as $H_{\beta T} : \beta_T = \frac{\beta}{T}$ for integer values of $\beta \in [0, 20]$ and various choices of $\gamma \in [0.75, 1)$.

4.5.1 Single regressor case

Simulations accommodate the LSTUR regressor case ($c_1 \neq 0$ and $a_1 \neq 0$) and the STUR case ($c_1 = 0$ and $a_1 \neq 0$). Accordingly, distance parameters $c_1 \in \{-5, 0, 5\}$ and $a_1 \in \{-10 - 5, 5, 10\}$. The variance matrix of innovations has the form as,

$$\Omega = \begin{pmatrix} 1 & -0.75 & 0.40 \\ -0.75 & 1 & -0.50 \\ 0.40 & -0.50 & 1 \end{pmatrix},$$

where we accommodate the endogeneity case ($\mathbb{E}(u_{at}u_{xt}) \neq 0$).

The following table summarizes the size performance of predictability tests using a short-horizon QR-IVX estimator under the case of STUR regressor with various choices of quantiles τ and persistence parameters γ. The empirical size is the rejection frequency of a chi-square distributed Wald test under the null hypothesis

[1] The details of the QR method are provided in the Appendix. The experiment of this chapter also uses the code of the alternative QR procedure provided by Professor Ji Hyung Lee on his website: https://sites.google.com/site/jihyung412/. This declaration is to honor the contribution of Professor Ji Hyung Lee to this chapter and the originality of his invention.

$\mathcal{H}_0 : \beta = 0$. The nominal test size is 0.05. The sample size $T = 200$. The number of replication is 2,500.

Table 4.1: Empirical Size with STUR Regressor ($T = 200$)

$c_1 = 0$, $a_1 = -10$	τ=0.05	τ=0.1	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=0.95
γ=0.75	0.1100	0.0720	0.0588	0.0460	0.0500	0.0400	0.0364	0.0484	0.0492	0.0784	0.0956
γ=0.80	0.1092	0.0728	0.0624	0.0436	0.0516	0.0412	0.0376	0.0480	0.0496	0.0788	0.0956
γ=0.85	0.1076	0.0708	0.0648	0.0424	0.0568	0.0420	0.0388	0.0516	0.0496	0.0756	0.0936
γ=0.90	0.1116	0.0748	0.0632	0.0436	0.0540	0.0456	0.0404	0.0512	0.0484	0.0740	0.0952
γ=0.95	0.1116	0.0752	0.0628	0.0440	0.0516	0.0480	0.0412	0.0524	0.0472	0.0748	0.0964
$c_1 = 0$, $a_1 = 10$	τ=0.05	τ=0.1	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=0.95
γ=0.75	0.0964	0.0724	0.0540	0.0440	0.0428	0.0412	0.0444	0.0472	0.0608	0.0644	0.1032
γ=0.80	0.0944	0.0732	0.0556	0.0424	0.0408	0.0400	0.0444	0.0508	0.0584	0.0692	0.1004
γ=0.85	0.0936	0.0740	0.0540	0.0428	0.0408	0.0408	0.0452	0.0520	0.0560	0.0720	0.0976
γ=0.90	0.0928	0.0736	0.0576	0.0452	0.0400	0.0404	0.0428	0.0512	0.0528	0.0720	0.0940
γ=0.95	0.0900	0.0724	0.0576	0.0444	0.0384	0.0392	0.0448	0.0524	0.0544	0.0732	0.0956
$c_1 = 0$, $a_1 = 5$	τ=0.05	τ=0.1	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=0.95
γ=0.75	0.0940	0.0800	0.0588	0.0588	0.0552	0.0464	0.0592	0.0592	0.0620	0.0920	0.1212
γ=0.80	0.0996	0.0872	0.0636	0.0640	0.0552	0.0464	0.0616	0.0624	0.0648	0.0928	0.1204
γ=0.85	0.1056	0.0872	0.0632	0.0640	0.0556	0.0484	0.0660	0.0672	0.0644	0.0908	0.1168
γ=0.90	0.1076	0.0840	0.0648	0.0716	0.0632	0.0536	0.0680	0.0704	0.0660	0.0936	0.1176
γ=0.95	0.1124	0.0852	0.0660	0.0712	0.0672	0.0548	0.0676	0.0712	0.0684	0.0952	0.1184
$c_1 = 0$, $a_1 = -5$	τ=0.05	τ=0.1	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=0.95
γ=0.75	0.1132	0.0812	0.0700	0.0540	0.0548	0.0552	0.0456	0.0464	0.0544	0.0812	0.0912
γ=0.80	0.1160	0.0828	0.0692	0.0560	0.0608	0.0580	0.0532	0.0472	0.0600	0.0880	0.0916
γ=0.85	0.1132	0.0828	0.0712	0.0580	0.0592	0.0576	0.0548	0.0524	0.0660	0.0892	0.0964
γ=0.90	0.1208	0.0856	0.0772	0.0584	0.0652	0.0576	0.0552	0.0524	0.0660	0.0928	0.1012
γ=0.95	0.1260	0.0884	0.0808	0.0620	0.0672	0.0588	0.0600	0.0568	0.0720	0.0940	0.1016

The size of the QR-IVX test is well controlled, as the empirical size is slightly smaller 0.05. Besides, empirical sizes are well controlled robustly across $\gamma \in [0.75, 1)$. Such properties hold for the LSTUR regressor as well. The details for LSTUR cases are omitted here.

To investigate the local power behavior, we employ a sequence of local alternatives as $\mathcal{H}_{\beta_T} = \frac{\beta}{T}$ for integers values between $\beta \in [0, 20]$ and various $\gamma \in [0.75, 1)$. Our discussions have four cases, $c_1 = -5, 5, -10$, or 10.

The local power functions approach unity in the LSTUR cases with the fastest convergence occurring when $\gamma = 0.95$. Besides, when there is a larger $|a_1|$, the local power function reaches unity more rapidly. This result is not surprising since an LSTUR regressor with a larger $|a_1|$ expects to have stronger signals. Similarly, the above phenomena hold for the STUR case.

These results confirm that the QR-IVX inference procedure is robust for $\gamma \in [0.75, 1)$ and performs well with both STUR and LSTUR regressors.

4.5.2 Multiple regressor case

This subsection considers short-horizon quantile predictive regressions with multiple regressors in which the STUR and LSTUR regressors are included.

Three examples are considered: (i) STUR and MIR regressors ($c_1 = 0$, $a_1 = 10$, $c_2 = \{-5, -2\}$ and $\alpha_2 \in \{0.25, 0.50, 0.75\}$); (ii) STUR and LUR ($c_1 = 0$, $a_1 = 10$, $c_2 = \{-5, -2, 0, 2, 5\}$ and $\alpha_2 = 1$); (iii) STUR and MER regressors ($c_1 = 0$, $a_1 = 10$, $c_2 = \{1, 2, 5\}$ and $\alpha_2 \in \{0.25, 0.50\}$). Case (i) represents the normal periods, while Case (iii) stands for the bubble period, and Case (ii) demonstrates the bubble collapse period. The model setup is given as (4.4.4) with the following covariance matrix as

$$\Omega = \begin{pmatrix} 1 & -0.75 & -0.4 & -0.2 \\ -0.75 & 1 & 0.5 & 0.20 \\ -0.4 & 0.5 & 1 & 0.15 \\ -0.2 & 0.20 & 0.15 & 1 \end{pmatrix}.$$

We accommodate the endogeneity case ($\mathbb{E}(u_{at}u_{xt}) \neq 0$) in the covariance structure. For simplicity, we only consider the STUR case where $c_1 = 0$ and $a_1 = 10$.

The following table reports size performances in testing $\mathcal{H}_0 : \beta_1 = \beta_2 = 0$ using the short-horizon QR-IVX test statistics with STUR & MER regressors or STUR & LUR regressors. In this table, we briefly list several selected cases of all DGPs. The IVX persistence parameter γ is selected from $[0.75, 1)$. The localizing coefficients $C_z = \{-5, 5\}$. The nominal size is 0.05. The sample size is selected as 200. The number of replication is 2,500.

Similar results are observed for Case (iii) with both STUR & MER regressors. In this case, the size distortions of the short-horizon QR test are much more substantial than the short-horizon QR-IVX, although the sizes of short-horizon QR-IVX test

Table 4.2: Empirical Size with STUR & MIR or STUR & LUR Regressors ($T = 200$)

$c1 = 0$, $a1 = 10$, $c2 = -5$, $\alpha 2 = 0.25$	τ=0.05	τ=0.1	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=0.95
γ=0.75	0.1556	0.0972	0.0636	0.0488	0.0416	0.0408	0.0368	0.0468	0.0484	0.0892	0.1396
γ=0.80	0.1572	0.0984	0.0652	0.0472	0.0420	0.0408	0.0344	0.0464	0.0480	0.0872	0.1412
γ=0.85	0.1596	0.1008	0.0612	0.0496	0.0440	0.0384	0.0352	0.0476	0.0472	0.0884	0.1388
γ=0.90	0.1604	0.0988	0.0636	0.0492	0.0460	0.0408	0.0360	0.0452	0.0496	0.0892	0.1408
γ=0.95	0.1636	0.0968	0.0652	0.0476	0.0500	0.0392	0.0376	0.0444	0.0472	0.0852	0.1380
$c1 = 0$, $a1 = 10$, $c2 = -5$, $\alpha 2 = 0.5$	τ=0.05	τ=0.1	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=0.95
γ=0.75	0.1652	0.1056	0.0572	0.0476	0.0484	0.0384	0.0364	0.0424	0.0548	0.0760	0.1368
γ=0.80	0.1636	0.1032	0.0588	0.0508	0.0476	0.0400	0.0376	0.0436	0.0540	0.0772	0.1344
γ=0.85	0.1692	0.1024	0.0588	0.0508	0.0448	0.0400	0.0352	0.0468	0.0556	0.0748	0.1304
γ=0.90	0.1732	0.1044	0.0608	0.0504	0.0452	0.0424	0.0368	0.0468	0.0592	0.0796	0.1376
γ=0.95	0.1696	0.1068	0.0604	0.0488	0.0464	0.0436	0.0396	0.0496	0.0556	0.0796	0.1380
$c1 = 0$, $a1 = 10$, $c2 = -2$, $\alpha 2 = 1$	τ=0.05	τ=0.1	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=0.95
γ=0.75	0.1780	0.1032	0.0716	0.0540	0.0476	0.0520	0.0368	0.0508	0.0572	0.1008	0.1380
γ=0.80	0.1836	0.1064	0.0760	0.0556	0.0508	0.0564	0.0384	0.0580	0.0636	0.0980	0.1364
γ=0.85	0.1876	0.1148	0.0772	0.0636	0.0596	0.0580	0.0388	0.0676	0.0660	0.1076	0.1392
γ=0.90	0.1820	0.1200	0.0848	0.0676	0.0632	0.0652	0.0508	0.0708	0.0704	0.1116	0.1444
γ=0.95	0.1920	0.1300	0.0960	0.0688	0.0728	0.0732	0.0552	0.0820	0.0760	0.1164	0.1568
$c1 = 0$, $a1 = 10$, $c2 = 0$, $\alpha 2 = 1$	τ=0.05	τ=0.1	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=0.95
γ=0.75	0.2012	0.1380	0.0952	0.0680	0.0756	0.0684	0.0584	0.0768	0.0768	0.1176	0.1668
γ=0.80	0.2084	0.1420	0.1072	0.0800	0.0820	0.0772	0.0632	0.0864	0.0820	0.1292	0.1692
γ=0.85	0.2196	0.1560	0.1120	0.0924	0.0896	0.0880	0.0696	0.0988	0.0912	0.1364	0.1844
γ=0.90	0.2316	0.1644	0.1280	0.1044	0.1008	0.0996	0.0800	0.1136	0.1044	0.1488	0.1988
γ=0.95	0.2432	0.1820	0.1424	0.1144	0.1140	0.1084	0.0920	0.1268	0.1152	0.1604	0.2156

might be comparatively larger than Case (i) and (ii).

Table 4.3: Empirical Size with STUR & MER Regressors ($T = 200$)

$c1 = 0$, $a1 = 10$, $c2 = 5$, $\alpha2 = 0.5$	τ=0.05	τ=0.1	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=0.95
γ=0.1500	0.1240	0.0672	0.0516	0.0512	0.0544	0.0604	0.0656	0.0664	0.0996	0.1452	
γ=0.80	0.1552	0.1268	0.0676	0.0532	0.0552	0.0568	0.0552	0.0676	0.0700	0.1016	0.1472
γ=0.85	0.1564	0.1244	0.0692	0.0580	0.0588	0.0580	0.0572	0.0660	0.0736	0.1024	0.1480
γ=0.90	0.1612	0.1272	0.0708	0.0592	0.0644	0.0636	0.0580	0.0656	0.0772	0.1012	0.1532
γ=0.95	0.1704	0.1264	0.0716	0.0621	0.0652	0.0616	0.0576	0.0724	0.0772	0.1048	0.1564
$c1 = 0$, $a1 = 10$, $c2 = 5$, $\alpha2 = 0.25$	τ=0.05	τ=0.1	τ=0.2	τ=0.3	τ=0.4	τ=0.5	τ=0.6	τ=0.7	τ=0.8	τ=0.9	τ=0.95
γ=0.75	0.0956	0.1144	0.0956	0.0856	0.0808	0.0828	0.0756	0.0900	0.1012	0.1040	0.0964
γ=0.80	0.1080	0.1148	0.0944	0.0924	0.0804	0.0860	0.0856	0.0816	0.0864	0.1008	0.0972
γ=0.85	0.0988	0.1144	0.0884	0.0756	0.0864	0.0716	0.0880	0.0872	0.0952	0.0988	0.0976
γ=0.90	0.1008	0.1072	0.0968	0.0864	0.0944	0.0792	0.0768	0.0836	0.0920	0.1000	0.1040
γ=0.95	0.0948	0.1088	0.0888	0.0864	0.0832	0.0852	0.0816	0.0792	0.1036	0.1036	0.0952

With similar local alternatives, the power functions show relatively faster convergence to unity, analog to the single regressor case. In all cases, the power curves are steep. The speeds of convergence in Case (iii) are much faster since explosive roots have more significant signals.

4.6 Empirical Study

We apply our QR-IVX inference procedure to check the predictability of economic or market fundamentals. If we define y_t as the S&P500 excess return and denote x_{t-1} as the economic or market fundamental, the empirical model has the formulation as

$$y_t = \alpha + \beta x_{t-1} + \epsilon_t.$$

To measure market fundamentals, we employ the monthly financial data sets from Welch and Goyal (2008), ranging from February 1920 to December 2017. Following Lee (2016), the excess stock return, y_t is calculated as

$$y_t = \log\left(\frac{P_t + D_t}{P_{t-1}}\right) - \log\left(Rfree_t + 1\right),$$

where P_t and D_t indicate the S&P 500 index returns ($Index$) and dividends ($D12$) at time t. In this paper, the persistent regressor $x_{t-1}(:= tbl_t)$ is the Treasury Bills Rates. The stochastic component, u_{at}, is $corpr_t$, the Long-term Corporate Bond Returns. The instrument, Z_t, is $200 \times \log(BM_t/BM_{t-1})$ where BM_t is the Book-to-Market Ratio (b/m) as the ratio of book value to market value for the Dow Jones Industrial Average. To check the validity of the proposed instrument, the observations for variables Z_t, x_{t-1}, and u_{at} are between January 1926 and December 2017. The sample correlations in the data are $\widehat{\rho}_{a,\Delta x} = -0.31$, $\widehat{\rho}_{a,Z} = -0.17$, and $\widehat{\rho}_{Z,\Delta x} = 0.05$. Since $\widehat{\rho}_{\Delta x,a}$ is far away from zero, the NLLS estimates of the STUR model are not consistent. Observations of sample correlations reinforce the need and validity of the instrument Z_t. By applying the IV-assisted test of Lieberman and Phillips (2018), we statistically justify that tbl is a persistent regressor with a stochastic component $corpr_t$ in the autoregressive slope. As STUR and LSTUR regressors share the identical property, there is no need to distinguish between them.

Table 4.4: Short-horizon QR-IVX Estimations & Test(1920:03-2017:12) with tbl_t for Monthly Data

τ=	0.05	0.1	0.2	0.3	0.4	0.5
estimate	**0.3409**	0.1111	0.1639	0.0852	0.091	**0.1707**
test	**2.0548**	0.9767	1.832	1.2733	1.4579	**2.6879**

Table 4.5: Short-horizon QR-IVX Estimations & Test(1920:03-2017:12) with tbl_t for Monthly Data

τ=	0.6	0.7	0.8	0.9	0.95
estimate	**0.1456**	**0.1306**	0.0738	-0.123	**-0.2455**
test	**2.2611**	**1.9933**	1.0245	-1.3383	**-2.2446**

Table 4.6: Short-horizon QR-IVX Estimations & Test(1921:Q1-2017:Q4) with tbl_t for Quarterly Data

τ=	0.05	0.1	0.2	0.3	0.4	0.5
estimate	0.3739	0.3777	-0.2743	-0.1477	0.0629	0.1102
test	0.536	0.8604	-1.0043	-0.6854	0.3164	0.5831

As shown in Table 4.4 and 4.5, ranging from March 1920 to December 2017, the persistent predictor tbl_t shows significant predictive power at certain quantiles.

τ=	0.6	0.7	0.8	0.9	0.95
estimate	0.1473	0.088	-0.1589	-0.1107	-0.1805
test	0.7803	0.4589	-0.8351	-0.5263	-0.7214

Table 4.7: Short-horizon QR-IVX Estimations & Test(1921:Q1-2017:Q4) with tbl_t for Quarterly Data

The results shown with bold letters imply the rejection rate of the null hypothesis of no predictability at 5% level. For each item, the first line indicates the short-horizon QR-IVX estimates, and the corresponding second line demonstrates the square root of the test statistics. The 5% critical value is 1.96 according to the derived limiting theory. We can easily justify that on 5%, 50%, 60%, 70%, 95% quantiles, the strong predictability of tbl_t on the excess returns is shown in monthly data. Since at extreme quantiles there are severe size distortions, we can only confidently justify that at 50% 60% and 70% quantiles, tbl_t shows strong predictability for monthly data. Related results are demonstrated in Table 4.4 and 4.5. A similar analysis applies to the quarterly data between 1921:Q1 and 2017:Q4 for all quantiles. There is no strong evidence that tbl_t shows predictability for the S&P500 excess return on quarterly data. Results are demonstrated at Table 4.6 and 4.7.

4.7 Conclusion

This paper shows that the IVX instruments developed in Phillips and Magdalinos (2009), the QR-IVX method developed in Lee (2016), and the LHIVX method established in Phillips and Lee (2013) can extend to STUR and LSTUR regressors. Since the Wald test built here has an asymptotic chi-square distribution, no numerical simulations are in need to justify the critical values of the limit distributions. Another advantage of IVX methods is that it can easily extend to the multivariate case with both mixed roots and random departures from unity.

Bibliography

Aitkin, A. C. (1936). IV.-On least squares and linear combination of observations. Proceedings of the Royal Society of Edinburgh, 55, 42-48.

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. The Annals of Mathematical Statistics, 676-687.

Bai, J. (2003). Inferential theory for factor models of large dimensions. Econometrica, 71(1), 135-171.

Bai, J. (2009). Panel data models with interactive fixed effects. Econometrica, 77(4), 1229-1279.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica, 70(1), 191-221.

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.

Bickel, P. J. (1975). One-step Huber estimates in the linear model. Journal of the American Statistical Association, 70(350), 428-434.

Blanchard, O. J., & Watson, M. W. (1982). Bubbles, rational expectations and financial markets (Tech. Rep.). National Bureau of Economic Research.

Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. Econometrica, 83(3), 1147-1184.

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.

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

Cavanagh, C. L., Elliott, G., & Stock, J. H. (1995). Inference in models with nearly integrated regressors. Econometric Theory, 11(5), 1131-1147.

Chen, K., & Wen, Y. (2017). The great housing boom of China. American Economic Journal: Macroeconomics, 9(2), 73-114.

Cochrane, J. H. (1991). Volatility tests and efficient markets: A review essay (Tech.Rep.). National Bureau of Economic Research.

Cogley, T., & Sargent, T. J. (2001). Evolving post-world war II US inflation dynamics. NBER macroeconomics annual, 16, 331-373.

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. Review of Economic Dynamics, 8(2), 262-302.

Diba, B. T., & Grossman, H. I. (1987). On the inception of rational bubbles. The Quarterly Journal of Economics, 102(3), 697-700.

Diba, B. T., & Grossman, H. I. (1988). Explosive rational bubbles in stock prices? The American Economic Review, 78(3), 520-530.

Doob, J. L. (2012). Measure theory (Vol. 143). Springer Science & Business Media.

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.

Fan, R., & Lee, J. H. (2019). Predictive quantile regressions under persistence and conditional heteroskedasticity. Journal of Econometrics, 213(1), 261-280.

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

Freedman, D. A. (1975). On tail probabilities for martingales. the Annals of Probability, 3(1), 100-118.

Granger, C. W., & Swanson, N.R. (1997). An introduction to stochastic unit-root processes. Journal of Econometrics, 80(1), 35-62.

Grenander, U., & Rosenblatt, M. (2008). Statistical analysis of stationary time series (Vol. 320). American Mathematical Soc.

Grout, P. A. (2003). Public and private sector discount rates in public–private partnerships. The Economic Journal, 113(486), C62-C68.

Hahn, J., & Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. Econometrica, 70(4), 1639-1657.

Hannan, E. J. (2009). Multiple time series (Vol. 38). John Wiley & Sons.

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory, 8(4), 489-500.

Hansen, B. E. (1999). The grid bootstrap and the autoregressive model. Review of Economics and Statistics, 81(4), 594-607.

Harris, D. (1996). Fully Modified Least Squares in 1 (2) Regression. Econometric Theory, 12(1), 201-204.

Harrison, G. W., Lau, M. I., & Williams, M. B. (2002). Estimating individual discount rates in Denmark: A field experiment. American Economic Review, 92(5), 1606-1617.

Harvey, D. I., Leybourne, S. J., & Sollis, R. (2015). Recursive right-tailed unit root tests for an explosive asset price bubble. Journal of Financial Econometrics, 13(1), 166-187.

Harvey, D. I., Leybourne, S. J., Sollis, R., & Taylor, A. R. (2016). Tests for explosive financial bubbles in the presence of non-stationary volatility. Journal of Empirical Finance, 38, 548-574.

Homm, U., & Breitung, J. (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. Journal of Financial Econometrics, 10(1), 198-231.

Huang, W., Jin, S., & Su, L. (2019). Panel Cointegration with Latent Group Structures and an Application to the PPP Theory. Econometric Theory, forthcoming.

Jegadeesh, N. (1991). Seasonality in stock price mean reversion: Evidence from the US and the UK. The Journal of Finance, 46(4), 1427-1444.

Kallenberg, O. (2006). Foundations of modern probability. Springer Science & Business Media.

Ke, Z. T., Fan, J., & Wu, Y. (2015). Homogeneity pursuit. Journal of the American Statistical Association, 110(509), 175-194.

Khintchine, A. (1924). Über einen satz der wahrscheinlichkeitsrechnung. Fundamenta Mathematicae, 6(1), 9-20.

Knight, K. (1989). Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. Canadian Journal of Statistics, 17(3), 261-278.

Kolmogoroff, A. (1928). Über die Summen durch den Zufall bestimmter unabhängiger Grössen. Mathematische Annalen, 99(1), 309-319.

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

Krämer, W., & Hassler, U. (1998). Limiting efficiency of OLS vs. GLS when regressors are fractionally integrated. Economics Letters, 60(3), 285-290.

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

Leybourne, S. J., McCabe, B. P., & Mills, T. C. (1996). Randomized unit root processes for modelling and forecasting financial time series: theory and applications. Journal of Forecasting, 15(3), 253-270.

Li, Y., Yu, J., & Zeng, T. (2019). Deviance Information Criterion for Latent Variable Models and Misspecified Models. Journal of Econometrics, 21(6), 450-493.

Lieberman, O., & Phillips, P. C. (2017). A multivariate stochastic unit root model with an application to derivative pricing. Journal of Econometrics, 196(1), 99-110.

Lieberman, O., & Phillips, P. C. (2018). IV and GMM inference in endogenous stochastic unit root models. Econometric Theory, 34(5), 1065-1100.

Lieberman, O., & Phillips, P. C. (2020). Hybrid stochastic local unit roots. Journal of Econometrics, 215(1), 257-285.

Lin, C.-C., & Ng, S. (2012). Estimation of panel data models with parameter heterogeneity when group membership is unknown. Journal of Econometric Methods, 1(1), 42-55.

Liptser, R. S., & Shiryaev, A. N. (2013). Statistics of random processes II: Applications (Vol. 6). Springer Science & Business Media.

Magdalinos, T. (2012). Mildly explosive autoregression under weak and strong dependence. Journal of Econometrics, 169(2), 179-187.

Magdalinos, T., & Phillips, P. C. (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. Econometric Theory, 25(2), 482-526.

Magnus, J. R. (1986). The exact moments of a ratio of quadratic forms in normal variables. Annales d'Economie et de Statistique, 95-109.

Marcinkiewicz, J., & Zygmund, A. (1937). Remarque sur la loi du logarithme itéré Fundamenta Mathematicae, 29(1), 215-222.

McCabe, B. P., & Smith, R.J. (1998). The power of some tests for difference stationarity under local heteroscedastic integration. Journal of the American Statistical Association, 93(442), 751-761.

Narayan, P. K., Mishra, S., Sharma, S., & Liu, R. (2013). Determinants of stock price bubbles. Economic Modelling, 35, 661-667.

Park, J. Y., & Phillips, P. C. (1988). Statistical inference in regressions with integrated processes: Part 1. Econometric Theory, 4(3), 468-497.

Pavlidis, E., Yusupova, A., Paya, I., Peel, D., Martínez-García, E., Mack, A., & Grossman, V. (2016). Episodes of exuberance in housing markets: in search of the smoking gun. The Journal of Real Estate Finance and Economics, 53(4), 419-449.

Perron, P. (1991). A continuous time approximation to the unstable first-order autoregressive process: the case without an intercept. Econometrica, 211-236.

Phillips, P. C. (1987a). Time series regression with a unit root. Econometrica, 277-301.

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

Phillips, P. C. (1989). Partially identified econometric models. Econometric Theory, 5(2), 181-240.

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

Phillips, P. C. (2014b). Dynamic panel GMM with roots near unity (Tech. Rep.). Working Paper, Yale University Technical Report.

Phillips, P. C., & Chang, Y. (1994). Fully modified least squares in I (2) regression. Econometric Theory, 10(5), 967-967.

Phillips, P. C., & Durlauf, S. N. (1986). Multiple time series regression with integrated processes. The Review of Economic Studies, 53(4), 473-495.

Phillips, P. C., & Lee, C. C. (1996). Efficiency gains from quasi-differencing under nonstationarity. Athens Conference on Applied Probability and Time Series Analysis, Springer, 300-314.

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

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

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

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

Phillips, P. C., Shi, S., & Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. International Economic Review, 56(4), 1043-1078.

Phillips, P. C., Shi, S., & Yu, J. (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. International Economic Review, 56(4), 1079-1134.

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

Phillips, P. C., Wu, Y., & Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? International Economic Review, 52(1), 201-226.

Phillips, P. C., & Yu, J. (2009). Limit theory for dating the origination and collapse of mildly explosive periods in time series data. Research Collection School Of Economics 2009-09.

Phillips, P. C., & Yu, J. (2011). Dating the timeline of financial bubbles during the subprime crisis. Quantitative Economics, 2(3), 455-491.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. Econometric Theory, 7(2), 186-199.

Santos, M. S., & Woodford, M. (1997). Rational asset pricing bubbles. Econometrica: Journal of the Econometric Society, 19-57.

Sarafidis, V., & Weber, N. (2015). A partially heterogeneous framework for analyzing panel data. Oxford Bulletin of Economics and Statistics, 77(2), 274-296.

Shi, S. (2017). Speculative bubbles or market fundamentals? An investigation of US regional housing markets. Economic Modelling, 66, 101-111.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(4), 583-639.

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

Stout, W. F. (1970). A martingale analogue of Kolmogorov's law of the iterated logarithm. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 15(4), 279-290.

Su, L., Shi, Z., & Phillips, P. C. (2016). Identifying latent structures in panel data. Econometrica, 84(6), 2215-2264.

Su, L., Wang, W., & Zhang, Y. (2019). Strong consistency of spectral clustering for stochastic block models. IEEE Transactions on Information Theory, 66(1), 324-338.

Tao, Y., Phillips, P. C., & Yu, J. (2019). Random coefficient continuous systems: Testing for extreme sample path behavior. Journal of Econometrics, 209(2), 208-237.

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.

White, H. (2014). Asymptotic theory for econometricians. Academic press.

White, J. S. (1961). Asymptotic expansions for the mean and variance of the serial correlation coefficient. Biometrika, 48(1-2), 85-94.

Wittmann, R. (1985). A general law of iterated logarithm. Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete, 68(4), 521-543.

Xiao, Z. (2009). Quantile cointegrating regression. Journal of Econometrics, 150(2), 248-260.

Xiao, Z., & Phillips, P. C. (2002). Higher order approximations for Wald statistics in time series regressions with integrated processes. Journal of Econometrics, 108(1), 157-198.

Yoon, G. (2006). A note on some properties of STUR processes. Oxford Bulletin of Economics and Statistics, 68(2), 253-260.

Appendix A: Appendix

A.1 Proofs in Chapter 2

A.1.1 Technical lemma in pure explosive model

Lemma A.1.1 *Let $X_{n,ols}$ and $Y_{n,ols}$ be defined as in (2.2.9).*

$$\mathbb{E}(Y_{n,ols}^2) = \frac{a^2\sigma^2}{(a-1)^3(a+1)} + o(1),$$
$$\mathbb{E}(X_{n,ols}^2) = \frac{\sigma^2}{(a-1)^2} + o(1),\ \mathbb{E}(X_nY_n) = o(1).$$

Proof.

$$\begin{aligned}\mathbb{E}(Y_{n,ols}^2) &= \mathbb{E}\left(\sum_{j=1}^n a^{-j}u_j\right)^2 = \sigma^2\left(\sum_{j=1}^n a^{-j}\sum_{i=1}^n a^{-i}\min\{i,j\}\right) \\ &= \sigma^2\sum_{j=1}^n a^{-j}\sum_{i=1}^j a^{-i}i + \sigma^2\sum_{j=1}^n ja^{-j}\sum_{i=j+1}^n a^{-i}. \qquad \text{(A.1.1)}\end{aligned}$$

The first part of Equation (A.1.1) can be computed as,

$$\begin{aligned}&\sigma^2\sum_{j=1}^n a^{-j}\sum_{i=1}^j a^{-i}i \\ =\ &\frac{a^{-1}\sigma^2}{(1-a^{-1})^2}\sum_{j=1}^n a^{-j}\left[1-(j+1)a^{-j}+ja^{-j-1}\right] \\ =\ &\frac{a^{-1}\sigma^2}{(1-a^{-1})^2}\sum_{j=1}^n a^{-j} - \frac{a^{-1}\sigma^2}{(1-a^{-1})^2}\sum_{j=1}^n (j+1)a^{-2j} + \frac{a^{-1}\sigma^2}{(1-a^{-1})^2}\sum_{j=1}^n ja^{-2j-1} \\ =\ &\frac{a^{-1}\sigma^2}{(1-a^{-1})^2}\frac{a^{-1}(1-a^{-n})}{1-a^{-1}} - \frac{a^{-1}\sigma^2}{(1-a^{-1})^2}\frac{a^{-2}(1-a^{-2n})}{1-a^{-2}}\end{aligned}$$

$$
\begin{aligned}
& -\frac{a^{-1}\sigma^2}{(1-a^{-1})}\frac{a^{-2}\left[1-(n+1)a^{-2n}+na^{-2n-2}\right]}{(1-a^{-2})^2} \\
= \quad & \frac{a^{-1}\sigma^2}{(1-a^{-1})^2}\frac{a^{-1}}{1-a^{-1}} - \frac{a^{-1}\sigma^2}{(1-a^{-1})^2}\frac{a^{-2}}{1-a^{-2}} - \frac{a^{-1}\sigma^2}{(1-a^{-1})}\frac{a^{-2}}{(1-a^{-2})^2} + o(1).
\end{aligned}
$$

The second part of Equation (A.1.1) is

$$
\begin{aligned}
\sigma^2 \sum_{j=1}^{n} ja^{-j} \sum_{i=j+1}^{n} a^{-i} \quad &= \quad \sigma^2 \sum_{j=1}^{n} ja^{-j} \frac{a^{-(j+1)}(1-a^{-(n-j)})}{1-a^{-1}} \\
&= \quad \frac{\sigma^2}{1-a^{-1}} \sum_{j=1}^{n} ja^{-2j-1} - \frac{\sigma^2 a^{-n-1}}{1-a^{-1}} \sum_{j=1}^{n} ja^{-j} \\
&= \quad \frac{\sigma^2 a^{-1}}{1-a^{-1}} a^{-2} \frac{1-(n+1)a^{-2n}+na^{-2n-2}}{(1-a^{-2})^2} \\
& \quad\ -\frac{\sigma^2 a^{-n-1}}{1-a^{-1}} \frac{a^{-1}\left[1-(n+1)a^{-n}+na^{-n-1}\right]}{(1-a^{-1})^2} \\
&= \quad \frac{\sigma^2 a^{-1}}{1-a^{-1}} a^{-2} \frac{1}{(1-a^{-2})^2} + o(1).
\end{aligned}
$$

Therefore we have

$$
\mathbb{E}(Y_{n,ols}^2) = \frac{a^2\sigma^2}{(a-1)^3(a+1)} + o(1).
$$

Similarly,

$$
\begin{aligned}
\mathbb{E}(X_{n,ols}^2) \quad &= \quad \frac{1}{n}\mathbb{E}\left(\sum_{j=1}^{n} a^{-(n-j)-1}u_j\right)\left(\sum_{i=1}^{n} a^{-(n-i)-1}u_i\right) \\
&= \quad \frac{1}{n}\mathbb{E}\left(\sum_{j=1}^{n} a^{-j}u_{n-j+1} \sum_{i=1}^{n} a^{-i}u_{n-i+1}\right) \\
&= \quad \frac{1}{n}\sigma^2 \sum_{j=1}^{n} a^{-j} \sum_{i=1}^{n} a^{-i} \min\{n-j+1, n-i+1\} \\
&= \quad \frac{1}{n}\sum_{i=1}^{n} a^{-i} \sum_{j=1}^{n} a^{-j}(n+1) - \frac{1}{n}\sum_{i=1}^{n} a^{-i} \sum_{j=1}^{n} a^{-j} \max\{i,j\} \\
&= \quad \frac{1}{n}\sum_{i=1}^{n} a^{-i} \sum_{j=1}^{n} a^{-j}(n+1) + O\left(\frac{1}{n}\right) \\
&= \quad \frac{n+1}{n} a^{-2} \left(\frac{1-a^{-n}}{1-a^{-1}}\right)^2 \sigma^2 + o(1) \\
&= \quad \frac{a^{-2}\sigma^2}{(1-a^{-1})^2} + o(1).
\end{aligned}
$$

Therefore it is shown that $\mathbb{E}(X_{n,ols}^2) = \frac{\sigma^2}{(a-1)^2} + o_p(1)$. For $\mathbb{E}(X_{n,ols}Y_{n,ols})$, we have

$$\begin{aligned}
\mathbb{E}(X_{n,ols}Y_{n,ols}) &= \frac{1}{\sqrt{n}}\mathbb{E}\left[\sum_{t=1}^{n} a^{-(n-t)-1}u_t \sum_{j=1}^{n} a^{-j}u_j\right] \\
&= \frac{a^{-n-1}\sigma^2}{\sqrt{n}} \sum_{t=1}^{n} a^t \sum_{j=1}^{n} a^{-j}\min\{t,j\} \\
&= \frac{a^{-n-1}\sigma^2}{\sqrt{n}} \sum_{t=1}^{n} a^t \sum_{j=1}^{t} ja^{-j} + \frac{a^{-n-1}\sigma^2}{\sqrt{n}} \sum_{t=1}^{n} ta^t \sum_{j=t+1}^{n} a^{-j}.
\end{aligned}$$

For the first part of $\mathbb{E}(X_{n,ols}Y_{n,ols})$, we have

$$\begin{aligned}
\frac{a^{-n-1}\sigma^2}{\sqrt{n}} \sum_{t=1}^{n} a^t \sum_{j=1}^{t} ja^{-j} &= \frac{a^{-n-1}}{\sqrt{n}} \sum_{t=1}^{n} a^t a^{-1} \frac{1-(t+1)a^{-t}+ta^{-t-1}}{(1-a^{-1})^2} \\
&= \frac{1}{\sqrt{n}} \frac{a^{-n-2}}{(1-a^{-1})^2} \sum_{t=1}^{n} a^t(1-(t+1)a^{-t}+ta^{-t-1}) \\
&= \frac{1}{\sqrt{n}} \frac{a^{-n-2}}{(1-a^{-1})^2} \sum_{t=1}^{n} a^t + o(1) \\
&= \frac{1}{\sqrt{n}} \frac{-a^{-n-1}+a^{-1}}{(1-a^{-1})^2(a-1)} + o(1) \\
&= O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

For the second part of $\mathbb{E}(X_{n,ols}Y_{n,ols})$, we have

$$\begin{aligned}
\frac{a^{-n-1}\sigma^2}{\sqrt{n}} \sum_{t=1}^{n} ta^t \sum_{j=t+1}^{n} a^{-j} &= \frac{1}{\sqrt{n}} a^{-n-1}\left(\sum_{t=1}^{n} ta^t\right) \frac{a^{-t-1}(1-a^{-(n-t)})}{1-a^{-1}} \\
&= \frac{1}{\sqrt{n}} \frac{a^{-n-2}}{1-a^{-1}} \sum_{t=1}^{n} t(1-a^{-(n-t)}) \\
&= \frac{1}{\sqrt{n}} \frac{a^{-n-2}}{1-a^{-1}} \sum_{t=1}^{n} t - \frac{1}{\sqrt{n}} \frac{a^{-2n-2}}{1-a^{-1}} \sum_{t=1}^{n} ta^t \\
&= -\frac{1}{\sqrt{n}} \frac{a^{-2n-2}}{1-a^{-1}} \frac{a\left[1-(n+1)a^n+na^n\right]}{(1-a)^2} + o(1) \\
&= o(1),
\end{aligned}$$

where the exponential rates dominate polynomial rates. Denote $\sigma_X^2 := \frac{\sigma^2}{(a-1)^2}$, and $\sigma_Y^2 := \frac{a^2\sigma^2}{(a-1)^3(a+1)}$. ■

Lemma A.1.2 *As $n \to \infty$,*

$$\frac{1}{a^{2n-1}} \mathbb{E} \sum_{t=1}^{n} \left(\sum_{j=t}^{n} a^{t-1-j} u_j u_t \right) = o_p(1).$$

Proof. The derivations are as

$$\begin{aligned}
& \frac{1}{a^{2n-1}} \mathbb{E} \sum_{t=1}^{n} \left(\sum_{j=t}^{n} a^{t-1-j} u_j u_t \right) \\
= \;& \sigma^2 a^{-2n+1} \sum_{t=1}^{n} \left(\sum_{j=t}^{n} a^{t-1-j} t \right) = \sigma^2 a^{-2n} \sum_{t=1}^{n} t \cdot a^t \sum_{j=t}^{n} a^{-j} \\
= \;& \sigma^2 a^{-2n} \sum_{t=1}^{n} t a^t a^{-t} \frac{1 - a^{-(n-t)}}{1 - a^{-1}} < \frac{\sigma^2 a^{-2n}}{1 - a^{-1}} \sum_{t=1}^{n} t = O\left(\frac{n^2}{a^{2n}} \right) \\
= \;& o(1).
\end{aligned}$$

■

A.1.2 Proof of Lemma 2.2.1

Proof. We complete the proof in three steps: Step (i) shows the asymptotic normality of $Y_{n,ols}$; Step (ii) shows the asymptotic normality of $X_{n,ols}$; Step (iii) demonstrates the joint normality of $X_{n,ols}$ and $Y_{n,ols}$ using Cramér-Wold device (Kallenberg, O., 2006).

(i) As innovations are normally distributed, for any fixed n, the random variable $Y_{n,ols}$ is normally distributed. By Lemma A.1.1, note the fact that $\mathbb{E}(Y_{n,ols}^2) = \frac{a^2\sigma^2}{(a-1)^3(a+1)} + o_p(1)$. By the virtue of Cauchy sequence, for any m and n that diverge with restriction $\frac{n}{m} + \frac{1}{n} \to 0$,

$$|Y_{m,ols} - Y_{n,ols}| = \left| \sum_{j=n+1}^{m} a^{-j} \sum_{i=1}^{j} \epsilon_i \right| \preceq O_p\left(a^{-n} \cdot \sqrt{m} \right) = o_p(1).$$

Note the decomposition that

$$\sum_{j=1}^{n} a^{-j} u_j = \sum_{j=1}^{n} a^{-j} \left(\sum_{i=1}^{j} \epsilon_i \right) = \sum_{i=1}^{n} \left(\sum_{j=i}^{n} a^{-j} \right) \epsilon_i,$$

where $\left(\sum_{j=i}^{n} a^{-j}\right) \epsilon_i$, for $i = 1, 2, ..., n$, is a martingale difference sequence. Therefore the martingale convergence theorem is applied and convergence of $Y_{n,ols}$ is shown below:

$$Y_{n,ols} \overset{a.s}{\to} Y =_d \mathcal{N}\left(0, \frac{a^2\sigma^2}{(a-1)^3(a+1)}\right).$$

(ii) For $X_{n,ols}$,

$$X_{n,ols} := \sum_{t=1}^{n} \left\{ \left[\frac{1}{\sqrt{n}} a^{-(n-t)-1} \right] \sum_{s=1}^{t} \epsilon_s \right\} = \sum_{s=1}^{n} \left\{ \sum_{t=s}^{n} \left[\frac{1}{\sqrt{n}} a^{-(n-t)-1} \right] \epsilon_s \right\}.$$

Let $\zeta_{ns} := \sum_{t=s}^{n} \left[\frac{1}{\sqrt{n}} a^{-(n-t)-1} \right] \epsilon_s$ so that $X_{n,ols} = \sum_{s=1}^{n} \zeta_{ns}$. Note that ζ_{ns} is an independent but not indentically distributed sequence. The Lindeberg-Feller central limit theorem (Kallenberg, O., 2006) can be applied to obtain the limiting distribution of $X_{n,ols}$. The stability condition is provided in Lemma A.1.1. To check the Lindeberg condition, $\forall \eta > 0$,

$$\begin{aligned}
& \sum_{s=1}^{n} \mathbb{E}(\zeta_{ns}^2 1_{|\zeta_{ns}|>\eta}) \\
= & \sum_{s=1}^{n} \left(\sum_{t=s+1}^{n} \left[\frac{1}{\sqrt{n}} a^{-(n-t)-1} \right] \right)^2 \mathbb{E}\left(\epsilon_s^2 1_{\left\{ \left| \sum_{t=s}^{n} \left[\frac{1}{\sqrt{n}} a^{-(n-t)-1} \right] \right| |\epsilon_s| > \eta \right\}} \right) \\
\leq & \sum_{s=1}^{n} \left(\sum_{t=s}^{n} \left[\frac{1}{\sqrt{n}} a^{-(n-t)-1} \right] \right)^2 \mathbb{E}\left[\epsilon_s^2 1_{\left\{ \left[\sum_{t=1}^{n} \frac{1}{\sqrt{n}} a^{-(n-t)-1} \right]^2 \epsilon_s^2 > \eta^2 \right\}} \right] \\
\leq & K \max_{1 \leq s \leq n} \mathbb{E}\left[\epsilon_s^2 1_{\left\{ \left[\sum_{t=1}^{n} \frac{1}{\sqrt{n}} a^{-(n-t)-1} \right]^2 \epsilon_s^2 > \eta^2 \right\}} \right],
\end{aligned}$$

where some $K < \infty$. The last step follows $\sum_{s=1}^{n} \left(\sum_{t=s}^{n} \left[\frac{1}{\sqrt{n}} a^{-(n-t)-1} \right] \right)^2 < K$ because $\sum_{t=s}^{n} a^{-2(n-t)-2} = a^{-2s} \frac{1-a^{-2(n-s)}}{1-a^{-2}} \leq \frac{a^{-2}}{(1-a^{-2})}$. Consequently, we have

$$\begin{aligned}
& \sum_{s=1}^{n} \mathbb{E}(\zeta_{ns}^2 1_{|\zeta_{ns}|>\eta}) \leq K \max_{1 \leq s \leq n} \mathbb{E}\left[\epsilon_s^2 1_{\left\{ \left[\sum_{t=s}^{n} \frac{1}{\sqrt{n}} a^{-(n-t)-1} \right]^2 \epsilon_s^2 > \eta^2 \right\}} \right] \\
\leq & K \max_{1 \leq s \leq n} \mathbb{E}\left(\epsilon_s^2 1_{\left\{ \frac{2a^{-2}}{n(1-a^{-2})}) \epsilon_i^2 > \eta^2/2 \right\}} \right) \to 0.
\end{aligned}$$

As $n \to \infty$, $X_n \overset{d}{\to} X =_d \mathcal{N}\left(0, \frac{\sigma^2}{(a-1)^2}\right)$.

Case (iii) by the Cramér-Wold device, the following condition is needed to show the convergence to joint normality,

$$pX_{n,ols} + qY_{n,ols} \xrightarrow{d} pX + qY,$$

where Y and X are two independent normal variables that follow $\mathcal{N}\left(0, \frac{a^2\sigma^2}{(a-1)^3(a+1)}\right)$ and $\mathcal{N}\left(0, \frac{\sigma^2}{(a-1)^2}\right)$ respectively and p, q are any two real numbers. If we define Z as $Z =_d \mathcal{N}\left(0, q^2\frac{a^2\sigma^2}{(a-1)^3(a+1)} + p^2\frac{\sigma^2}{(a-1)^2}\right)$, then $pX_{n,ols} + qY_{n,ols} \xrightarrow{d} Z$. We can write $pX_{n,ols} + qY_{n,ols} = \sum_{t=1}^{T} \varrho_t$ where the array

$$\begin{aligned}
\varrho_t &:= p\sum_{j=t}^{n} a^{-j}\epsilon_t + \frac{q}{\sqrt{n}}\sum_{i=t}^{n} a^{-(n-i)-1}\epsilon_t \\
&= \sum_{j=t}^{n}\left(pa^{-j} + \frac{q}{\sqrt{n}}a^{-(n-j)-1}\right)\epsilon_t,
\end{aligned}$$

consists of independent and non-identically distributed random variables. As we assume the joint normality distribution for ϵ_t, then the term $pX_{n,ols} + qY_{n,ols}$ is also normally distributed as

$$pX_{n,ols} + qY_{n,ols} \sim \mathcal{N}\left(0, p^2\mathbb{E}X_{n,ols}^2 + q^2\mathbb{E}Y_{n,ols}^2 + 2pq\mathbb{E}X_{n,ols}Y_{n,ols}\right).$$

By Lemma A.1.1, we have asymptotic uncorrelation for $X_{n,ols}$ and $Y_{n,ols}$. The argument of Cauchy sequence and martingale convergence theorem in (i) can be applied again to shown convergence to the Gaussian variate Z. As for the joint normality, the asymptotic uncorrelation implies asymptotic independence:

$$pX_{n,ols} + qY_{n,ols} \xrightarrow{d} \mathcal{N}\left(0, q^2\frac{a^2\sigma^2}{(a-1)^3(a+1)} + p^2\frac{\sigma^2}{(a-1)^2}\right).$$

We complete the proof. ■

A.1.3 Proof of Theorem 2.2.2

Proof. The explosive model has the following decomposition.

$$a^{-2n}\sum_{t=1}^{n} y_{t-1}^2 = \frac{1}{a^2-1}\left\{a^{-2n}(y_n^2-y_0^2) - 2a^{-2n+1}\sum_{t=1}^{n} y_{t-1}u_t - a^{-2n}\sum_{t=1}^{n} u_t^2\right\},$$

where

$$\begin{aligned} a^{-2n+1}\sum_{t=1}^{n} y_{t-1}u_t &= y_0 a^{-n}\sum_{t=1}^{n} a^{-(n-t)}u_t + a^{-2n+1}\sum_{t=1}^{n}\left(\sum_{j=1}^{t-1} a^{t-1-j}u_j\right)u_t \\ &= O_p(a^{-n}) + a^{-n+1}\sqrt{n}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} a^{-(n-t)-1}u_t\right)\left(\sum_{j=1}^{n} a^{-j}u_j\right) + o_p(1) \\ &= O_p(a^{-n}\sqrt{n}) + O_p(a^{-n}) = o_p(1). \end{aligned}$$

Besides, $a^{-2n}\sum_{t=1}^{n} u_t^2 = O_p\left(\frac{n^2}{a^{-2n}}\right)$. In all, we have the following approximation,

$$\begin{aligned} a^{-2n}\sum_{t=1}^{n} y_{t-1}^2 &= \frac{1}{a^2-1}(a^{-n}y_n)^2 + o_p(1) = \frac{1}{(a^2-1)}\left(\sum_{j=1}^{n} a^{-j}u_j\right)^2 + o_p(1) \\ &= \frac{1}{(a^2-1)}Y_{n,ols}^2 + o_p(1). \end{aligned}$$

For the term $\frac{a^{-n}}{\sqrt{n}}\sum_{t=1}^{n} y_{t-1}u_t$,

$$\begin{aligned} \frac{a^{-n}}{\sqrt{n}}\sum_{t=1}^{n} y_{t-1}u_t &= \frac{y_0}{\sqrt{n}}\sum_{t=1}^{n} a^{-(n-t)}u_t + \frac{a^{-n}}{\sqrt{n}}\sum_{t=1}^{n}\left(\sum_{j=1}^{t-1} a^{t-1-j}u_j\right)u_t \\ &= \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} a^{-(n-t)-1}u_t\right)\left(\sum_{t=1}^{n} a^{-j}u_j\right) + o_p(1) \\ &= X_{n,ols}Y_{n,ols} + o_p(1). \end{aligned}$$

■

A.1.4 Technical lemmas in mildly explosive model

The paper defines $X_n := \frac{1}{k_n\sqrt{n}}\sum_{t=1}^{n} \rho_n^{-(n-t)-1}u_t$, and $Y_n := \frac{1}{k_n^{\frac{3}{2}}}\sum_{j=1}^{n} \rho_n^{-j}u_j$.

Lemma A.1.3 *As $n \to \infty$, we have $\rho_n^{-n} \cdot \frac{n}{k_n} = o(1)$, $\rho_n^{-n} \cdot \frac{n^2}{k_n^2} = o(1)$, and $\rho_n^{-n} \cdot \frac{\sqrt{n}}{\sqrt{k_n}} = o(1)$.*

Proof. For $\rho_n^{-n} \cdot \frac{n}{k_n} = o(1)$, see Phillips and Magdalinos (2007). For the other two results,

$$\begin{aligned}
\log\left(\rho_n^{-n} \cdot \frac{n^2}{k_n^2}\right) &= -n \log \rho_n + 2\log\left(\frac{n}{k_n}\right) = -n \log\left(1 + \frac{c}{k_n}\right) + 2\log\left(\frac{n}{k_n}\right) \\
&= -n\left[\frac{c}{k_n} + O\left(\frac{1}{k_n^2}\right)\right] + 2\log\left(\frac{n}{k_n}\right) \\
&= -\frac{nc}{k_n}\left[1 + \frac{2}{c}\frac{\log(n/k_n)}{n/k_n} + O\left(\frac{1}{k_n}\right)\right] = -\frac{cn}{k_n}\left[1 + o(1)\right] \\
&= o(1).
\end{aligned}$$

Similarly we prove $\rho_n^{-n} \cdot \frac{\sqrt{n}}{\sqrt{k_n}} = o(1)$. ■

Lemma A.1.4 *As $n \to \infty$, the following results are shown,*

(1)$\frac{1}{k_n}\sum_{j=1}^n \rho_n^{-j} = O(1)$, and $\lim_{n\to\infty} \frac{1}{k_n}\sum_{j=1}^n \rho_n^{-j} = \frac{1}{c}$,

(2)$\frac{1}{k_n}\sum_{j=1}^n \rho_n^{-2j} = O(1)$, and $\lim_{n\to\infty} \frac{1}{k_n}\sum_{j=1}^n \rho_n^{-2j} = \frac{1}{2c}$,

(3)$\frac{1}{k_n^2}\sum_{j=1}^n j\rho_n^{-j} = O(1)$, and $\lim_{n\to\infty} \frac{1}{k_n^2}\sum_{j=1}^n j\rho_n^{-j} = \frac{1}{c^2}$

(4)$\frac{1}{k_n^2}\sum_{j=1}^n j\rho_n^{-2j} = O(1)$, and $\lim_{n\to\infty} \frac{1}{k_n^2}\sum_{j=1}^n j\rho_n^{-2j} = \frac{1}{4c^2}$.

Proof. When $n \to \infty$,

(1)

$$\frac{1}{k_n}\sum_{j=1}^n \rho_n^{-j} = \frac{\rho_n^{-1}}{k_n}\frac{1-\rho_n^{-n}}{1-\rho_n^{-1}} \to \frac{1}{c}.$$

since

$$1 - \rho_n^{-1} \sim \frac{c}{k_n\left[1 + O(\frac{1}{k_n})\right]}.$$

(2)

$$\frac{1}{k_n}\sum_{j=1}^n \rho_n^{-2j} = \frac{\rho_n^{-2}}{k_n}\frac{1-\rho_n^{-2n}}{1-\rho_n^{-2}} \to \frac{1}{2c},$$

since

$$1 - \rho_n^{-2} \sim \frac{2c}{k_n\left[1 + O(\frac{1}{k_n})\right]}.$$

(3)

$$\begin{aligned}
\frac{1}{k_n^2}\sum_{j=1}^{n} j\rho_n^{-j} &= \frac{1}{k_n^2}\frac{(\rho_n^{-1})(1-(n+1)\rho_n^{-n}+n\rho_n^{-n-1})}{(1-\rho_n^{-1})^2} \\
&\sim \frac{k_n^2\rho_n^{-1}}{c^2k_n^2}(1-\rho_n^{-n}-n\rho_n^{-n}+n\rho_n^{-n-1}) \\
&\to \frac{1}{c^2}.
\end{aligned}$$

(4)

$$\begin{aligned}
\frac{1}{k_n^2}\sum_{j=1}^{n} j\rho_n^{-2j} &= \frac{1}{k_n^2}\frac{(\rho_n^{-2})(1-(n+1)\rho_n^{-2n}+n\rho_n^{-2n-2})}{(1-\rho_n^{-2})^2} \\
&\sim \frac{k_n^2\rho_n^{-2}}{4c^2k_n^2}(1-\rho_n^{-2n}-n\rho_n^{-2n}+n\rho_n^{-2n-2}) \\
&\to \frac{1}{4c^2}.
\end{aligned}$$

■

Lemma A.1.5 *As* $n \to \infty$*,*

$$\begin{aligned}
\mathbb{E}(X_{n,ols}^2) &= \frac{\sigma^2}{c^2} + o(1), \\
\mathbb{E}(Y_{n,ols}^2) &= \frac{\sigma^2}{2c^3} + o(1), \\
\mathbb{E}(X_{n,ols}Y_{n,ols}) &= o(1).
\end{aligned}$$

Proof. As $n \to \infty$, we derive the following arguments,

(1)

$$\mathbb{E}(Y_n^2) = \frac{1}{k_n^3}\mathbb{E}\left[\left(\sum_{j=1}^{n}\rho_n^{-j}u_j\right)\left(\sum_{i=1}^{n}\rho_n^{-i}u_i\right)\right] = \frac{\sigma^2}{k_n^3}\sum_{i=1}^{n}\rho_n^{-i}\sum_{j=1}^{n}\rho_n^{-j}\min\{i,j\}$$

$$= \frac{\sigma^2}{k_n^3}\sum_{i=1}^{n}\rho_n^{-i}\sum_{j=1}^{i}\rho^{-j}j + \frac{\sigma^2}{k_n^3}\sum_{i=1}^{n}i\rho_n^{-i}\sum_{j=i+1}^{n}\rho^{-j}. \quad \text{(A.1.2)}$$

Two terms in (A.1.2) are computed separately. The first term in (A.1.2) follows

$$\begin{aligned}
&\frac{\sigma^2}{k_n^3}\sum_{i=1}^{n}\rho_n^{-i}\sum_{j=1}^{i}\rho^{-j}j = \frac{\sigma^2}{k_n^3}\sum_{i=1}^{n}\rho_n^{-i}\frac{\rho_n^{-1}\left[1-(i+1)\rho_n^{-i}+i\rho_n^{-i-1}\right]}{(1-\rho_n^{-1})^2}\\
&= \frac{\rho_n^{-1}\sigma^2}{c^2k_n}\sum_{i=1}^{n}\rho_n^{-i}(1-(i+1)\rho_n^{-i}+i\rho_n^{-i-1})\\
&= \frac{\rho_n^{-1}\sigma^2}{c^2k_n}\sum_{i=1}^{n}\rho_n^{-i} - \frac{\rho_n^{-1}\sigma^2}{c^2k_n}\sum_{i=1}^{n}(i+1)\rho_n^{-2i} + \frac{\rho_n^{-1}\sigma^2}{c^2k_n}\sum_{i=1}^{n}i\rho_n^{-2i-1}\\
&= \frac{\rho_n^{-1}\sigma^2}{c^2k_n}\sum_{i=1}^{n}\rho_n^{-i} - \frac{\rho_n^{-1}\sigma^2}{c^2k_n}\sum_{i=1}^{n}\rho_n^{-2i} + \frac{\rho_n^{-1}\sigma^2}{c^2k_n}\left(\sum_{i=1}^{n}i\rho_n^{-2i-1} - \sum_{i=1}^{n}i\rho_n^{-2i}\right)\\
&= \frac{\rho_n^{-1}\sigma^2}{c^2k_n}\sum_{i=1}^{n}\rho_n^{-i} - \frac{\rho_n^{-1}\sigma^2}{c^2k_n}\sum_{i=1}^{n}\rho_n^{-2i} - \frac{\rho_n^{-1}\sigma^2}{ck_n^2}\sum_{i=1}^{n}i\rho_n^{-2i}\\
&= \frac{\sigma^2}{4c^3}+o(1).
\end{aligned}$$

Similarly, the second term of (A.1.2) is computed as,

$$\begin{aligned}
&\frac{\sigma^2}{k_n^3}\sum_{i=1}^{n}i\rho_n^{-i}\sum_{j=i+1}^{n}\rho^{-j} = \frac{\sigma^2}{k_n^3}\left(\sum_{i=1}^{n}i\rho_n^{-i}\right)\frac{\rho^{-i-1}(1-\rho_n^{-(n-i)})}{1-\rho_n^{-1}}\\
&= \frac{\sigma^2\rho_n^{-1}}{ck_n^2}\sum_{i=1}^{n}i\rho_n^{-2i} - \rho_n^{-n-1}\left(\frac{\sigma^2}{ck_n^2}\sum_{i=1}^{n}i\rho_n^{-i}\right)\\
&= \frac{\sigma^2}{4c^3}+o(1).
\end{aligned}$$

In all, the following result is derived as,

$$\mathbb{E}(Y_n^2) = \frac{\sigma^2}{2c^3}+o(1).$$

(2) As $n\to\infty$,

$$\begin{aligned}
\mathbb{E}(X_n^2) &= \frac{1}{nk_n^2}\mathbb{E}\left[\left(\sum_{t=1}^{n}\rho_n^{-(n-t)-1}u_t\right)\left(\sum_{s=1}^{n}\rho_n^{-(n-s)-1}u_s\right)\right]\\
&= \frac{1}{nk_n^2}\mathbb{E}\left[\left(\sum_{t=1}^{n}\rho_n^{-t}u_{n-t+1}\right)\left(\sum_{s=1}^{n}\rho_n^{-s}u_{n-s+1}\right)\right]\\
&= \frac{1}{nk_n^2}\sum_{t=1}^{n}\rho_n^{-t}\sum_{s=1}^{n}\rho_n^{-s}\min\{n-t+1,n-s+1\}
\end{aligned}$$

$$\begin{aligned}
&= \frac{\sigma^2}{nk_n^2}\sum_{t=1}^{n}\rho_n^{-t}\sum_{s=1}^{n}\rho^{-s}(n+1) - \frac{\sigma^2}{nk_n^2}\sum_{t=1}^{n}\rho_n^{-t}\sum_{s=1}^{n}\rho_n^{-s}\max\{t,s\} \\
&= \frac{\sigma^2}{nk_n^2}\sum_{t=1}^{n}\rho_n^{-t}\sum_{s=1}^{n}\rho^{-s}(n+1) - \left(\frac{k_n}{n}\right)\left(\frac{\sigma^2}{k_n^3}\sum_{t=1}^{n}t\rho_n^{-t}\right)\left(\sum_{s=1}^{t}\rho_n^{-s}\right) \\
&\quad - \left(\frac{k_n}{n}\right)\left(\frac{\sigma^2}{k_n^3}\sum_{t=1}^{n}\rho_n^{-t}\right)\left(\sum_{s=t+1}^{n}s\rho_n^{-s}\right). \qquad \text{(A.1.3)}
\end{aligned}$$

Similarly three terms are discussed separately. The second (A.1.3) is calculated as,

$$\begin{aligned}
\left(\frac{\sigma^2}{k_n^3}\sum_{t=1}^{n}t\rho_n^{-t}\right)\left(\sum_{s=1}^{t}\rho_n^{-s}\right) &= \frac{\sigma^2}{k_n^3}\sum_{t=1}^{n}\rho_n^{-t}t\frac{\rho_n^{-1}(1-\rho_n^{-t})}{1-\rho^{-1}} \sim \frac{\sigma^2\rho_n^{-1}}{ck_n^2}\sum_{t=1}^{n}t(\rho_n^{-t}-\rho_n^{-2t}) \\
&= \frac{\sigma^2\rho_n^{-1}}{ck_n^2}\sum_{t=1}^{n}t\rho_n^{-t} - \frac{\sigma^2\rho_n^{-t}}{ck_n^2}\sum_{t=1}^{n}t\rho_n^{-2t} = \frac{3\sigma^2}{4c^3}+o(1) \\
&= O(1).
\end{aligned}$$

The third term in (A.1.3) is calculated as,

$$\begin{aligned}
\left(\frac{\sigma^2}{k_n^3}\sum_{t=1}^{n}\rho_n^{-t}\right)\left(\sum_{s=t+1}^{n}s\rho_n^{-s}\right) &\le \left(\frac{\sigma^2}{k_n^3}\sum_{t=1}^{n}\rho_n^{-t}\right)\left(\sum_{s=1}^{n}s\rho_n^{-s}\right) \\
&= \left(\frac{\sigma^2}{k_n}\sum_{t=1}^{n}\rho_n^{-t}\right)\left(\frac{1}{k_n^2}\sum_{s=1}^{n}s\rho_n^{-s}\right) \\
&= O(1).
\end{aligned}$$

In all, we have the following argument as,

$$\begin{aligned}
\mathbb{E}(X_n^2) &= \frac{(n+1)\sigma^2}{nk_n^2}\sum_{t=1}^{n}\rho_n^{-t}\sum_{s=1}^{n}\rho_n^{-s} + O\left(\frac{k_n}{n}\right) = \frac{(n+1)\rho_n^{-2}}{n}\frac{\sigma^2}{c^2}(1-\rho_n^{-n})^2 + o(1) \\
&= \frac{\sigma^2}{c^2}+o(1).
\end{aligned}$$

(3) As $n \to \infty$,

$$\mathbb{E}(X_nY_n) = \mathbb{E}\left[\frac{1}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}\rho_n^{-(n-t)-1}u_t\sum_{j=1}^{n}\rho_n^{-j}u_j\right] = \frac{\rho_n^{-n-1}}{\sqrt{n}k_n^{\frac{5}{2}}}\left[\sum_{t=1}^{n}\rho_n^{t}\sum_{j=1}^{n}\sigma^2\rho_n^{-j}\min\{t,j\}\right] \qquad \text{(A.1.4)}$$

$$= \frac{\sigma^2\rho_n^{-n-1}}{\sqrt{n}k_n^{\frac{5}{2}}}\left(\sum_{t=1}^{n}\rho_n^t\right)\left(\sum_{j=1}^{t}\rho_n^{-j}j\right) + \frac{\sigma^2\rho_n^{-n-1}}{\sqrt{n}k_n^{\frac{5}{2}}}\left(\sum_{t=1}^{n}t\rho_n^t\right)\left(\sum_{j=t+1}^{n}\rho_n^{-j}\right).$$

The first term of (A.1.4) is computed as,

$$\begin{aligned}
&\frac{\sqrt{k_n}}{\sqrt{n}}\frac{\sigma^2\rho_n^{-n-1}}{k_n^3}\left(\sum_{t=1}^{n}\rho_n^t\right)\left(\sum_{j=1}^{t}\rho_n^{-j}j\right) = \frac{\sqrt{k_n}}{\sqrt{n}}\frac{\sigma^2\rho_n^{-n-1}}{k_n^3}\sum_{t=1}^{n}\rho_n^t\left(\rho_n^{-1}\frac{1-(t+1)\rho_n^{-t}+t\rho_n^{-(t+1)}}{(1-\rho_n^{-1})^2}\right)\\
&= \frac{\sqrt{k_n}}{\sqrt{n}}\frac{\sigma^2\rho_n^{-n-2}}{c^2k_n}\sum_{t=1}^{n}\rho_n^t(1-(t+1)\rho_n^{-t}+t\rho_n^{-(t+1)})\\
&= \frac{\sqrt{k_n}}{\sqrt{n}}\frac{\sigma^2\rho_n^{-n-2}}{c^2k_n}\sum_{t=1}^{n}(\rho_n^t+t\rho_n^{-1}-(t+1))\\
&= \frac{\sqrt{k_n}}{\sqrt{n}}\frac{\sigma^2\rho_n^{-n-1}}{c^2k_n}\frac{\rho_n^n-1}{\rho_n-1}+o\left(\frac{\sqrt{k_n}}{\sqrt{n}}\right)\\
&= O\left(\frac{\sqrt{k_n}}{\sqrt{n}}\right) = o(1),
\end{aligned}$$

where exponential rates dominate polynomial rates. The second part of (A.1.4) can be derived similarly.

$$\begin{aligned}
&\frac{\sigma^2\rho_n^{-n-1}}{\sqrt{n}k_n^{\frac{5}{2}}}\left(\sum_{t=1}^{n}t\rho_n^t\right)\left(\sum_{j=t+1}^{n}\rho_n^{-j}\right) = \frac{\sigma^2\rho_n^{-n-1}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}t\rho_n^t\frac{\rho_n^{-(t+1)}(1-\rho_n^{-(n-t)})}{1-\rho_n^{-1}}\\
&= \frac{\sigma^2\rho_n^{-n-2}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}\frac{t(1-\rho_n^{-(n-t)})}{1-\rho_n^{-1}} = \frac{\sigma^2\rho_n^{-n-2}}{ck_n^{\frac{3}{2}}\sqrt{n}}\sum_{t=1}^{n}t(1-\rho_n^{-(n-t)})\\
&= \frac{\sigma^2\rho_n^{-n-2}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}t - \frac{\sigma^2\rho_n^{-n-2}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}t\rho_n^{-(n-t)} = o(1),
\end{aligned}$$

due to Lemma A.1.3 and the dominance of exponential rates over polynomial rates. Finally, it is justified that

$$\mathbb{E}(X_nY_n) = o(1).$$

We complete the proof. ■

Lemma A.1.6 *As* $n\to\infty$, $\frac{\rho_n^{-n}}{\sqrt{n}k_n^{\frac{5}{2}}}\mathbb{E}\sum_{t=1}^{n}\sum_{j=t}^{n}\rho_n^{t-j}u_tu_j = o_p(1)$.

Proof. As $n\to\infty$,

$$\frac{\rho_n^{-n}}{\sqrt{n}k_n^{\frac{5}{2}}}\mathbb{E}\sum_{t=1}^{n}\sum_{j=t}^{n}\rho_n^{t-j}u_tu_j \;=\; \sigma^2\frac{\rho_n^{-n}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}\sum_{j=t}^{n}\rho_n^{t-j}t = \frac{\sigma^2\rho_n^{-n}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}t\frac{1-\rho_n^{-(n-t)}}{1-\rho_n^{-1}}$$

$$\leq \frac{\sigma^2 \rho_n^{-n}}{\sqrt{n} k_n^{\frac{5}{2}}} \sum_{t=1}^{n} \frac{t}{1-\rho_n^{-1}} = O_p\left(\rho_n^{-n} \frac{n^{\frac{3}{2}}}{k_n^{\frac{3}{2}}}\right) = o_p(1),$$

due to Lemma A.1.3. ■

A.1.5 Proof of Lemma 2.2.2

Proof. By Cramér-Wold device, it is sufficient to show that $pX_{n,ols} + qY_{n,ols} \xrightarrow{d} pX + qY$, equivalent to show that $pX_{n,ols} + qY_{n,ols} \xrightarrow{d} \mathcal{N}(0, \frac{(2cp^2+q^2)\sigma^2}{2c^3})$. The expression is rewritten as

$$\begin{aligned} pX_n + qY_n &= \sum_{t=1}^{n} \left\{ \left[\frac{1}{k_n\sqrt{n}} p\rho_n^{-(n-t)-1} + \frac{1}{k_n^{\frac{3}{2}}} q\rho_n^{-t} \right] \sum_{s=1}^{t} \epsilon_s \right\} \\ &= \sum_{s=1}^{n} \left\{ \sum_{t=s+1}^{n} \left[\frac{1}{k_n\sqrt{n}} p\rho_n^{-(n-t)-1} + \frac{1}{k_n^{\frac{3}{2}}} q\rho_n^{-t} \right] \epsilon_s \right\}. \end{aligned}$$

Define $\zeta_{ns} := \sum_{t=s+1}^{n} \left[\frac{1}{k_n\sqrt{n}} p\rho_n^{-(n-t)-1} + \frac{1}{k_n^{\frac{3}{2}}} q\rho_n^{-t} \right] \epsilon_s$. The term ζ_{ns} is independent but not identically distributed. The Lindeberg-Feller central limit theorem (White, 2014) is applied to derive the asymptotic normality. First, the stability condition is shown:

$$\mathbb{E}\left[\sum_{s=1}^{n} \zeta_{ns}\right]^2 = p^2\mathbb{E}(X_{n,ols})^2 + q^2\mathbb{E}(Y_{n,ols})^2 + 2pq\mathbb{E}(X_{n,ols}Y_{n,ols}) = \frac{(2cp^2+q^2)\sigma^2}{2c^3} + o(1).$$

Then, the Lindeberg condition needs to be checked. For any $\eta > 0$,

$$\begin{aligned} &\sum_{s=1}^{n} \mathbb{E}\left(\zeta_{ns}^2 1_{|\zeta_{ns}|>\eta}\right) \\ &= \sum_{s=1}^{n} \left(\sum_{t=s+1}^{n} \left[\frac{1}{\sqrt{n}k_n} p\rho_n^{-(n-t)-1} + \frac{1}{k_n^{\frac{3}{2}}} q\rho_n^{-t} \right] \right)^2 \mathbb{E}\left(\epsilon_s^2 1_{\left\{ \left| \sum_{t=s+1}^{n} \left[p\rho_n^{-(n-t)-1} + k_n^{-\frac{3}{2}} q\rho_n^{-t} \right] \right| |\epsilon_s| > \eta \right\}} \right) \\ &\leq \sum_{s=1}^{n} \left[\frac{2}{k_n^2 n} \left(\sum_{t=s+1}^{n} p\rho_n^{-(n-t)-1} \right)^2 + \frac{2}{k_n^3} \left(\sum_{t=s+1}^{n} q\rho_n^{-t} \right)^2 \right] \mathbb{E}\left[\epsilon_s^2 1_{\left\{ \left[\sum_{t=1}^{n} \left(\frac{1}{k_n\sqrt{n}} p\rho_n^{-(n-t)-1} + k_n^{-\frac{3}{2}} q\rho_n^{-1} \right) \right]^2 \epsilon_s^2 > \eta^2 \right\}} \right] \\ &\leq K \max_{1\leq s\leq n} \mathbb{E}\left(\epsilon_s^2 1_{\left\{ 2\left[\frac{1}{k_n^2 n} \left(\sum_{t=s+1}^{n} p\rho_n^{-(n-t)-1} \right)^2 + \frac{1}{k_n^3} \left(\sum_{t=s+1}^{n} q\rho_n^{-t} \right)^2 \right] \epsilon_s^2 > \eta^2 \right\}} \right), \end{aligned}$$

where K is some constant value since $\frac{1}{k_n^2 n}\sum_{s=1}^n \left(p\sum_{t=s+1}^n \rho_n^{-(n-t)-1}\right)^2 \leq \frac{p^2}{c^2}$ and $\frac{1}{k_n^3}\sum_{s=1}^n \left(q\sum_{t=s+1}^n \rho_n^{-t}\right)^2 \leq \frac{q^2}{2c^3}$. For any $s \geq 1$, $\left(\sum_{t=s+1}^n \rho_n^{-(n-t)-1}\right)^2 \leq \frac{k_n^2}{c^2}$ and $\left(\sum_{t=s+1}^n \rho_n^{-t}\right)^2 \leq \frac{k_n^2}{c^2}$. Further, it follows that

$$\begin{aligned}
\sum_{s=1}^n \mathbb{E}(\zeta_{ns}^2 1_{|\zeta_{ns}|>\eta}) &\leq K \max_{1\leq s\leq n} \mathbb{E}\left(\epsilon_i^2 1_{\left\{2\left(\frac{p^2}{nk_n^2}+\frac{q^2}{k_n^3}\right)\frac{k_n^2}{c^2}\epsilon_s^2>\eta^2\right\}}\right) \\
&\leq K \max_{1\leq s\leq n} \mathbb{E}\left(\epsilon_s^2 1_{\left\{2\frac{p^2}{nk_n^2}\frac{k_n^2}{c^2}\epsilon_s^2>\eta^2/2\right\}}\right) + K \max_{1\leq s\leq n} \mathbb{E}\left(\epsilon_i^2 1_{\left\{2\frac{q^2}{k_n^3}\frac{k_n^2}{c^2}\epsilon_s^2>\eta^2/2\right\}}\right) \\
&\leq K \max_{1\leq s\leq n} \mathbb{E}\left(\epsilon_s^2 1_{\left\{\frac{2p^2}{c^2}\epsilon_s^2>\eta^2 n/2\right\}}\right) + K \max_{1\leq s\leq n} \mathbb{E}\left(\epsilon_s^2 1_{\left\{2\frac{q^2}{c^2}\epsilon_s^2>\eta^2 k_n/2\right\}}\right) \to 0,
\end{aligned}$$

where $n \to \infty$. Proofs end here. ■

A.1.6 Proof of Theorem 2.2.4

Proof. First we discuss the denominator as,

$$\frac{\rho_n^{-2n}}{k_n^4}\sum_{t=1}^n y_{t-1}^2 = \frac{1}{k_n(\rho_n^2-1)}\left\{\frac{\rho_n^{-2n}}{k_n^3}y_n^2 - \frac{\rho_n^{-2n}}{k_n^3}y_0^2 - \frac{\rho_n^{-2n+1}}{k_n^3}\sum_{t=1}^n y_{t-1}u_t - \frac{\rho_n^{-2n}}{k_n^3}\sum_{t=1}^n u_t^2\right\}.$$

Since $\sum_{t=1}^n u_t^2 = O_p(n^2)$, $\frac{\rho_n^{-2n}}{k_n^3}\sum_{t=1}^n u_t^2 = o_p(1)$. As $y_0 = o_p(k_n^{\frac{1}{2}})$, $\frac{\rho_n^{-2n}}{k_n^3}y_0^2 = o_p(1)$. In addition, we have

$$\frac{\rho_n^{-2n+1}}{k_n^3}\sum_{t=1}^n y_{t-1}u_t = \frac{y_0}{k_n^{\frac{3}{2}}}\frac{\rho_n^{-n}}{k_n^{\frac{3}{2}}}\sum_{t=1}^n \rho_n^{-(n-t)}u_t + \frac{\rho_n^{-2n+1}}{k_n^3}\sum_{t=1}^n\left(\sum_{j=1}^{t-1}\rho_n^{t-1-j}u_j\right)u_t.$$

By the initial condition for y_0, $\frac{y_0}{k_n^{\frac{1}{2}}} = o_p(1)$. By the dominance of exponential rates, $\frac{\rho_n^{-n}}{k_n^{\frac{5}{2}}}\sum_{t=1}^n \rho_n^{-(n-t)}u_t = o_p(1)$. So, $\frac{\rho_n^{-2n}}{k_n^3}\sum_{t=1}^n y_{t-1}u_t = o_p(1)$.

It follows that

$$\begin{aligned}
\frac{\rho_n^{-2n}}{k_n^4}\sum_{t=1}^n y_{t-1}^2 &= \frac{1}{k_n(\rho_n^2-1)}\left(\frac{\rho_n^{-n}}{k_n^{\frac{3}{2}}}y_n\right)^2 + o_p(1) = \frac{1}{k_n(\rho_n^2-1)}\left(\frac{y_0}{k_n^{\frac{3}{2}}} + \frac{1}{k^{\frac{3}{2}}}\sum_{j=1}^n \rho_n^{-j}u_j\right)^2 + o_p(1) \\
&= \frac{1}{k_n(\rho_n^2-1)}\left(\frac{1}{k^{\frac{3}{2}}}\sum_{j=1}^n \rho_n^{-j}u_j\right)^2 + o_p(1) = \frac{1}{2c}Y_{n,ols}^2 + o_p(1).
\end{aligned}$$

For the numerator, the following decomposition is shown

$$
\begin{aligned}
&\frac{\rho_n^{-n}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n} y_{t-1}u_t = \frac{\rho_n^{-n}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}\sum_{j=1}^{t-j}\rho_n^{t-j}u_t u_j = \frac{\rho_n^{-n}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}\sum_{j=1}^{t-j}\rho_n^{t-j}u_t u_j + o_p(1)\\
&= \frac{\rho_n^{-n}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}\sum_{j=1}^{n}\rho_n^{t-j}u_t u_j - \frac{\rho_n^{-n}}{\sqrt{n}k_n^{\frac{5}{2}}}\sum_{t=1}^{n}\sum_{j=t}^{n}\rho_n^{t-j}u_t u_j + o_p(1)\\
&= \left(\frac{1}{\sqrt{n}k_n}\sum_{t=1}^{n}\rho_n^{-(n-t)-1}u_t\right)\left(\frac{1}{k_n^{\frac{3}{2}}}\sum_{j=1}^{n}\rho_n^{-j}u_j\right) + o_p(1)\\
&= X_{n,ols}Y_{n,ols} + o_p(1),
\end{aligned}
$$

by Lemma A.1.6. Then the proof is complete. ■

A.1.7 Proof of Theorem 2.2.6

Proof. For the mildly stationary case, it is shown that

$$
\begin{aligned}
&(1-\rho_n^2)\sum_{t=1}^{n} y_{t-1}^2 = y_0^2 - y_n^2 + \sum_{t=1}^{n} u_t^2 + 2\rho_n\sum_{t=1}^{n} y_{t-1}u_t \qquad \text{(A.1.5)}\\
&\Rightarrow\\
&\frac{1}{k_n n^2}\sum_{t=1}^{n} y_{t-1}^2 = \frac{1}{(1-\rho_n^2)}\left\{\frac{1}{k_n n^2}y_0^2 - \frac{1}{k_n n^2}y_n^2 + \frac{1}{k_n n^2}\sum_{t=1}^{n}u_t^2 + \frac{2}{k_n n^2}\rho_n\sum_{t=1}^{n} y_{t-1}u_t\right\}.
\end{aligned}
$$

Specifically, the order of each term of (A.1.5) is discussed. It shows that $y_0^2 = o_p(k_n)$. It follows that

$$
\frac{1}{k_n n^2}y_n^2 = \frac{1}{k_n n^2}\left(\rho_n^n y_0 + \sum_{j=1}^{n}\rho_n^{n-j}u_j\right)^2 \le 2\frac{1}{k_n n^2}\rho^{2n}y_0^2 + 2\frac{1}{k_n n^2}\left(\sum_{j=1}^{n}\rho_n^{n-j}u_j\right)^2. \qquad \text{(A.1.6)}
$$

For (A.1.6), it is shown that $\frac{1}{k_n n^2}\rho^{2n}y_0^2 = o_p(1)$, and

$$
\begin{aligned}
&\frac{1}{k_n n^2}\mathbb{E}\left(\sum_{j=1}^{n}\rho_n^{2n-j}u_j\right)^2 = \frac{1}{k_n n^2}\mathbb{E}\left[\sum_{j=1}^{n}\rho_n^{n-j}u_j\right]\left[\sum_{i=1}^{n}\rho_n^{n-i}u_i\right]\\
&= \frac{\sigma^2}{n^2 k_n}\sum_{j=1}^{n}\rho_n^{j-1}\sum_{i=1}^{n}\rho_n^{i-1}\min\{n+1-j, n+1-i\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\sigma^2}{n^2 k_n} \sum_{j=1}^{n} \rho_n^{j-1} \sum_{i=1}^{n} \rho_n^{i-1}(n+1) - \frac{\sigma^2}{n^2 k_n} \sum_{j=1}^{n} \rho_n^{j-1} \sum_{i=1}^{n} \rho_n^{i-1} \max\{i,j\} \\
&= \frac{\sigma^2}{n^2 k_n} \sum_{j=1}^{n} \rho_n^{j-1} \sum_{i=1}^{n} \rho_n^{i-1}(n+1) + O\left(\frac{k_n^2}{n^2}\right) \\
&= \frac{\sigma^2}{n^2 k_n} \left(\sum_{j=1}^{n} \rho_n^{j-1}\right)^2 (n+1) + o(1) = \left(\frac{k_n}{n}\right) \frac{\sigma^2}{c^2} + o(1) = o(1).
\end{aligned}
$$

In summary,

$$
\frac{1}{n^2 k_n} y_n^2 = \frac{k_n}{n} \frac{\sigma^2}{c^2} + o(1) = o(1).
$$

For the term $\frac{1}{n^2 k_n} \sum_{t=1}^{n} y_{t-1} u_t$ of (A.1.5), it is shown that

$$
\frac{1}{n^2 k_n} \sum_{t=1}^{n} y_{t-1} u_t = \frac{y_0}{n^2 k_n} \sum_{t=1}^{n} \rho_n^{t-1} u_t + \frac{1}{n^2 k_n} \sum_{t=1}^{n} \left(\sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t. \qquad \text{(A.1.7)}
$$

For the first part in (A.1.7), we have ,

$$
\begin{aligned}
&\frac{y_0^2}{n^4 k_n^2} \mathbb{E} \left(\sum_{t=1}^{n} \rho_n^{t-1} u_t \right)^2 = \frac{y_0^2 \rho_n^{-2}}{n^4 k_n^2} \mathbb{E} \left[\sum_{i=1}^{n} \rho_n^{i} \sum_{j=1}^{n} \rho_n^{j} u_i u_j \right] \\
&= \frac{\sigma^2 y_0^2 \rho_n^{-2}}{n^4 k_n^2} \sum_{i=1}^{n} \rho_n^{i} \sum_{j=1}^{n} \rho_n^{j} \min\{i,j\} = o_p(k_n) \times O(k_n^3) \times \frac{1}{n^4 k_n^2} = o_p\left(\frac{k_n^2}{n^4}\right) \\
&= o_p(1).
\end{aligned}
$$

For the second part of (A.1.7),

$$
\begin{aligned}
&\frac{1}{n^2 k_n} \mathbb{E} \left[\sum_{t=1}^{n} \left(\sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j u_t \right) \right] = \frac{\sigma^2}{n^2 k_n} \frac{1}{\rho_n} \sum_{t=1}^{n} \rho_n^{t} \left(\sum_{j=1}^{t-1} \rho_n^{-j} \right) \min\{t,j\} \\
&= \frac{\sigma^2}{n^2 k_n} \frac{1}{\rho_n^2} \sum_{t=1}^{n} \rho_n^{t} \frac{1 - t\rho_n^{-(t-1)} - (t-1)\rho_n^{-t}}{(1-\rho_n^{-1})^2} \sim \frac{k_n \sigma^2}{c^2 n^2} \frac{1}{\rho_n^2} \sum_{t=1}^{n} \rho_n^{t} (1 - t\rho_n^{-t+1} + (t-1)\rho_n^{-t}) \\
&= \frac{k_n \sigma^2}{c^2 n^2} \frac{1}{\rho_n^2} \sum_{t=1}^{n} \rho_n^{t} - \frac{k_n \sigma^2}{c^2 n^2} \frac{1}{\rho_n} \sum_{t=1}^{n} t + \frac{k_n \sigma^2}{c^2 n^2} \frac{1}{\rho_n^2} \sum_{t=1}^{n} (t-1) \\
&= \frac{k_n \sigma^2 (1-\rho_n^n)}{c^2 n^2 \rho_n (1-\rho_n)} - \frac{k_n \sigma^2 n(n+1)}{2c^2 n^2 \rho_n} + \frac{k_n \sigma^2 n(n+1)}{2c^2 n^2 \rho_n^2} - \frac{k_n \sigma^2 n}{c^2 n^2 \rho_n^2} \\
&= \frac{k_n \sigma^2 \rho_n (1-\rho_n^n)}{c^2 n^2 \rho_n (1-\rho_n)} + \frac{k_n \sigma^2 n(n+1)}{2\rho_n^2 c^2 n^2} (1-\rho_n) - \frac{k_n \sigma^2 n}{c^2 n^2 \rho_n} \\
&= -\frac{\sigma^2}{2c} + O\left(\frac{k_n^2}{n^2}\right) + O\left(\frac{k_n}{n}\right) = -\frac{\sigma^2}{2c} + o(1).
\end{aligned}
$$

So $\frac{1}{n^2 k_n}\sum_{t=1}^{n} y_{t-1}u_t = -\frac{\sigma^2}{2c} + o_p(1)$. It is also easy to show $\frac{1}{n^2}\mathbb{E}\sum_{t=1}^{n} u_t^2 = \frac{\sigma^2}{2} + o_p(1)$. Based on all above discussions, it follows that

$$\frac{\rho_n \sum_{t=1}^{n} y_{t-1}u_t}{n^2 k_n} = O_p(1).$$

The proof is complete. ■

A.2 Proofs in Chapter 3

A.2.1 Technical lemmas of recursive k-means algorithm

We collect technical proofs for classifications, estimations and inferences on the recursive k-means classifications and the modified k-means classifications.

We denote $\widehat{g}_i := \widehat{g}_i\left(\widehat{\overline{c}}\right)$ as any k-means classification estimates for g_i^0 for each $i = 1, 2, ...n$. To demonstrate consistency of the recursive k-means classifications, we intend to establish the consistency of $\widehat{\overline{c}}$ in terms of the Hausdorff distance. The Hausdorff distance measures how far two compact subsets of a metric space are separated from each other:

$$d_H(a,b) = \max\left\{\max_{g\in\{1,2,...,K^0\}}\left(\min_{\widetilde{g}\in\{1,2,...,K^0\}}|a_{\widetilde{g}} - b_g|\right), \max_{\widetilde{g}\in\{1,2,...,K^0\}}\left(\min_{g\in\{1,2,...,K^0\}}|a_{\widetilde{g}} - b_g|\right)\right\}.$$

Lemma A.2.1 *If Assumption 1 and 2 hold,*

$$\sup_{(\overline{c},\delta)\in\mathcal{C}\times\Delta_{K^0}} T^{\gamma}\left|\widehat{Q}_{nT}(\overline{c},\delta) - \widetilde{Q}_{nT}(\overline{c},\delta)\right| = o_p(1),$$

where

$$\widehat{Q}_{nT}(\overline{c},\delta) := \frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^{T}\left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\overline{\rho}_{g_i}\right)^2,$$

and

$$\widetilde{Q}_{nT}(\overline{c},\delta) := \frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t-1}\overline{\rho}_{g_i}\right)^2 + \frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{u}_{it}^2,$$

and $\{\breve{\rho}_i\}_{i=1}^n$ *are the collection of individual least squares estimates and* $\overline{\rho}_{g_i} := \exp\left(\frac{\overline{c}_{g_i}}{T^\gamma}\right)$.

Proof. Define $\overline{\rho}^0_{g_i^0} := \exp\left(\overline{c}^0_{g_i^0}/T^\gamma\right)$ and $\overline{\rho}_{g_i} := \exp\left(\frac{\overline{c}_{g_i}}{T^\gamma}\right)$. Observe the following argument as,

$$\begin{aligned}
T^\gamma\left[\widehat{Q}_{nT}(\overline{c},\delta) - \widetilde{Q}_{nT}(\overline{c},\delta)\right] &= \frac{2}{nT^\gamma}\sum_{i=1}^n \frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}\left(\overline{\rho}^0_{g_i^0} - \overline{\rho}_{g_i}\right) \\
&= \frac{2}{nT^{2\gamma}}\sum_{i=1}^n \frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}\left(\overline{c}^0_{g_i^0} - \overline{c}_{g_i}\right) \\
&= \frac{2}{nT^{2\gamma}}\sum_{i=1}^n \frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}^0_{g_i^0} \\
&\quad -\frac{2}{nT^{2\gamma}}\sum_{i=1}^n \frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}_{g_i}.
\end{aligned}$$

We have

$$\frac{2}{nT^{2\gamma}}\sum_{i=1}^n \frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}^0_{g_i^0} = \sum_{\widetilde{g}=1}^{K^0}\mathbf{1}\left\{g_i^0 = \widetilde{g}\right\}\frac{1}{n}\sum_{i=1}^n \frac{1}{\breve{\rho}_i^{2T}T^{2\gamma}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}^0_{\widetilde{g}}.$$

For any $\widetilde{g} \in \{1, 2, ..., K^0\}$,

$$\begin{aligned}
\mathbf{1}\left\{g_i^0 = \widetilde{g}\right\}\frac{1}{n}\sum_{i=1}^n \frac{1}{\breve{\rho}_i^{2T}T^{2\gamma}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}_{\widetilde{g}} &\leq \left|\overline{c}^0_{\widetilde{g}}\right|\mathbf{1}\left\{g_i^0 = \widetilde{g}\right\}\frac{1}{n}\sum_{i=1}^n \frac{1}{\breve{\rho}_i^{2T}T^{2\gamma}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it} \\
&= O_p\left(\frac{1}{\overline{\rho}_{\widetilde{g}}^T T^\gamma\sqrt{n}}\right),
\end{aligned}$$

since $|\overline{c}_{\widetilde{g}}| \leq c_u$ due to the compact support of distance parameters. Define $\overline{\rho}_{\widetilde{g}} := \exp\left(\frac{\overline{c}_{\widetilde{g}}}{T^\gamma}\right)$. Therefore

$$\frac{2}{nT^{2\gamma}}\sum_{i=1}^n \frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}^0_{g_i^0} = O_p\left(\frac{1}{(\rho_l)^T T^\gamma\sqrt{n}}\right), \qquad \text{(A.2.1)}$$

where we define $\rho_l := \left(\exp\left(\frac{c_l}{T^\gamma}\right)\right)$.

Similar argument can be applied to the term $\frac{2}{nT^{2\gamma}}\sum_{i=1}^n \frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}_{g_i}$ and we can justify

$$\frac{2}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}_{g_i}=O_p\left(\frac{1}{(\rho_l)^T T^{\gamma}\sqrt{n}}\right). \tag{A.2.2}$$

By combining above results (A.2.1) and (A.2.2), we have

$$\sup_{(\overline{c},\delta)\in\mathcal{C}\times\Delta_{K^0}} T^{\gamma}\left|\widehat{Q}_{nT}(c,\delta)-\widetilde{Q}_{nT}(c,\delta)\right|=O_p\left(\frac{1}{(\rho_l)^T T^{\gamma}\sqrt{n}}\right)=o_p(1),$$

due to the dominance of exponential rates. ■

Lemma A.2.2 *(Estimation Consistency) Assumption 1 and 2 hold, and* $c_i^0>0$ *for each i. Under* $(n,T)\to\infty$*,*

$$d_H(\overline{c}^0,\widehat{\overline{c}})\overset{p}{\to}0.$$

By Lemma A.2.2, there is one permutation $\tau:\{1,2,...K^0\}\to\{1,2,...,K^0\}$ such that parameter estimates converge to the true values,

$$\left|\widehat{\overline{c}}_{\tau(g)}-\overline{c}_g^0\right|\overset{p}{\to}0.$$

By relabelling $\widehat{\overline{c}}$, we take $\tau(g)=g$. For any $\eta>0$, we define

$$\mathcal{N}_{\eta}:=\left\{\overline{c}\in\mathcal{C}:\left|\overline{c}_g^0-\overline{c}_g\right|<\eta,\forall g\in\left\{1,2,...,K^0\right\}\right\}. \tag{A.2.3}$$

Let $\widehat{g}_i(\widehat{\overline{c}})=\arg\min_{g\in\{1,2,...,K^0\}}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\exp\left(\frac{\widehat{\overline{c}}_g}{T^{\gamma}}\right)\right)^2$. After verifying the consistency of $\widehat{\overline{c}}$ for $\overline{c}^0$, we provide the individual and uniform consistency of recursive k-means classifications in the following theorems.

Proof. Define $\overline{\rho}_{g_i^0}^0:=\exp\left(\overline{c}_{g_i^0}^0/T^{\gamma}\right)$ and $\overline{\rho}_{g_i}:=\exp\left(\frac{\overline{c}_{g_i}}{T^{\gamma}}\right)$. Observe the following argument as,

$$\widetilde{Q}_{nT}(\widehat{\overline{c}},\widehat{\delta})=\widehat{Q}_{nT}(\widehat{\overline{c}},\widehat{\delta})+o_p(T^{-\gamma})\le\widehat{Q}_{nT}(\overline{c}^0,\delta^0)+o_p(T^{-\gamma})=\widetilde{Q}_{nT}(\overline{c}^0,\delta^0)+o_p(T^{-\gamma}),$$

where the equalities come from Lemma A.2.1. Because $\widetilde{Q}_{nT}(c,\delta)$ is minimized at $\overline{c}=\overline{c}^0$ and $\delta=\delta^0$, we have

$$\widetilde{Q}_{nT}(\widehat{\overline{c}},\widehat{\delta})-\widetilde{Q}_{nT}(\overline{c}^0,\delta^0)=o_p(T^{-\gamma}).$$

On the other hand, for any $\overline{c}$ and $\overline{c}_g^0>\overline{c}_{\widetilde{g}}$ (The case $\overline{c}_g^0\leq\overline{c}_{\widetilde{g}}$ can be discussed in the identical way), we have

$$\begin{aligned}
o_p(T^{-\gamma}) &= \widetilde{Q}_{nT}(\overline{c},\delta)-\widetilde{Q}_{nT}(\overline{c}^0,\delta^0)\\
&= \frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left((\overline{\rho}^0_{g_i^0})^2-(\overline{\rho}_{g_i})^2\right)\\
&= \sum_{g=1}^{K^0}\sum_{\widetilde{g}=1}^{K^0}\left(\left(\overline{\rho}_g^0\right)^2-\left(\overline{\rho}_{\widetilde{g}}\right)^2\right)\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T^{2\gamma}\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right]\\
&\geq \underline{M}\sum_{g=1}^{K^0}\sum_{\widetilde{g}=1}^{K^0}\left(\left(\overline{\rho}_g^0\right)^2-\left(\overline{\rho}_{\widetilde{g}}\right)^2\right)\frac{\sigma^2}{4c_u^2}+o_p\left(\frac{1}{T^\gamma}\right)\\
&\geq \frac{\sigma^2}{4c_u^2}\frac{\underline{\underline{M}}}{T^\gamma}\sum_{g=1}^{K^0}\min_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left(\left(\overline{c}_g^0-\overline{c}_{\widetilde{g}}\right)\left(\overline{\rho}_g^0+\overline{\rho}_{\widetilde{g}}\right)\right)+o_p\left(\frac{1}{T^\gamma}\right)\\
&\geq \frac{\sigma^2}{4c_u^2}\frac{\underline{\underline{M}}}{T^\gamma}\max_{g\in\{1,2,\ldots,K^0\}}\left(\min_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left(\left(\overline{c}_g^0\right)-\left(\overline{c}_{\widetilde{g}}\right)\left(\overline{\rho}_g^0+\overline{\rho}_{\widetilde{g}}\right)\right)\right)+o_p\left(\frac{1}{T^\gamma}\right).
\end{aligned}$$

Note that $\frac{\sigma^2}{4c_u^2}$ is bounded away from zeros by Assumption 1 and 2. As a result,

$$\max_{g\in\{1,2,\ldots,K^0\}}\left(\min_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left|\overline{\rho}_g^0-\overline{\rho}_{\widetilde{g}}\right|\right)=o_p\left(T^{-\gamma}\right), \tag{A.2.4}$$

or

$$\max_{g\in\{1,2,\ldots,K^0\}}\left(\min_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left|\overline{c}_g^0-\overline{c}_{\widetilde{g}}\right|\right)=o_p\left(1\right). \tag{A.2.5}$$

Let

$$\tau(g)=\arg\min_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left|\overline{c}_g^0-\overline{c}_{\widetilde{g}}\right|.$$

Then we have for $\widetilde{g}\neq g$,

$$T^\gamma\left|\frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\left(\widehat{\overline{\rho}}_{\tau(g)}\right)^2-\left(\widehat{\overline{\rho}}_{\tau(\widetilde{g})}\right)^2\right)\right|$$

$$
\begin{aligned}
\geq\ & T^{\gamma}\left|\frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\left(\overline{\rho}_g^0\right)^2-\left(\overline{\rho}_{\widetilde{g}}^0\right)^2\right)\right| \\
& -T^{\gamma}\left|\frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\left(\widehat{\overline{\rho}}_{\tau(g)}\right)^2-\left(\overline{\rho}_g^0\right)^2\right)\right| \\
& -T^{\gamma}\left|\frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\left(\widehat{\overline{\rho}}_{\tau(\widetilde{g})}\right)^2-\left(\overline{\rho}_{\widetilde{g}}^0\right)^2\right)\right|. \qquad \text{(A.2.6)}
\end{aligned}
$$

The first term on the right hand side of (A.2.6) is bounded away from zero since

$$
\frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\breve{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \geq \underline{M}\frac{\sigma^2}{4c_u^2},
$$

w.p.a.1 and $\left|\overline{c}_g^0-\overline{c}_{\widetilde{g}}^0\right|\geq c^*>0$ as $\widetilde{g}\neq g$. Due to (A.2.4) or (A.2.5), we have the 2nd and 3rd term of (A.2.6) are $o_p(1)$. Therefore, we have $\tau(g)\neq\tau(\widetilde{g})$ with probability approaching one. We note that asymptotically τ is not only an onto mapping but also one-to-one mapping. Hence τ has the inverse mapping denoted as τ^{-1}. Then we have

$$
\min_{g\in\{1,2,\ldots,K^0\}}\left|\overline{\rho}_g^0-\widehat{\overline{\rho}}_{\widetilde{g}}\right| \geq \left|\overline{\rho}_{\tau^{-1}(\widetilde{g})}^0-\widehat{\overline{\rho}}_{\widetilde{g}}\right| = \min_{h\in\{1,2,\ldots,K^0\}}\left|\overline{\rho}_{\tau^{-1}(h)}^0-\widehat{\overline{\rho}}_h\right| = o_p\left(T^{-\gamma}\right),
$$

where the last equality comes from (A.2.4) or (A.2.5). So we have

$$
\max_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left(\min_{g\in\{1,2,\ldots,K^0\}}\left|\overline{\rho}_g^0-\widehat{\overline{\rho}}_{\widetilde{g}}\right|\right)=o_p(T^{-\gamma}). \qquad \text{(A.2.7)}
$$

In all, by combining the results for two terms (A.2.4) and (A.2.7), we complete the proof. ■

Lemma A.2.3 *If Assumption 1 and 2 hold. For any $M>0$,*

$$
\max_{1\leq i\leq n}\Pr\left(\frac{1}{\rho_i^{2T}T^{\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\geq M\right)=o\left(\frac{1}{n}\right)
$$

Proof. Due to the dominance of exponential rates, the argument can be proved by the Markov inequality. ■

Lemma A.2.4 *If Assumption 1 and 2 hold. For arbitrary* $\widetilde{M} \geq \frac{5\sigma^2}{2c_l^2}$*,*

$$\max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right| \geq \widetilde{M}\right) = o\left(\frac{1}{n}\right)$$

Proof. Due to the uniform dominance of innovations over the fixed effect in (3.2.1), we only need to consider the case where there is no individual fixed effect. In other words, it is equivalent to consider the decomposition in (3.2.1) with $\mu_i = 0$ for all $i = 1, 2, ..., n$. It follows that

$$\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 = \frac{1}{2c_i^0}\left\{\frac{\rho_i^{-2T}}{T^\gamma}y_{i,T}^2 - \frac{2\rho_i^{-2T+1}}{T^\gamma}\sum_{t=1}^{T}y_{i,t-1}u_{it} - \frac{\rho_i^{-2T}}{T^\gamma}\sum_{t=1}^{T}u_{it}^2\right\} - \frac{1}{\rho_i^{2T}T^{2\gamma-1}}\overline{y}_{i,-1}^2.$$

Therefore we can derive the following uniform upper bound as

$$\begin{aligned}
n\max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right| \geq \widetilde{M}\right) \leq\ & n\max_{1\leq i\leq n}\Pr\left(\left|\frac{1}{2c_i^0}\frac{\rho_i^{-2T}}{T^\gamma}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| \geq \widetilde{M}\right) \\
& + n\max_{1\leq i\leq n}\Pr\left(\left|\frac{1}{2c_i^0}\frac{2\rho_i^{-2T+1}}{T^\gamma}\sum_{t=1}^{T}y_{i,t-1}u_{it}\right| \geq \widetilde{M}\right) \\
& + n\max_{1\leq i\leq n}\Pr\left(\left|\frac{1}{2c_i^0}\frac{\rho_i^{-2T}}{T^\gamma}\sum_{t=1}^{T}u_{it}^2\right| \geq \widetilde{M}\right) \\
& + n\max_{1\leq i\leq n}\Pr\left(\left|\frac{1}{\rho_i^{2T}T^{2\gamma-1}}\overline{y}_{i,-1}^2\right| \geq \widetilde{M}\right) \\
& + n\max_{1\leq i\leq n}\Pr\left(\frac{1}{2c_i^0}\frac{\rho_i^{-2T}}{T^\gamma}\mathbb{E}y_{i,T}^2 \geq \widetilde{M}\right). \qquad \text{(A.2.8)}
\end{aligned}$$

The details follow similar procedures of Lemma A.2.8. We need the following rate restrictions as in Lemma A.2.8: $T^{2\gamma-2}n(\log n)^2 \to 0$, and $T^{\frac{5\gamma-3}{4}}n(\log n)^2 \to 0$. The additional needed rate restriction here is

$$n(\log n)^2 T^{1-3\gamma} \to 0,$$

as we hope to make sure that for each $i = 1, 2, ..., n$, the adjustment rate $\rho_i^{2T}T^{2\gamma}$ is larger than the upper bound of (BB.2.1) in Lemma A.2.8 as $n^{\frac{1}{4}}\sqrt{\log n}T^{\frac{1+5\gamma}{4}}\rho_i^{2T}$. The first and third terms of (A.2.8) can be proved by exponential inequality for martingale difference sequence (Freedman, 1975). The second term has been proved in Lemma A.2.3. The probability of the fourth term is $o_p(1)$ due the Markov inequality and the dominating exponential rates. For the fifth term of (A.2.8), if $\widetilde{M} > \frac{5\sigma^2}{2c_l^2}$,

$$n \max_{1\leq i\leq n} \Pr\left(\left|\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^\gamma}\mathbb{E}y_{i,T}^2\right| \geq \widetilde{M}\right) = 0.$$

All in all, in order to make sure

$$\max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right| \geq \widetilde{M}\right) = o\left(\frac{1}{n}\right),$$

we just need to make sure Assumptions 1 and 2 hold and $\widetilde{M}$ is large enough as $\widetilde{M} > \frac{5\sigma^2}{2c_l^2}$. ■

Lemma A.2.5 *If Assumption 1 and 2 hold, for any $\overline{M}$ satisfying $0 < \overline{M} \leq \frac{\sigma^2}{8c_u^2}$,*

$$\max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right| \leq \overline{M}\right) = o\left(\frac{1}{n}\right)$$

Proof. The proof follows the fashion in Lemma A.2.4 and accommodates the following decomposition

$$\begin{aligned}
& n \max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \leq \overline{M}\right) \\
\leq\ & n \max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \leq \overline{M}, \frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\mathbb{E}\widetilde{y}_{i,t-1}^2 > 2\overline{M}\right) \\
& + n \max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\mathbb{E}\widetilde{y}_{i,t-1}^2 \leq 2\overline{M}\right) \\
=\ & n \max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t-1}^2 - \mathbb{E}\widetilde{y}_{i,t-1}^2\right) \leq -\overline{M}\right) \\
& + n \max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\mathbb{E}\widetilde{y}_{i,t-1}^2 \leq 2\overline{M}\right) \\
\leq\ & n \max_{1\leq i\leq n} \Pr\left(\left|\frac{1}{2c_i^0}\frac{\rho_i^{-2T}}{T^\gamma}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| \geq \overline{M}\right)
\end{aligned}$$

$$
\begin{aligned}
&+n \max_{1\le i\le n} \Pr\left(\left|\frac{1}{2c_i^0}\frac{2\rho_i^{-2T+1}}{T^\gamma}\sum_{t=1}^{T} y_{i,t-1}u_{it}\right| \ge \overline{M}\right)\\
&+n \max_{1\le i\le n} \Pr\left(\left|\frac{1}{2c_i^0}\frac{\rho_i^{-2T}}{T^\gamma}\sum_{t=1}^{T} u_{it}^2\right| \ge \overline{M}\right)\\
&+n \max_{1\le i\le n} \Pr\left(\left|\frac{1}{\rho_i^{2T}T^{2\gamma-1}}\overline{y}_{i,-1}^2\right| \ge \overline{M}\right)\\
&+n \max_{1\le i\le n} \Pr\left(\frac{1}{2c_i^0}\frac{\rho_i^{-2T}}{T^\gamma}\mathbb{E}y_{i,T}^2 \le 2\overline{M}\right). \qquad \text{(A.2.9)}
\end{aligned}
$$

The details follow similar procedures of Lemma A.2.8. We need the following rate restrictions as in Lemma A.2.8: $T^{2\gamma-2}n(\log n)^2 \to 0$, and $T^{\frac{5\gamma-3}{4}}n(\log n)^2 \to 0$. The additional needed rate restriction here is

$$
n(\log n)^2 T^{1-3\gamma} \to 0,
$$

as we hope to make sure that for each $i = 1, 2, ..., n$, the adjustment rate $\rho_i^{2T}T^{2\gamma}$ is larger than the upper bound of (BB.2.1) in Lemma A.2.8 as $n^{\frac{1}{4}}\sqrt{\log n}T^{\frac{1+5\gamma}{4}}$. For any positive constant $\overline{M} > 0$, the first four terms in (A.2.9) are $o(1)$. The asymptotic negligibility can be proved in the identical way to Lemma A.2.4. If we define $2\overline{M} < \frac{\sigma^2}{4c_u^2}$, then

$$
n \max_{1\le i\le n} \Pr\left(\frac{1}{2c_i^0}\frac{\rho_i^{-2T}}{T^\gamma}\mathbb{E}y_{i,T}^2 \le 2\overline{M}\right) = 0.
$$

In all, we complete the proof. ■

Lemma A.2.6 *Suppose Assumption 1 and 2 hold. For some $\eta = O\left(\frac{1}{T^\gamma}\right)$, under joint convergence $(n,T) \to \infty$,*

$$
\sup_{\overline{c}\in\mathcal{N}_\eta} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{\widehat{g}_i(\overline{c}) \ne g_i^0\right\} \xrightarrow{p} 0.
$$

Proof. As the incidental parameter μ_i is of lower order, it is equivalent to consider the model where $\mu_i = 0$. For the definition of $\widehat{g}_i(\cdot)$, we have for all

$g \in \{1, 2, ..., K^0\}$,

$$\mathbf{1}\{\widehat{g}_i(\overline{c}) = g\} \leq \mathbf{1}\left\{\sum_{t=1}^{T}(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\overline{\rho}_g)^2 \leq \sum_{t=1}^{T}(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\overline{\rho}_{g_i^0})^2\right\}.$$

Define $\overline{\rho}_{g_i^0} := \exp\left(\overline{c}_{g_i^0}/T^\gamma\right)$ and $\overline{\rho}_{g_i} := \exp\left(\frac{\overline{c}_{g_i}}{T^\gamma}\right)$. For simplicity, we write $\rho_i^0 \left(:= \exp\left(\frac{c_i^0}{T^\gamma}\right)\right)$ as ρ_i. We derive the following transformation

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{\widehat{g}_i(\overline{c}) = g\} &= \sum_{g=1}^{K^0}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{g_i^0 \neq g\}\mathbf{1}\{\widehat{g}_i(\overline{c}) = g\} \\
&\leq \sum_{g=1}^{K^0}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{g_i^0 \neq g\}\mathbf{1}\left\{\sum_{t=1}^{T}(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\overline{\rho}_g)^2 \leq \sum_{t=1}^{T}(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\overline{\rho}_{g_i^0})^2\right\} \\
&=: \sum_{g=1}^{K^0}\frac{1}{n}\sum_{i=1}^{n}Z_{ig}(\overline{c}),
\end{aligned}$$

where $Z_{ig}(\overline{c}) := \mathbf{1}\{g_i^0 \neq g\}\mathbf{1}\left\{\sum_{t=1}^{T}(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\overline{\rho}_g)^2 \leq \sum_{t=1}^{T}(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\overline{\rho}_{g_i^0})^2\right\}$. We intend to bound $Z_{ig}(\overline{c})$ for all $\overline{c} \in \mathcal{N}_\eta$ by the quantity irrelevant to $\overline{c}$. Therefore for all i, it has

$$\begin{aligned}
Z_{ig}(\overline{c}) &\leq \max_{\widetilde{g} \neq g}\mathbf{1}\left\{\sum_{t=1}^{T}\left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\overline{\rho}_g\right)^2 \leq \sum_{t=1}^{T}\left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}\right)^2\right\} \\
&= \max_{\widetilde{g} \neq g}\mathbf{1}\left\{\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} - \overline{\rho}_g\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 + 2\widetilde{u}_{it} - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} + \overline{\rho}_g\right)\right) \leq 0\right\}.
\end{aligned}$$

Let us define

$$\begin{aligned}
H_T &:= \left|\begin{array}{c} \sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} - \overline{\rho}_g\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 + 2\widetilde{u}_{it} - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} + \overline{\rho}_g\right)\right) \\ -\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 - \overline{\rho}_g^0\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 + 2\widetilde{u}_{it} - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 + \overline{\rho}_g^0\right)\right) \end{array}\right| \\
&\leq \left|2\sum_{t=1}^{T}(\overline{\rho}_{\widetilde{g}} - \overline{\rho}_g)\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| + \left|2\sum_{t=1}^{T}(\overline{\rho}_{\widetilde{g}}^0 - \overline{\rho}_g^0)\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \\
&\quad + \left|\begin{array}{c} \sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} - \overline{\rho}_g\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} + \overline{\rho}_g\right)\right) \\ -\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 - \overline{\rho}_g^0\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 + \overline{\rho}_g^0\right)\right) \end{array}\right| \\
&=: H_{1T} + H_{2T} + H_{3T},
\end{aligned}$$

where $H_{1T} := \left|2\sum_{t=1}^{T}(\overline{\rho}_{\widetilde{g}} - \overline{\rho}_g)\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|$, $H_{2T} := \left|2\sum_{t=1}^{T}(\overline{\rho}_{\widetilde{g}}^0 - \overline{\rho}_g^0)\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|$ and

$$H_{3T} := \left| \begin{array}{c} \sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} - \overline{\rho}_g\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} + \overline{\rho}_g\right)\right) \\ -\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 - \overline{\rho}_g^0\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 + \overline{\rho}_g^0\right)\right) \end{array} \right|.$$

By the compactness of the parameter support, we have

$$H_{1T} = \left|2\sum_{t=1}^{T}(\overline{\rho}_{\widetilde{g}} - \overline{\rho}_g)\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \leq 2\left|\overline{\rho}_{\widetilde{g}} - \overline{\rho}_g\right|\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \leq \frac{B_1}{T^\gamma}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|,$$

where B_1 is a constant independent of η and T. We have the above argument by the definition of η. Similarly we can justify that $H_{2T} \leq \frac{B_2}{T^\gamma}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|$ where B_2 is a constant independent of η and T. For H_{3T}, we have with B_3 as a constant independent of η and T,

$$\begin{aligned} H_{3T} &= \left| \begin{array}{c} \sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} - \overline{\rho}_g\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} + \overline{\rho}_g\right)\right) \\ -\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 - \overline{\rho}_g^0\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 + \overline{\rho}_g^0\right)\right) \end{array} \right| \\ &= \left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right|\left|\overline{\rho}_{\widetilde{g}}^0\left(\overline{\rho}_{\widetilde{g}} - \overline{\rho}_{\widetilde{g}}^0 - \overline{\rho}_g + \overline{\rho}_g^0\right) + \frac{1}{2}\left(\left(\overline{\rho}_{\widetilde{g}}^0\right)^2 - \overline{\rho}_{\widetilde{g}}^2 + \overline{\rho}_g^2 - \left(\overline{\rho}_g^0\right)^2\right)\right| \\ &\leq \frac{B_3}{T^\gamma}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right|. \end{aligned}$$

By combining available results, we obtain the following,

$$\begin{aligned} Z_{ig}(\overline{c}) &\leq \max_{\widetilde{g}\neq g}\mathbf{1}\left\{\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} - \overline{\rho}_g\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 + 2\widetilde{u}_{it} - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}} + \overline{\rho}_g\right)\right) \leq 0\right\} \\ &\leq \max_{\widetilde{g}\neq g}\mathbf{1}\left\{ \begin{array}{c} \sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 - \overline{\rho}_g^0\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 + 2\widetilde{u}_{it} - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 + \overline{\rho}_g^0\right)\right) \\ \leq \frac{B_1+B_2}{T^\gamma}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| + \frac{B_3}{T^\gamma}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right| \end{array} \right\}. \end{aligned}$$

Based on the following fact that

$$\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 - \overline{\rho}_g^0\right)\left(2\widetilde{y}_{i,t-1}\overline{\rho}_{\widetilde{g}}^0 - \widetilde{y}_{i,t-1}\left(\overline{\rho}_{\widetilde{g}}^0 + \overline{\rho}_g^0\right)\right) = \sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\overline{\rho}_{\widetilde{g}}^0 - \overline{\rho}_g^0\right)^2,$$

we define the following argument as

$$\widetilde{Z}_{ig} := \max_{\widetilde{g}\neq g} \mathbf{1}\left\{ \begin{array}{c} \left(\overline{\rho}_{\widetilde{g}}^{0} - \overline{\rho}_{g}^{0}\right)^{2} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^{2} + 2\left(\overline{\rho}_{\widetilde{g}}^{0} - \overline{\rho}_{g}^{0}\right) \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \\ \leq \frac{B_1+B_2}{T^{\gamma}}\eta \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| + \frac{B_3}{T^{\gamma}}\eta \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}^{2}\right| \end{array} \right\}.$$

Consequently, we can bound $Z_{ig}(\overline{c})$ by $\sup_{\overline{c}\in\mathcal{N}_{\eta}} Z_{ig}(\overline{c}) \leq \widetilde{Z}_{ig}$. Note that

$$\sup_{\overline{c}\in\mathcal{N}_{\eta}} \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{g}_i(\overline{c}) \neq g_i^0\right\} \leq \frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{K^0}\widetilde{Z}_{ig}.$$

For $i = 1, 2, ..., n$, $g_i^0 = \widetilde{g} \neq g$, we have equivalent representations $\rho_i = \overline{\rho}_{\widetilde{g}}$. For all $g \in \{1, 2, ...K^0\}$, we have

$$\begin{aligned}
& \Pr\left(\widetilde{Z}_{ig} = 1\right) \\
\leq\ & \sum_{\widetilde{g}\neq g}\Pr\left(\begin{array}{c} \left(\overline{\rho}_{\widetilde{g}}^{0} - \overline{\rho}_{g}^{0}\right)^{2} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^{2} + 2\left(\overline{\rho}_{\widetilde{g}}^{0} - \overline{\rho}_{g}^{0}\right) \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \\ \leq \frac{B_1+B_2}{T^{\gamma}}\eta \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| + \frac{B_3}{T^{\gamma}}\eta \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}^{2}\right| \end{array} \right) \\
\leq\ & \sum_{\widetilde{g}\neq g}\Pr\left(\begin{array}{c} 2\left(\overline{\rho}_{\widetilde{g}}^{0} - \overline{\rho}_{g}^{0}\right) \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \leq -\left(\overline{\rho}_{\widetilde{g}}^{0} - \overline{\rho}_{g}^{0}\right)^{2} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^{2} \\ +\frac{B_1+B_2}{T^{\gamma}}\eta \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| + \frac{B_3}{T^{\gamma}}\eta \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}^{2}\right| \end{array} \right) \\
\leq\ & \sum_{\widetilde{g}\neq g}\Pr\left(\begin{array}{c} 2\frac{\left(\overline{c}_{\widetilde{g}}^{0} - \overline{c}_{g}^{0}\right)}{T^{\gamma}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \leq -\left(\overline{c}_{\widetilde{g}}^{0} - \overline{c}_{g}^{0}\right)^{2} \left(\overline{\rho}_{\widetilde{g}}^{0}\right)^{2T} \overline{M} \\ +(B_1+B_2)\left(\overline{\rho}_{\widetilde{g}}^{0}\right)^{2T}\eta M + B_3 T^{\gamma}\left(\overline{\rho}_{\widetilde{g}}^{0}\right)^{2T}\eta\widetilde{M} \end{array} \right) \\
& +\sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^{\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2} \leq \overline{M}\right) + \sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^{\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \geq M\right) \\
& +\sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^{\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\right| \geq \widetilde{M}\right) \\
\leq\ & \sum_{\widetilde{g}\neq g}\Pr\left(\begin{array}{c} 2\frac{c^*}{T^{\gamma}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \leq -(c^*)^{2} \left(\overline{\rho}_{\widetilde{g}}^{0}\right)^{2T} \overline{M} \\ +(B_1+B_2)\left(\overline{\rho}_{\widetilde{g}}^{0}\right)^{2T}\eta M + B_3 T^{\gamma}\left(\overline{\rho}_{\widetilde{g}}^{0}\right)^{2T}\eta\widetilde{M} \end{array} \right) + \sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^{\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2} \leq \overline{M}\right) \\
& +\sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^{\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \geq M\right) + \sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^{\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\right| \geq \widetilde{M}\right). \qquad \text{(A.2.10)}
\end{aligned}$$

Based on Lemma A.2.3, A.2.4 and A.2.5 we can argue the 2^{nd}, 3^{rd} and 4^{th} terms of (A.2.10) are $o\left(\frac{1}{n}\right)$. We restrict η as $\eta \leq \frac{(c^*)^2\overline{M}}{2B_3T^\gamma\widetilde{M}}$. For example, we can set $\eta = \frac{(c^*)^2\overline{M}}{4B_3T^\gamma\widetilde{M}}$. Therefore we have

$$\begin{aligned}
& \sum_{\widetilde{g}\neq g} \Pr\left(\begin{array}{c} 2\frac{c^*}{T^\gamma}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it} \leq -\left(c^*\right)^2\left(\overline{\rho}_{\widetilde{g}}^0\right)^{2T}\overline{M} \\ +(B_1+B_2)\left(\overline{\rho}_{\widetilde{g}}^0\right)^{2T}\eta M + B_3T^\gamma\left(\overline{\rho}_{\widetilde{g}}^0\right)^{2T}\eta\widetilde{M} \end{array}\right) \\
= & \sum_{\widetilde{g}\neq g} \Pr\left(\begin{array}{c} 2\frac{c^*}{T^\gamma\left(\overline{\rho}_{\widetilde{g}}^0\right)^{2T}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it} \leq -\left(c^*\right)^2\overline{M} \\ +(B_1+B_2)\eta M + B_3T^\gamma\eta\widetilde{M} \end{array}\right) \\
\leq & \sum_{\widetilde{g}\neq g} \Pr\left(2\frac{c^*}{T^\gamma\rho_i^{2T}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it} \leq \frac{-\left(c^*\right)^2\overline{M}}{2}\right) \\
\leq & \sum_{\widetilde{g}\neq g} \Pr\left(\frac{1}{T^\gamma\rho_i^{2T}}\left|\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \geq \frac{c^*\overline{M}}{4}\right) \\
= & \; o\left(\frac{1}{n}\right).
\end{aligned}$$

The last equality is due to Lemma A.2.3. Combining the above results, we obtain $\Pr\left(\widetilde{Z}_{ig}=1\right) = o\left(\frac{1}{n}\right)$. This implies that

$$\begin{aligned}
\sup_{\overline{c}\in\mathcal{N}_\eta}\frac{1}{n}\sum_{i=1}^n \mathbf{1}\left\{\widehat{g}_i(\overline{c})\neq g_i^0\right\} &\leq \frac{1}{n}\sum_{g=1}^{K^0}\sum_{i=1}^n \mathbb{E}\widetilde{Z}_{ig} = \frac{1}{n}\sum_{g=1}^{K^0}\sum_{i=1}^n \Pr\left(\widetilde{Z}_{ig}=1\right) = K^0\left(K^0-1\right)o\left(\frac{1}{n}\right) \\
&= o\left(\frac{1}{n}\right).
\end{aligned}$$

■

Assume that $\widehat{G}_g$ represents the g-th estimated classification group and G_g^0 denotes the g-th true group in the population. To rigorously state the uniform consistency of the recursive k-means algorithm, we define the following sequences of events as

$$\widehat{E}_{g,i} := \left\{i \notin \widehat{G}_g | i \in G_g^0\right\} \text{ and } \widehat{F}_{g,i} := \left\{i \notin G_g^0 | i \in \widehat{G}_g\right\}, \tag{A.2.11}$$

for $g = 1, 2, ..., K^0$ and $i = 1, 2, ..., n$. Let $\widehat{E}_{g,nT} := \bigcup_{i\in G_g^0}\widehat{E}_{g,i}$ and $\widehat{F}_{g,nT} := \bigcup_{i\in\widehat{G}_g}\widehat{F}_{g,i}$. $\widehat{E}_{g,nT}$ demonstrates the error event of not classifying the individual unit of G_g^0 into $\widehat{G}_g$ as a type-I classification error; $\widehat{E}_{g,nT}$ demonstrates the error event of not clustering the cross-sectional unit of $\widehat{G}_g$ into G_g^0 as a type-II classification error. Furthermore, we establish the uniform consistency of the recursive k-means classifier.

Lemma A.2.7 *(Uniform Consistency of Classification) Let Assumption 1 and 2 hold and $c_i^0 > 0$ for each $i = 1, 2, ..., n$. Under joint convergence $(n, T) \to \infty$,*

(i) $\Pr\left(\bigcup_{g=1}^{K^0} \widehat{E}_{g,nT}\right) \leq \sum_{g=1}^{K^0} \Pr\left(\widehat{E}_{g,nT}\right) \to 0,$

(ii) $\Pr\left(\bigcup_{g=1}^{K^0} \widehat{F}_{g,nT}\right) \leq \sum_{g=1}^{K^0} \Pr\left(\widehat{F}_{g,nT}\right) \to 0.$

Lemma A.2.7 illustrates that for all $g \in \{1, 2, ..., K^0\}$ all cross-sectional units belonging to group G_g^0 are assigned into the same estimated group $\widehat{G}_g$ asymptotically. Meanwhile, all cross-sectional agents classified into the same group $\widehat{G}_g$ for all $g \in \{1, 2, ..., K^0\}$ belong to the same group G_g^0 in the probability limit. These observations show that the summation of classification errors is diminishing asymptotically.

Proof. For the uniform consistency, observe that $\Pr\left(\bigcup_{g=1}^{K^0} \widehat{E}_{g,nT}\right) \leq \sum_{g=1}^{K^0} \Pr\left(\widehat{E}_{g,nT}\right) \leq \sum_{g=1}^{K^0} \sum_{i \in G_g^0} \Pr\left(\widehat{E}_{g,i}\right)$. We have

$$\begin{aligned}
\sum_{g=1}^{K^0} \sum_{i \in G_g^0} \Pr\left(\widehat{E}_{g,i}\right) &\leq n \max_{1 \leq i \leq n} \mathbb{E}\mathbf{1}\{\widehat{g}_i(\widehat{\overline{c}}) \neq g_i^0\} = n \max_{1 \leq i \leq n} \Pr\{\left|\widehat{g}_i(\widehat{\overline{c}}) - g_i^0\right| > 0\} \\
&\leq n \max_{1 \leq i \leq n} \sup_{\overline{c} \in \mathcal{N}_\eta} \Pr\{\left|\widehat{g}_i(\overline{c}) - g_i^0\right| > 0\} + n \max_{1 \leq i \leq n} \Pr\{\left|\widehat{\overline{c}}_{\widehat{g}_i} - \overline{c}_{g_i^0}^0\right| > \eta\} \\
&= o(1) + n \max_{1 \leq i \leq n} \Pr\{\left|\widehat{\overline{c}}_{\widehat{g}_i} - \overline{c}_{g_i^0}^0\right| > \eta\}. \qquad \text{(A.2.12)}
\end{aligned}$$

It remains to show that the second term of (A.2.12) is asymptotically negligible. Note the fact that $\eta = O\left(\frac{1}{T^\gamma}\right)$. By Lemma A.1.1, for any $i = 1, 2, ..., n$, the estimate $\widehat{\overline{c}}$ generated by the recursive $k-$means algorithm converges to the true value at the rate of $O_p\left(\frac{1}{\sqrt{n}\left(\rho_i^0\right)^T T^\gamma}\right)$ which is smaller than the radius, η. Moreover, since the rate of $O_p\left(\frac{1}{\sqrt{n}\left(\rho_i^0\right)^T T^\gamma}\right)$ is dominating over the polynomial rate of n, the Markov inequality produces

$$n \max_{1\le i\le n} \Pr\{\left|\widehat{\overline{c}}_{\widehat{g}_i} - c^0_{g^0_i}\right| > \eta\} = o(1).$$

Then we successfully justify the argument of (i);

For (ii), we basically follow the derivations in Su et al. (2016) and can easily derive the need results ■

A.2.2 Proof of Theorem 3.4.1:

Proof. (i) Note the definition of $\widehat{g}_i$,

$$\Pr\left(\max_{1\le i\le n}\left|\widehat{g}_i(\widehat{\overline{c}}) - g^0_i\right| > 0\right) \le \Pr\left(\widehat{\overline{c}} \notin \mathcal{N}_\eta\right) + \mathbb{E}\left[\sup_{c\in\mathcal{N}_\eta} \Pr\left(\max_{1\le i\le n}\left|\widehat{g}_i(\overline{c}) - g^0_i\right| > 0\right)\right].$$

Based on the proof of Lemma A.2.6, we know that $\eta = O\left(\frac{1}{T^\gamma}\right)$ asymptotically. The convergence rates of our estimates are fast enough to satisfy this condition as $\widehat{\overline{c}}_g - \overline{c}^0_g = O_p\left(\frac{1}{\sqrt{n}\left(\overline{\rho}^0_g\right)^T T^\gamma}\right) = o_p\left(\frac{1}{T^\gamma}\right)$ for $g = 1, 2, ..., K^0$. Therefore we derive the following argument as,

$$\Pr\left(\widehat{\overline{c}} \notin \mathcal{N}_\eta\right) = o(1).$$

Besides,

$$\begin{aligned} \sup_{\overline{c}\in\mathcal{N}_\eta} \Pr\left(\max_{1\le i\le n}\left|\widehat{g}_i(\overline{c}) - g^0_i\right| > 0\right) &\le n \sup_{\overline{c}\in\mathcal{N}_\eta} \max_{1\le i\le n} \Pr\left(\left|\widehat{g}_i(\overline{c}) - g^0_i\right| > 0\right) \\ &= n \max_{1\le i\le n} \sup_{\overline{c}\in\mathcal{N}_\eta} \Pr\left(\left|\widehat{g}_i(\overline{c}) - g^0_i\right| > 0\right) \\ &= n \cdot o\left(\frac{1}{n}\right) = o\left(1\right). \end{aligned}$$

Then we completed the proof. ■

A.2.3 Proof of Theorem 3.4.2:

Proof. Based on the definition of the post-classification estimator, we have for each $g \in \{1, 2, ..., K^0\}$ and $\bar{c}_g > 0$

$$\sqrt{n_g} T^\gamma (\bar{\rho}_g^0)^T \left(\widehat{\bar{\rho}}_{\widehat{g}} - \bar{\rho}_g^0 \right) = \frac{\frac{1}{\sqrt{n_g} T^\gamma (\bar{\rho}_g^0)^T} \sum_{i \in \widehat{G}_{\widehat{g}}} \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it}}{\frac{1}{n_g T^{2\gamma} (\bar{\rho}_g^0)^{2T}} \sum_{i \in \widehat{G}_{\widehat{g}}} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2}.$$

The numerator and denominator can be decomposed as

$$\begin{aligned} & \frac{1}{\sqrt{n_g} T^\gamma (\bar{\rho}_g^0)^T} \sum_{i \in \widehat{G}_{\widehat{g}}} \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it} \\ = \; & \frac{1}{\sqrt{n_g} T^\gamma (\bar{\rho}_g^0)^T} \sum_{i \in G_g^0} \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it} + \frac{1}{\sqrt{n_g} T^\gamma (\bar{\rho}_g^0)^T} \sum_{\substack{\widetilde{g}=1, \\ \widetilde{g} \neq g}}^{K^0} \sum_{\substack{i \in \widehat{G}_{\widehat{g}} \setminus G_g^0 \\ i \in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0 < \alpha_g^0}} \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it} \\ & + \frac{1}{\sqrt{n_g} T^\gamma (\bar{\rho}_g^0)^T} \sum_{\substack{\widetilde{g}=1, \\ \widetilde{g} \neq g}}^{K^0} \sum_{\substack{i \in \widehat{G}_{\widehat{g}} \setminus G_g^0 \\ i \in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0 \geq \alpha_g^0}} \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it} - \frac{1}{\sqrt{n_g} T^\gamma (\bar{\rho}_g^0)^T} \sum_{\widehat{g}=1}^{K^0} \sum_{i \in G_g^0 \setminus \widehat{G}_{\widehat{g}}} \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it}, \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{n_g T^{2\gamma} (\bar{\rho}_g^0)^{2T}} \sum_{i \in \widehat{G}_g} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2 \\ = \; & \frac{1}{n_g T^{2\gamma} (\bar{\rho}_g^0)^{2T}} \sum_{i \in G_g^0} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2 + \frac{1}{n_g T^{2\gamma} (\bar{\rho}_g^0)^{2T}} \sum_{\substack{\widetilde{g}=1, \\ \widetilde{g} \neq g}}^{K^0} \sum_{\substack{i \in \widehat{G}_g \setminus G_g^0 \\ i \in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0 < \alpha_g^0}} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2 \\ & + \frac{1}{n_g T^{2\gamma} (\bar{\rho}_g^0)^{2T}} \sum_{\substack{\widetilde{g}=1, \\ \widetilde{g} \neq g}}^{K^0} \sum_{\substack{i \in \widehat{G}_g \setminus G_g^0 \\ i \in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0 \geq \alpha_g^0}} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2 - \frac{1}{n_g T^{2\gamma} (\bar{\rho}_g^0)^{2T}} \sum_{\widehat{g}=1}^{K^0} \sum_{i \in G_g^0 \setminus \widehat{G}_g} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2. \end{aligned}$$

Therefore it remains to justify these arguments:

(i) For any $g = 1, 2, ..., K^0$,

$$\frac{1}{\sqrt{n_g} T (\bar{\rho}_g^0)^T} \sum_{\substack{i \in \widehat{G}_g \setminus G_g^0 \\ i \in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0 < \alpha_g^0}} \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it} = o_p(1),$$

$$\frac{1}{n_g T^{2\gamma}(\overline{\rho}_g^0)^{2T}} \sum_{\substack{i \in \widehat{G}_g \setminus G_g^0 \\ i \in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0 < \alpha_g^0}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 = o_p(1);$$

(ii) For any $g = 1, 2, ..., K^0$,

$$\frac{1}{\sqrt{n_g} T^{\gamma}(\overline{\rho}_g^0)^{T}} \sum_{\substack{i \in \widehat{G}_g \setminus G_g^0 \\ i \in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0 \geq \alpha_g^0}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} = o_p(1),$$

$$\frac{1}{n_g T^{2\gamma}(\overline{\rho}_g^0)^{2T}} \sum_{\substack{i \in \widehat{G}_g \setminus G_g^0 \\ i \in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0 \geq \alpha_g^0}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 = o_p(1);$$

(iii) For any $g = 1, 2, ..., K^0$,

$$\frac{1}{\sqrt{n_g} T^{\gamma}(\overline{\rho}_g^0)^{T}} \sum_{i \in G_g^0 \setminus \widehat{G}_{\widehat{g}}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} = o_p(1),$$

$$\frac{1}{n_g T^{2\gamma}(\overline{\rho}_g^0)^{2T}} \sum_{i \in G_g^0 \setminus \widehat{G}_{\widehat{g}}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 = o_p(1).$$

For (i) (ii) and (iii), second terms can be proved identically as the first ones. Without losing generality we just focus on the first terms. For (iii), we have for any $\varepsilon > 0$

$$\Pr\left(\left|\frac{1}{\sqrt{n_g} T^{\gamma}(\overline{\rho}_g^0)^{T}} \sum_{i \in G_g^0 \setminus \widehat{G}_g} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it}\right| > \varepsilon\right) \leq \Pr\left(\sum_{g=1}^{K^0} \widehat{E}_{g,nT}\right) \to 0,$$

under joint asymptotics. For (i), for any $\widetilde{g} \neq g = 1, 2, ..., K^0$ and for any $\varepsilon > 0$, we have

$$\Pr\left(\left|\frac{1}{\sqrt{n_g} T^{\gamma}(\overline{\rho}_g^0)^{T}} \sum_{\substack{\widetilde{g}=1, \\ \widetilde{g} \neq g}}^{K^0} \sum_{\substack{i \in \widehat{G}_g \setminus G_g^0 \\ i \in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0 < \alpha_g^0}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it}\right| > \varepsilon\right) \leq \Pr\left(\sum_{\widetilde{g}=1}^{K^0} \widehat{F}_{\widetilde{g},nT}\right) \to 0,$$

under joint asymptotic framework. For (ii), for any $\varepsilon > 0$, we have

$$\Pr\left(\left|\frac{1}{\sqrt{n_g}T^{\gamma}(\overline{\rho}_g^0)^T}\sum_{\substack{\widetilde{g}=1,\\ \widetilde{g}\neq g}}^{K^0}\sum_{\substack{i\in\widehat{G}_g\setminus G_g^0\\ i\in G_{\widetilde{g}}^0,\alpha_{\widetilde{g}}^0\geq\alpha_g^0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|>\varepsilon\right)\leq\Pr\left(\sum_{\widetilde{g}=1}^{K^0}\widehat{F}_{\widetilde{g},nT}\right)\to 0.$$

Summarizing above results,

$$\frac{1}{\sqrt{n_g}T^{\gamma}(\overline{\rho}_g^0)^T}\sum_{i\in\widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}=\frac{1}{\sqrt{n_g}T^{\gamma}(\overline{\rho}_g^0)^T}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}+o_p(1).$$

and

$$\frac{1}{n_gT^{2\gamma}(\overline{\rho}_g^0)^{2T}}\sum_{i\in\widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2=\frac{1}{n_gT^{2\gamma}(\overline{\rho}_g^0)^{2T}}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2+o_p(1).$$

The asymptotics for the post-classification estimator is asymptotically equivalent to the infeasible within estimator. The joint limit $(n, T) \to \infty$ for the infeasible within estimator follows Phillips and Moon (1999). ■

A.2.4 Technical lemmas of modified k-means algorithm

We also intend to establish the asymptotic properties of the mixed-root panel autoregressions. We collect the individual least squares $\{\widehat{c}_i\}_{i=1}^n$ and discuss their uniform bound.

Lemma A.2.8 *If Assumptions 1 and 2 hold,*

$$\sup_{1\leq i\leq n,c_i^0>0}\left|\widehat{c}_i-c_i^0\right|=O_p\left((\rho_u)^{-T}T^{\frac{-\gamma}{2}}\right),$$

where $\rho_u := \exp\left(\frac{c_u}{T^{\gamma}}\right)$.

Proof. If $c_i > 0$, there is one explosive root. The time series estimator is as,

$$\widehat{c}_i - c_i^0 = T^\gamma \left(\sum_{t=1}^T \widetilde{y}_{i,t-1}^2 \right)^{-1} \left(\sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right),$$

and

$$\begin{aligned} \sup_{1\le i\le n} \left| \widehat{c}_i - c_i^0 \right| &\le T^\gamma \left(\inf_{1\le i\le n} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2 \right)^{-1} \left(\sup_{1\le i\le n} \left| \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \right) \\ &= T^\gamma \left(\inf_{1\le i\le n} \frac{T^\gamma}{2c_i^0} \mathbb{E} y_{i,T}^2 - \sup_{1\le i\le n} \left(\frac{T^\gamma}{2c_i^0} \mathbb{E} y_{i,T}^2 - \sum_{t=1}^T \widetilde{y}_{i,t-1}^2 \right) \right)^{-1} \left(\sup_{1\le i\le n} \left| \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \right). \end{aligned}$$

(i) To justify the uniform upper bound for $\left| \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right|$, we apply the exponential inequality of Freedman (1975) as

$$\begin{aligned} &\Pr \left(\sup_{1\le i\le n} \left| \sum_{t=1}^T \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \ge M_{nT} \right) \\ \le\ &\Pr \left(\sup_{1\le i\le n} \left| \sum_{t=1}^T y_{i,t-1} u_{it} \right| \ge M_{nT} \right) + \Pr \left(\sup_{1\le i\le n} \left| T \overline{y}_{i,-1} \overline{u}_i \right| \ge M_{nT} \right) \\ \le\ &\underbrace{n \sup_{1\le i\le n} \Pr \left(\left| \sum_{t=1}^T y_{i,t-1} u_{it} \right| \ge M_{nT} \right)}_{\text{(AA.1)}} + \underbrace{n \sup_{1\le i\le n} \Pr \left(\left| T \overline{y}_{i,-1} \overline{u}_i \right| \ge M_{nT} \right)}_{\text{(AA.2)}}, \end{aligned}$$

where the innovations uniformly dominate the fixed effect. For (AA.1) we apply the exponential inequality of Freedman (1975). The process $\{y_{i,t-1} u_{it}\}_{t=1}^T$ is a martingale difference sequence as $\mathbb{E}\left(y_{i,t-1} u_{it} | \mathcal{F}_{i,t-1} \right) = 0$ with $\mathcal{F}_{i,t-1} := \sigma \left\{ u_{i,t-1}, u_{i,t-2}, ... \right\}$. We set a truncation rate $d_{i,nT} = \rho_i^T n^{\frac{1}{4}} T^{\frac{1+2\gamma}{4}}$. We define $z_{it} := y_{i,t-1} u_{it}$ and make the following decomposition $\sum_{t=1}^T z_{it} = \sum_{t=1}^T z_{1it} + \sum_{t=1}^T z_{2it} - \sum_{t=1}^T \mathbb{E}\left[z_{2it} | \mathcal{F}_{i,t-1} \right]$. Define $z_{1it} := z_{it} \mathbf{1}_{it} - \mathbb{E}\left[z_{it} \mathbf{1}_{it} | \mathcal{F}_{i,t-1} \right]$ and $z_{2it} := z_{it} \overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it} := \mathbf{1}\left\{ |z_{it}| \le d_{i,nT} \right\}$ and $\overline{\mathbf{1}}_{it} = 1 - \mathbf{1}_{it}$. We need to derive the uniform upper bound from the following negligible conditions. It suffices to find M_{nT} to ensure

$$n \sup_{1\le i\le n} \Pr \left(\left| \sum_{t=1}^T z_{1it} \right| \ge M_{nT} \right) = o(1),$$

$$
\begin{aligned}
n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} z_{2it}\right| \ge M_{nT}\right) &= o(1), \\
n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \mathbb{E}\left[z_{2it}|\mathcal{F}_{i,t-1}\right]\right| \ge M_{nT}\right) &= o(1). \qquad \text{(A.2.13)}
\end{aligned}
$$

The second and third arguments of (A.2.13) share identical derivations, and without losing generality we only focus on the second term. We define $V_{iT} := \sum_{t=1}^{T} \mathbb{E}\left[z_{1it}^2|\mathcal{F}_{i,t-1}\right]$, and $v_{i,nT} = \rho_i^{2T} n^{\frac{1}{2}} T^{\frac{1}{2}+\frac{3\gamma}{2}}$ as a truncation rate for V_{iT}.

$$
\begin{aligned}
\mathbb{E}\left[V_{iT}^2\right] &= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[z_{1it}^2|\mathcal{F}_{i,t-1}\right]\right]^2 \le T\sum_{t=1}^{T} \mathbb{E}\left[z_{1it}^4\right] \le 16T\sum_{t=1}^{T} \mathbb{E}\left[z_{it}^4\right] \\
&\le CT^{1+2\gamma}\sum_{t=1}^{T} \rho_i^{4t} = O_p\left(T^{1+3\gamma}\rho_i^{4T}\right).
\end{aligned}
$$

By Proposition 2.1 in Freedman (1975),

$$
\begin{aligned}
& n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} z_{1it}\right| \ge M_{nT}\right) \\
\le\ & n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} z_{1it}\right| \ge M_{nT}, V_{iT} \le v_{i,nT}\right) + n \sup_{1\le i\le n} \Pr\left(V_{iT} > v_{i,nT}\right) \\
\le\ & \sup_{1\le i\le n} \exp\left(\frac{-M_{nT}^2 + 2v_{i,nT}\log(n) + 4M_{nT}d_{i,nT}\log(n)}{2v_{i,nT} + 4M_{nT}d_{i,nT}}\right) + \sup_{1\le i\le n} o\left(nT^{1+3\gamma}\rho_i^{4T}v_{i,nT}^{-2}\right) \\
=\ & o(1). \qquad \text{(A.2.14)}
\end{aligned}
$$

To show asymptotic negligibility of (A.2.14), we need $M_{nT} \succeq \sup_{1\le i\le n} \rho_i^T n^{\frac{1}{4}} T^{\frac{1+3\gamma}{4}} \sqrt{\log(n)}$. For the second term of (A.2.13), we have

$$
\begin{aligned}
& n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} z_{2it}\right| \ge M_{nT}\right) \\
\le\ & n \sup_{1\le i\le n} \Pr\left(\max_{1\le t\le T} |z_{it}| \ge d_{i,nT}\right) \le \frac{nT}{d_{i,nT}^4} \sup_{1\le i\le n} \max_{1\le t\le T} \mathbb{E}\left[|z_{it}|^4 1\left\{|z_{it}| > d_{i,nT}\right\}\right] \\
=\ & \sup_{1\le i\le n} \max_{1\le t\le T} o\left(\frac{nT^{1+2\gamma}}{d_{i,nT}^4}\rho_i^{4t}\right) = o(1),
\end{aligned}
$$

which is guaranteed by our assumption for $d_{i,nT}$. It is easily justified that

$$\sup_{1\leq i\leq n} \rho_i^T n^{\frac{1}{4}} T^{\frac{1+3\gamma}{4}} \sqrt{\log(n)} \succ \sup_{1\leq i\leq n} \rho_i^T n^{\frac{1}{4}} T^{\frac{1+2\gamma}{4}} \log(n).$$

Therefore, for (AA.1) we need $M_{nT} \succeq \rho_u^T n^{\frac{1}{4}} T^{\frac{1+3\gamma}{4}+\epsilon} \sqrt{\log(n)}$ for any $\epsilon > 0$.

For (AA.2) term, the uniform upper bound follows decompositions as

$$\sup_{1\leq i\leq n} \left|T\overline{y}_{i,-1}\overline{u}_i\right| \leq T \sup_{1\leq i\leq n} \left|\overline{y}_{i,-1}\right| \sup_{1\leq i\leq n} \left|\overline{u}_i\right|.$$

For $\sup_{1\leq i\leq n} |\overline{u}_i|$ term, the exponential inequality of Freedman (1975) is easily applied, and shows $\sup_{1\leq i\leq n} |\overline{u}_i| = O_p\left(T^{-\frac{1}{2}} n^{\frac{1}{4}} \sqrt{\log n}\right)$. For the term $\sup_{1\leq i\leq n} \left|\overline{y}_{i,-1}\right|$,

$$\sum_{t=1}^{T} y_{i,t-1} = \sum_{t=1}^{T}\left(\sum_{s=0}^{t-1} \rho_i^{t-1-s} u_{is}\right) = \frac{T^\gamma}{c_i^0} \sum_{s=0}^{T-1} \left(\rho_i^{T-s} u_{is} - u_{is}\right).$$

Therefore,

$$\begin{aligned}\sup_{1\leq i\leq n} \left|\sum_{t=1}^{T} y_{i,t-1}\right| &= \sup_{1\leq i\leq n} \left|\frac{T^\gamma}{c_i^0} \sum_{s=0}^{T-1} \left(\rho_i^{T-s} u_{is} - u_{is}\right)\right| \\ &\leq \sup_{1\leq i\leq n} \left|\frac{T^\gamma \rho_i^T}{c_i^0} \sum_{s=0}^{T-1} \rho_i^{-s} u_{is}\right| + \sup_{1\leq i\leq n} \left|\frac{T^\gamma}{c_i^0} \sum_{s=0}^{T-1} u_{is}\right|.\end{aligned}$$

For the term of $\sup_{1\leq i\leq n} \left|\frac{T^\gamma}{c_i^0} \sum_{s=0}^{T-1} u_{is}\right|$, the upper bound is $O_p\left(n^{\frac{1}{4}} T^{\frac{1}{2}+\gamma}\right)$. For the term of $\sup_{1\leq i\leq n} \left|\frac{T^\gamma \rho_i^T}{c_i^0} \sum_{s=0}^{T-1} \rho_i^{-s} u_{is}\right|$, the exponential inequality of martingale (Freedman, 1975) applies and shows

$$\sup_{1\leq i\leq n} \left|\frac{T^\gamma \rho_i^T}{c_i^0} \sum_{s=0}^{T-1} \rho_i^{-s} u_{is}\right| = O_p\left(n^{\frac{1}{4}} T^{\frac{1+5\gamma}{4}} \left(\rho_u\right)^T \sqrt{\log(n)}\right),$$

and

$$\sup_{1\leq i\leq n} T\left|\overline{y}_{i,-1}\overline{u}_i\right| = O_p\left(n^{\frac{1}{2}} \left(\rho_u\right)^T T^{\frac{5\gamma-1}{4}} \log(n)\right).$$

Once $T^{2\gamma-2}n\left(\log n\right)^2 \to 0$, we have

$$\sup_{1\le i\le n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| = O_p\left(n^{\frac{1}{4}}T^{\frac{3\gamma+1}{4}+\epsilon}\left(\rho_u\right)^T\sqrt{\log(n)}\right).$$

(ii) To justify the uniform lower bound for the denominator $\inf_{1\le i\le n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right|$, we employ the following decomposition as

$$\inf_{1\le i\le n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 = \underbrace{\inf_{1\le i\le n}\frac{T^\gamma}{2c_i^0}\mathbb{E}y_{i,T}^2}_{\text{(BB.1)}} - \underbrace{\sup_{1\le i\le n}\left(\frac{T^\gamma}{2c_i^0}\mathbb{E}y_{i,T}^2 - \sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right)}_{\text{(BB.2)}}.$$

For (BB.1) term, expectation operator removes randomness. We have

$$\inf_{1\le i\le n}\frac{T^\gamma}{2c_i^0}\mathbb{E}y_{i,T}^2 = O\left(\rho_l^{2T}T^{2\gamma}\right),$$

where $\rho_l := \exp\left(\frac{c_l}{T^\gamma}\right)$.

For (BB.2) term, the martingale exponential inequality is not directly applicable. Therefore, we employ the following decomposition,

$$\begin{aligned}
&\Pr\left(\sup_{1\le i\le n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 - \frac{T^\gamma}{2c_i^0}\mathbb{E}y_{i,T}^2\right| \ge \widetilde{M}_{nT}\right)\\
\le\ & n\sup_{1\le i\le n}\Pr\left(\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 - \frac{T^\gamma}{2c_i^0}\mathbb{E}y_{i,T}^2\right| \ge \widetilde{M}_{nT}\right)\\
\le\ & \underbrace{n\sup_{1\le i\le n}\Pr\left(\left|\frac{T^\gamma}{2c_i^0}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| \ge \widetilde{M}_{nT}\right)}_{\text{(BB.2.1)}} + \underbrace{n\sup_{1\le i\le n}\Pr\left(\left|\frac{T^\gamma}{2c_i^0}\sum_{t=1}^{T}y_{i,t-1}u_{it}\right| \ge \widetilde{M}_{nT}\right)}_{\text{(BB.2.2)}}\\
&+\underbrace{n\sup_{1\le i\le n}\Pr\left(\left|\frac{T^\gamma}{2c_i^0}\sum_{t=1}^{T}u_{it}^2\right| \ge \widetilde{M}_{nT}\right)}_{\text{(BB.2.3)}} + \underbrace{n\sup_{1\le i\le n}\Pr\left(\left|T\overline{y}_{i,-1}^2\right| \ge \widetilde{M}_{nT}\right)}_{\text{(BB.2.4)}},
\end{aligned}$$

where the innovations dominate the fixed effect uniformly. For (BB.2.1) term, we have

$$n\sup_{1\le i\le n}\Pr\left(\left|\frac{T^\gamma}{2c_i^0}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| \ge \widetilde{M}_{nT}\right) \le n\sup_{1\le i\le n}\Pr\left(\left|\frac{T^\gamma\rho_i^{2T}}{2c_i^0}\sum_{t=1}^{T}\rho_i^{-2t}\left(u_{it}^2 - \mathbb{E}u_{it}^2\right)\right| \ge \widetilde{M}_{nT}\right).$$

Note that $\left\{\rho_i^{-2t}\left(u_{it}^2-\mathbb{E}u_{it}^2\right)\right\}_{t=1}^T$ is a martingale difference sequence as $\mathbb{E}\left(\rho_i^{-2t}\left(u_{it}^2-\mathbb{E}u_{it}^2\right)|\mathcal{F}_{i,t-1}\right)=0$ with $\mathcal{F}_{i,t-1}:=\sigma\left\{u_{i,t-1},u_{i,t-2},...\right\}$. We set a truncation rate $\phi_{i,nT}=n^{\frac{1}{4}}T^{\frac{1}{4}}$. We define $x_{it}:=\rho_i^{-2t}\left(u_{it}^2-\mathbb{E}u_{it}^2\right)$ and make the following decomposition $\sum_{t=1}^T x_{it}=\sum_{t=1}^T x_{1it}+\sum_{t=1}^T x_{2it}-\sum_{t=1}^T\mathbb{E}\left[x_{2it}|\mathcal{F}_{i,t-1}\right]$. Define $x_{1it}:=x_{it}\mathbf{1}_{it}-\mathbb{E}\left[x_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\right]$ and $x_{2it}:=x_{it}\overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it}:=\mathbf{1}\left\{|x_{it}|\leq\phi_{i,nt}\right\}$ and $\overline{\mathbf{1}}_{it}=1-\mathbf{1}_{it}$. We need to derive the uniform upper bound from the following negligible conditions. It suffices to find $\widetilde{M}_{nT}$ to ensure that

$$\begin{aligned}
n\sup_{1\leq i\leq n}\Pr\left(\frac{T^\gamma\rho_i^{2T}}{2c_i^0}\left|\sum_{t=1}^T x_{1it}\right|\geq\widetilde{M}_{nT}\right) &= o(1),\\
n\sup_{1\leq i\leq n}\Pr\left(\frac{T^\gamma\rho_i^{2T}}{2c_i^0}\left|\sum_{t=1}^T x_{2it}\right|\geq\widetilde{M}_{nT}\right) &= o(1),\\
n\sup_{1\leq i\leq n}\Pr\left(\frac{T^\gamma\rho_i^{2T}}{2c_i^0}\left|\sum_{t=1}^T\mathbb{E}\left[x_{2it}|\mathcal{F}_{i,t-1}\right]\right|\geq\widetilde{M}_{nT}\right) &= o(1). \qquad (A.2.15)
\end{aligned}$$

The second and third arguments of (A.2.15) share identical derivations, and without losing generality we only focus on the second term. We define $\widetilde{V}_{iT}:=\sum_{t=1}^T\mathbb{E}\left[x_{1it}^2|\mathcal{F}_{i,t-1}\right]$, and $\widetilde{v}_{i,nT}=n^{\frac{1}{2}}T^{\frac{1}{2}+\frac{\gamma}{2}}$ is a truncation rate for $\widetilde{V}_{iT}$. With some constant $\widetilde{C}>0$, we have

$$\begin{aligned}
\mathbb{E}\left[\widetilde{V}_{iT}^2\right] &= \mathbb{E}\left[\sum_{t=1}^T\mathbb{E}\left[x_{1it}^2|\mathcal{F}_{i,t-1}\right]\right]^2\leq T\sum_{t=1}^T\mathbb{E}\left[x_{1it}^4\right]\leq 16T\sum_{t=1}^T\mathbb{E}\left[x_{it}^4\right]\\
&\leq \widetilde{C}T\sum_{t=1}^T\rho_i^{-8t}=O_p\left(T^{1+\gamma}\right).
\end{aligned}$$

By Proposition 2.1 in Freedman (1975), we have

$$\begin{aligned}
& n\sup_{1\leq i\leq n}\Pr\left(\frac{T^\gamma\rho_i^{2T}}{2c_i^0}\left|\sum_{t=1}^T x_{1it}\right|\geq\widetilde{M}_{nT}\right)\\
\leq\; & n\sup_{1\leq i\leq n}\Pr\left(\frac{T^\gamma\rho_i^{2T}}{2c_i^0}\left|\sum_{t=1}^T x_{1it}\right|\geq\widetilde{M}_{nT},\widetilde{V}_{iT}\leq\widetilde{v}_{i,nT}\right)+n\sup_{1\leq i\leq n}\Pr\left(\widetilde{V}_{iT}>\widetilde{v}_{i,nT}\right)\\
\leq\; & \sup_{1\leq i\leq n}\exp\left(\frac{-4(c_i^0)^2\widetilde{M}_{nT}^2/\left(T^{2\gamma}\rho_i^{4T}\right)+2\widetilde{v}_{i,nT}\log(n)+8c_i^0\widetilde{M}_{nT}\phi_{i,nt}\log(n)/\left(T^\gamma\rho_i^{2T}\right)}{2\widetilde{v}_{i,nT}+8c_i^0\widetilde{M}_{nT}\phi_{i,nt}/\left(T^\gamma\rho_i^{2T}\right)}\right)\\
& +\sup_{1\leq i\leq n}o\left(nT^{1+\gamma}\widetilde{v}_{i,nT}^{-2}\right)
\end{aligned}$$

$$= \quad o(1). \tag{A.2.16}$$

By our assumptions for $\widetilde{v}_{i,nT}$, we have $\sup_{1\leq i\leq n} o\left(nT^{1+\gamma}\widetilde{v}_{i,nT}^{-2}\right) = o(1)$. The asymptotic negligibility of the exponential term in (A.2.16) follows

$$\begin{aligned}
\widetilde{M}_{nT} &\succ \quad \sup_{1\leq i\leq n} T^{\gamma}\sqrt{\log(n)}T^{\frac{1+\gamma}{4}}\rho_i^{2T}n^{\frac{1}{4}}, \\
\widetilde{M}_{nT} &\succ \quad \sup_{1\leq i\leq n} T^{\gamma}\log(n)T^{\frac{1}{4}}\rho_i^{2T}n^{\frac{1}{4}}.
\end{aligned}$$

Since $\sup_{1\leq i\leq n} T^{\gamma}\sqrt{\log(n)}T^{\frac{1+\gamma}{4}}\rho_i^{2T}n^{\frac{1}{4}} \succ \sup_{1\leq i\leq n} T^{\gamma}\log(n)T^{\frac{1}{4}}\rho_i^{2T}n^{\frac{1}{4}}$, we only require $\widetilde{M}_{nT} \succ \sup_{1\leq i\leq n} T^{\gamma}\sqrt{\log(n)}T^{\frac{1+\gamma}{4}}\rho_i^{2T}n^{\frac{1}{4}}$. For the second term of (A.2.15), we show

$$\begin{aligned}
n\sup_{1\leq i\leq n}\Pr\left(\frac{T^{\gamma}\rho_i^{2T}}{2c_i}\left|\sum_{t=1}^{T}x_{2it}\right| \geq \widetilde{M}_{nT}\right) &\leq n\sup_{1\leq i\leq n}\Pr\left(\max_{1\leq t\leq T}|x_{it}| \geq \phi_{i,nt}\right) \\
&\leq \frac{nT}{\phi_{i,nT}^4}\sup_{1\leq i\leq n}\max_{1\leq t\leq T}\mathbb{E}\left[|x_{it}|^4 1\left\{|x_{it}| > \phi_{i,nT}\right\}\right] \\
&= \sup_{1\leq i\leq n} o\left(\frac{nT}{\phi_{i,nT}^4}\right),
\end{aligned}$$

where the assumption for $\phi_{i,nt}$ ensures that $\sup_{1\leq i\leq n} o\left(\frac{nT}{\phi_{i,nT}^4}\right) = o(1)$.

Therefore, for (BB.2.1) term, we have

$$\sup_{1\leq i\leq n}\left|\frac{T^{\gamma}}{2c_i^0}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| = O_p\left(\sup_{1\leq i\leq n}\sqrt{\log(n)}T^{\frac{1+5\gamma}{4}}\rho_i^{2T}n^{\frac{1}{4}}\right).$$

For (BB.2.2) term, we have, for any $\epsilon > 0$,

$$\sup_{1\leq i\leq n}\left|\frac{T^{\gamma}}{2c_i^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| = O_p\left(n^{\frac{1}{4}}T^{\frac{7\gamma+1}{4}+\epsilon}\left(\rho_u\right)^T\sqrt{\log(n)}\right),$$

based on our derivations for the numerator.

For (BB.2.3) term, we have

$$\sup_{1\leq i\leq n}\frac{T^{\gamma}}{2c_i^0}\sum_{t=1}^{T}u_{it}^2 = \sup_{1\leq i\leq n}\frac{T^{\gamma}}{2c_i^0}\sum_{t=1}^{T}\mathbb{E}u_{it}^2 - \sup_{1\leq i\leq n}\frac{T^{\gamma}}{2c_i^0}\sum_{t=1}^{T}\left(\mathbb{E}u_{it}^2 - u_{it}^2\right).$$

We easily show $\sup_{1\leq i\leq n}\frac{T^{\gamma}}{2c_i^0}\sum_{t=1}^{T}\mathbb{E}u_{it}^2 = O\left(T^{1+\gamma}\right)$. Similarly by the law of iterated logarithm, we show $\sup_{1\leq i\leq n}\frac{T^{\gamma}}{2c_i^0}\sum_{t=1}^{T}\left(\mathbb{E}u_{it}^2 - u_{it}^2\right) = o_p\left(T^{1+\gamma}\right)$. Therefore we show

$$\sup_{1\leq i\leq n}\frac{T^{\gamma}}{2c_i^0}\sum_{t=1}^{T}u_{it}^2 = O_p\left(T^{1+\gamma}\right).$$

For (BB.2.4) term, we have the following decomposition as

$$\begin{aligned}
& n\sup_{1\leq i\leq n}\Pr\left(\left|T\overline{y}_{i,-1}^2\right| \geq \overline{M}_{nT}\right) \\
\leq\ & n\sup_{1\leq i\leq n}\Pr\left(\left|T^{\frac{1}{2}}\overline{y}_{i,-1}\right| \geq \sqrt{\overline{M}_{nT}}\right) \\
=\ & n\sup_{1\leq i\leq n}\Pr\left(\left|T^{\frac{-1}{2}}\sum_{t=1}^{T}\left(\sum_{s=0}^{t-1}\rho_i^{t-1-s}u_{is}\right)\right| \geq \sqrt{\overline{M}_{nT}}\right) \\
=\ & n\sup_{1\leq i\leq n}\Pr\left(\left|T^{\frac{-1}{2}}\sum_{s=0}^{T-1}\left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}\right| \geq \sqrt{\overline{M}_{nT}}\right) \\
=\ & n\sup_{1\leq i\leq n}\Pr\left(\left|T^{\frac{-1}{2}}\sum_{s=0}^{T-1}\frac{\rho_i^{T-s}-1}{\rho_i-1}u_{is}\right| \geq \sqrt{\overline{M}_{nT}}\right) \\
=\ & n\sup_{1\leq i\leq n}\Pr\left(\left|\frac{T^{\frac{2\gamma-1}{2}}}{\left(c_i^0\right)}\sum_{s=0}^{T-1}\left(\rho_i^{T-s}-1\right)u_{is}\right| \geq \sqrt{\overline{M}_{nT}}\right) \\
\leq\ & n\sup_{1\leq i\leq n}\Pr\left(\left|\frac{T^{\frac{2\gamma-1}{2}}\rho_i^T}{\left(c_i^0\right)}\sum_{s=0}^{T-1}\left(\rho_i^{-s}\right)u_{is}\right| \geq \sqrt{\overline{M}_{nT}}\right) \\
& +n\sup_{1\leq i\leq n}\Pr\left(\left|\frac{T^{\frac{2\gamma-1}{2}}}{\left(c_i^0\right)}\sum_{s=0}^{T-1}u_{is}\right| \geq \sqrt{\overline{M}_{nT}}\right),
\end{aligned}$$

where innovations dominate the fixed effect for each individual. Note that $\left\{\rho_i^{-s}u_{is}\right\}_{s=1}^{T}$ is a martingale difference sequence as $\mathbb{E}\left(\rho_i^{-s}u_{is}|\mathcal{F}_{i,s-1}\right) = 0$ with $\mathcal{F}_{i,s-1} := \sigma\left\{u_{i,s-1}, u_{i,s-2}, ...\right\}$. We set a truncation rate $o_{i,nT} = n^{\frac{1}{4}}T^{\frac{1}{4}}$. We define $\widetilde{x}_{it} := \rho_i^{-s}u_{is}$ and make the following decomposition $\sum\limits_{t=1}^{T}\widetilde{x}_{it} = \sum\limits_{t=1}^{T}\widetilde{x}_{1it} + \sum\limits_{t=1}^{T}\widetilde{x}_{2it} - \sum\limits_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{2it}|\mathcal{F}_{i,t-1}\right]$. Define $\widetilde{x}_{1it} := \widetilde{x}_{it}\mathbf{1}_{it} - \mathbb{E}\left[\widetilde{x}_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\right]$ and $\widetilde{x}_{2it} := \widetilde{x}_{it}\overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it} := \mathbf{1}\left\{\left|\widetilde{x}_{it}\right| \leq o_{i,nt}\right\}$ and $\overline{\mathbf{1}}_{it} = 1 - \mathbf{1}_{it}$. We need to derive the uniform upper bound from the following negligible conditions. It suffices to show to find $\overline{M}_{nT}$ to ensure that

$$n\sup_{1\leq i\leq n}\Pr\left(\frac{T^{\frac{2\gamma-1}{2}}\rho_i^T}{\left(c_i^0\right)}\left|\sum_{t=1}^{T}\widetilde{x}_{1it}\right| \geq \sqrt{\overline{M}_{nT}}\right) = o\left(1\right),$$

$$
\begin{aligned}
n \sup_{1\le i\le n} \Pr\left(\frac{T^{\frac{2\gamma-1}{2}}\rho_i^T}{(c_i^0)}\left|\sum_{t=1}^{T}\widetilde{x}_{2it}\right| \ge \sqrt{\overline{M}_{nT}}\right) &= o(1)\,,\\
n \sup_{1\le i\le n} \Pr\left(\frac{T^{\frac{2\gamma-1}{2}}\rho_i^T}{(c_i^0)}\left|\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{2it}|\mathcal{F}_{i,t-1}\right]\right| \ge \sqrt{\overline{M}_{nT}}\right) &= o(1)\,. \quad \text{(A.2.17)}
\end{aligned}
$$

The second and third arguments of (A.2.17) share identical derivations, and without losing generality we only focus on the second term. We define $\overline{V}_{iT} := \sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{1it}^2|\mathcal{F}_{i,t-1}\right]$, and $\overline{v}_{i,nT} = n^{\frac{1}{2}}T^{\frac{1}{2}+\frac{\gamma}{2}}$ is a truncation rate for $\overline{V}_{iT}$. With some constant $\overline{C} > 0$, we have

$$
\begin{aligned}
\mathbb{E}\left[\overline{V}_{iT}^2\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{1it}^2|\mathcal{F}_{i,t-1}\right]\right]^2 \le T\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{1it}^4\right] \le 16T\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{it}^4\right]\\
&\le \overline{C}T\sum_{t=1}^{T}\rho_i^{-4t} = O\left(T^{1+\gamma}\right).
\end{aligned}
$$

By Proposition 2.1 in Freedman (1975),

$$
\begin{aligned}
& n \sup_{1\le i\le n} \Pr\left(\frac{T^{\frac{2\gamma-1}{2}}\rho_i^T}{(c_i^0)}\left|\sum_{t=1}^{T}\widetilde{x}_{1it}\right| \ge \sqrt{\overline{M}_{nT}}\right)\\
\le\ & n \sup_{1\le i\le n} \Pr\left(\frac{T^{\frac{2\gamma-1}{2}}\rho_i^T}{(c_i^0)}\left|\sum_{t=1}^{T}\widetilde{x}_{1it}\right| \ge \sqrt{\overline{M}_{nT}}, \overline{V}_{iT} \le \overline{v}_{i,nT}\right) + n\sup_{1\le i\le n}\Pr\left(\overline{V}_{iT} > \overline{v}_{i,nT}\right)\\
\le\ & \sup_{1\le i\le n}\exp\left(\frac{-\left(c_i^0\right)^2\overline{M}_{nT}T^{1-2\gamma}/\rho_i^{2T} + 2\overline{v}_{i,nT}\log(n) + 4\sqrt{(c_i^0)^2}\sqrt{\overline{M}_{nT}}o_{i,nt}\log(n)\,T^{\frac{1-2\gamma}{2}}\rho_i^{-T}}{2\overline{v}_{i,nT} + 4\sqrt{(c_i^0)^2}\sqrt{\overline{M}_{nT}}o_{i,nt}T^{\frac{1-2\gamma}{2}}\rho_i^{-T}}\right)\\
& + \sup_{1\le i\le n} o\left(nT^{1+\gamma}\overline{v}_{i,nT}^{-2}\right)\\
=\ & o(1). \qquad \text{(A.2.18)}
\end{aligned}
$$

By our assumptions for $\widetilde{v}_{i,nT}$, we have $\sup_{1\le i\le n} o\left(nT^{1+\gamma}\overline{v}_{i,nT}^{-2}\right) = o(1)$. The asymptotic negligibility of the exponential term (A.2.18) is ensured by the following facts,

$$
\begin{aligned}
\overline{M}_{nT} &\succ \sup_{1\le i\le n} T^{\frac{4\gamma-1}{2}}\left(\log(n)\right)^2\rho_i^{2T}n^{\frac{1}{2}},\\
\overline{M}_{nT} &\succ \sup_{1\le i\le n} T^{\frac{5\gamma-1}{2}}\log(n)\rho_i^{2T}n^{\frac{1}{2}}.
\end{aligned}
$$

Since $\sup_{1\le i\le n} T^{\frac{5\gamma-1}{2}}\log(n)\rho_i^{2T}n^{\frac{1}{2}} \succ \sup_{1\le i\le n} T^{\frac{4\gamma-1}{2}}\left(\log(n)\right)^2\rho_i^{2T}n^{\frac{1}{2}}$, we only require $\overline{M}_{nT} \succ \sup_{1\le i\le n} T^{\frac{5\gamma-1}{2}}\log(n)\rho_i^{2T}n^{\frac{1}{2}}$. For the second term of (A.2.17), we show

$$
\begin{aligned}
n \sup_{1\leq i\leq n} \Pr\left(\frac{T^{\frac{2\gamma-1}{2}}\rho_i^T}{c_i^0}\left|\sum_{t=1}^{T}\widetilde{x}_{2it}\right| \geq \sqrt{\overline{M}_{nT}}\right) &\leq n \sup_{1\leq i\leq n} \Pr\left(\max_{1\leq t\leq T}|\widetilde{x}_{it}| \geq o_{i,nt}\right) \\
&\leq \frac{nT}{o_{i,nT}^4}\sup_{1\leq i\leq n}\max_{1\leq t\leq T}\mathbb{E}\left[|\widetilde{x}_{it}|^4 1\left\{|\widetilde{x}_{it}| > o_{i,nT}\right\}\right] \\
&= \sup_{1\leq i\leq n} o\left(\frac{nT}{o_{i,nT}^4}\right).
\end{aligned}
$$

The assumption for $o_{i,nt}$ ensures that $\sup_{1\leq i\leq n} o\left(\frac{nT}{o_{i,nT}^4}\right) = o(1)$.

Combining results of (BB.1), (BB.2.1), (BB.2.2), (BB.2.3), and (BB.2.4), if $T^{5\gamma-3}n\left(\log n\right)^2 \to 0$, we have

$$
\inf_{1\leq i\leq n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 = O_p\left(n^{\frac{1}{4}}\sqrt{\log n}T^{\frac{1+5\gamma}{4}+\epsilon}\left(\rho_u\right)^{2T}\right). \tag{A.2.19}
$$

for any $\epsilon > 0$.

All in all, based on (A.2.19) and (A.2.15), we have

$$
\sup_{\substack{1\leq i\leq n\\ c_i^0>0}}\left|\widehat{c}_i - c_i^0\right| = O_p\left(\left(\rho_u\right)^{-T}T^{\frac{-\gamma}{2}}\right).
$$

This concludes our proof. ■

Lemma A.2.9 *If Assumptions 1 and 2 hold,*

$$
\sup_{1\leq i\leq n, c_i^0<0}\left|\widehat{c}_i - c_i^0\right| = O_p\left(\frac{n^{\frac{1}{4}+\epsilon}\sqrt{\log(n)}}{T^{\frac{1+\gamma}{2}}}\right),
$$

with arbitrary $\epsilon > 0$.

Proof. Similarly, we demonstrate the uniform upper bounds and uniform lower bound of several terms as

$$
\underbrace{\inf_{1\leq i\leq n}\sum_{t=1}^{T}\mathbb{E}\widetilde{y}_{i,t-1}^2}_{\text{(CC.1)}},\ \underbrace{\sup_{1\leq i\leq n}\sum_{t=1}^{T}\left|\mathbb{E}\widetilde{y}_{i,t-1}^2 - \widetilde{y}_{i,t-1}^2\right|}_{\text{(CC.2)}},\ \underbrace{\sup_{1\leq i\leq n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|}_{\text{(CC.3)}}.
$$

For the term of (CC.1), we have

$$\inf_{1\le i\le n}\sum_{t=1}^{T}\mathbb{E}\widetilde{y}_{i,t-1}^{2}=O\left(T^{1+\gamma}\right),$$

since the expectation operator removes randomness. For (CC.2), by applying the exponential inequality (Freedman, 1975) and Markov inequality (White, 2014), we can readily show

$$\sup_{1\le i\le n}\sum_{t=1}^{T}\left|\mathbb{E}y_{i,t-1}^{2}-\widetilde{y}_{i,t-1}^{2}\right|=o_p\left(T^{1+\gamma}\right).$$

Therefore we show

$$\begin{aligned}\inf_{1\le i\le n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2} &= \inf_{1\le i\le n}\sum_{t=1}^{T}\mathbb{E}\widetilde{y}_{i,t-1}^{2}-\sup_{1\le i\le n}\sum_{t=1}^{T}\left(\mathbb{E}\widetilde{y}_{i,t-1}^{2}-\widetilde{y}_{i,t-1}^{2}\right)\\ &= O_p\left(T^{1+\gamma}\right).\end{aligned}$$

For the term of (CC.3), we accommodates the following decomposition as,

$$\begin{aligned}&\Pr\left(\sup_{1\le i\le n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\ge \check{M}_{nT}\right)\\ \le\ &\Pr\left(\sup_{1\le i\le n}\left|\sum_{t=1}^{T}y_{i,t-1}u_{it}\right|\ge \check{M}_{nT}\right)+\Pr\left(\sup_{1\le i\le n}\left|T\overline{y}_{i,-1}\overline{u}_{i}\right|\ge \check{M}_{nT}\right)\\ \le\ &\underbrace{n\sup_{1\le i\le n}\Pr\left(\left|\sum_{t=1}^{T}y_{i,t-1}u_{it}\right|\ge \check{M}_{nT}\right)}_{\text{(CC.3.1)}}+\underbrace{n\sup_{1\le i\le n}\Pr\left(\left|T\overline{y}_{i,-1}\overline{u}_{i}\right|\ge \check{M}_{nT}\right)}_{\text{(CC.3.2)}},\end{aligned}$$

where the innovations dominate the fixed effect uniformly. For (CC.3.1) we apply the exponential inequality of Freedman (1975). The process $\{y_{i,t-1}u_{it}\}_{t=1}^{T}$ is a martingale difference sequence as $\mathbb{E}\left(y_{i,t-1}u_{it}|\mathcal{F}_{i,t-1}\right)=0$ with $\mathcal{F}_{i,t-1}:=\sigma\left\{u_{i,t-1},u_{i,t-2},...\right\}$. We set a truncation rate $\check{d}_{i,nT}=n^{\frac{1}{4}}T^{\frac{1+2\gamma}{4}}$. We define $\check{z}_{it}:=y_{i,t-1}u_{it}$ and make the following decomposition $\sum_{t=1}^{T}\check{z}_{it}=\sum_{t=1}^{T}\check{z}_{1it}+\sum_{t=1}^{T}\check{z}_{2it}-\sum_{t=1}^{T}\mathbb{E}\left[\check{z}_{2it}|\mathcal{F}_{i,t-1}\right]$. Define $\check{z}_{1it}:=\check{z}_{it}\mathbf{1}_{it}-\mathbb{E}\left[\check{z}_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\right]$ and $\check{z}_{2it}:=\check{z}_{it}\overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it}:=\mathbf{1}\left\{|\check{z}_{it}|\le\check{d}_{i,nT}\right\}$ and $\overline{\mathbf{1}}_{it}=1-\mathbf{1}_{it}$. We need to derive the uniform upper bound from the following negligible conditions. It suffices to show find $\check{M}_{nT}$ to ensure that

$$\begin{aligned}
n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \check{z}_{1it}\right| \ge \check{M}_{nT}\right) &= o(1), \\
n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \check{z}_{2it}\right| \ge \check{M}_{nT}\right) &= o(1), \\
n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \mathbb{E}\left[\check{z}_{2it}|\mathcal{F}_{i,t-1}\right]\right| \ge \check{M}_{nT}\right) &= o(1). \qquad \text{(A.2.20)}
\end{aligned}$$

The second and third arguments of (A.2.20) share identical derivations, and without losing generality we only focus on the second term. We define $\check{V}_{iT} := \sum_{t=1}^{T} \mathbb{E}\left[\check{z}_{1it}^2|\mathcal{F}_{i,t-1}\right]$, and $\check{v}_{i,nT} = n^{\frac{1}{2}} T^{\frac{2+2\gamma}{2}}$ is a truncation rate for $\check{V}_{iT}$.

$$\begin{aligned}
\mathbb{E}\left[\check{V}_{iT}^2\right] &= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[\check{z}_{1it}^2|\mathcal{F}_{i,t-1}\right]\right]^2 \le T\sum_{t=1}^{T} \mathbb{E}\left[\check{z}_{1it}^4\right] \le 16T\sum_{t=1}^{T} \mathbb{E}\left[\check{z}_{it}^4\right] \\
&\le CT^{2\gamma+2} = O\left(T^{2\gamma+2}\right).
\end{aligned}$$

By Proposition 2.1 in Freedman (1975), we have

$$\begin{aligned}
& n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \check{z}_{1it}\right| \ge \check{M}_{nT}\right) \\
\le\ & n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \check{z}_{1it}\right| \ge \check{M}_{nT}, \check{V}_{iT} \le \check{v}_{i,nT}\right) + n \sup_{1\le i\le n} \Pr\left(\check{V}_{iT} > \check{v}_{i,nT}\right) \\
\le\ & \sup_{1\le i\le n} \exp\left(\frac{-\check{M}_{nT}^2 + 2\check{v}_{i,nT}\log(n) + 4\check{M}_{nT}\check{d}_{i,nT}\log(n)}{2\check{v}_{i,nT} + 4\check{M}_{nT}\check{d}_{i,nT}}\right) + \sup_{1\le i\le n} o\left(nT^{2+2\gamma}\check{v}_{i,nT}^{-2}\right) \\
=\ & o(1). \qquad \text{(A.2.21)}
\end{aligned}$$

To show asymptotic negligibility of (A.2.21), we need $\check{M}_{nT} \succ n^{\frac{1}{4}} T^{\frac{2+2\gamma}{4}} \sqrt{\log(n)}$ and $\check{M}_{nT} \succ n^{\frac{1}{4}} T^{\frac{1+2\gamma}{4}} \log(n)$. Our assumption for $\check{v}_{i,nT}$ ensures $\sup_{1\le i\le n} o\left(nT^{2+2\gamma}\check{v}_{i,nT}^{-2}\right) = o(1)$. For the second term of (A.2.20), we have

$$\begin{aligned}
& n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \check{z}_{2it}\right| \ge \check{M}_{nT}\right) \\
\le\ & n \sup_{1\le i\le n} \Pr\left(\max_{1\le t\le T} |\check{z}_{it}| \ge \check{d}_{i,nT}\right) \le \frac{nT}{\check{d}_{i,nT}^4} \sup_{1\le i\le n} \max_{1\le t\le T} \mathbb{E}\left[|z_{it}|^4 1\left\{|z_{it}| > \check{d}_{i,nT}\right\}\right]
\end{aligned}$$

$$= \sup_{1\leq i\leq n}\max_{1\leq t\leq T} o\left(\frac{nT^{1+2\gamma}}{\breve{d}_{i,nT}^{4}}\right) = o(1),$$

which is guaranteed by our assumption for $d_{i,nT}$. Since $n^{\frac{1}{4}}T^{\frac{2+2\gamma}{4}}\sqrt{\log(n)} \succ n^{\frac{1}{4}}T^{\frac{1+2\gamma}{4}}\log(n)$, we need $\breve{M}_{nT} \succ n^{\frac{1}{4}}T^{\frac{2+2\gamma}{4}}\sqrt{\log(n)}$. Therefore we have, for any $\epsilon > 0$,

$$\sup_{1\leq i\leq n}\left|\sum_{t=1}^{T} y_{i,t-1}u_{it}\right| = O_p\left(n^{\frac{1}{4}}T^{\frac{2+2\gamma}{4}+\epsilon}\sqrt{\log(n)}\right).$$

For (CC.3.2) term, we have

$$\sup_{1\leq i\leq n}\left|T\overline{y}_{i,-1}\overline{u}_i\right| \leq T\sup_{1\leq i\leq n}\left|\overline{y}_{i,-1}\right|\sup_{1\leq i\leq n}\left|\overline{u}_i\right|.$$

By applying exponential inequality, we have $\sup_{1\leq i\leq n}|T\overline{u}_i| = O_p\left(\sqrt{\log n}n^{\frac{1}{4}}T^{\frac{1}{2}}\right)$. Note that for $\left\{\left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}\right\}_{s=1}^{T}$,

$$\left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is} = \frac{1-\rho_i^{T-s}}{1-\rho_i}u_{is} = \frac{T^{\gamma}}{-c_i^0}\left(1-\rho_i^{T-s}\right)u_{is}.$$

Therefore,

$$\begin{aligned}
\sup_{1\leq i\leq n}\left|\sum_{t=1}^{T} y_{i,t-1}\right| &= \sup_{1\leq i\leq n}\left|\sum_{s=1}^{T}\left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}\right| = \sup_{1\leq i\leq n}\frac{T^{\gamma}}{-c_i^0}\left|\sum_{s=1}^{T}\left(1-\rho_i^{T-s}\right)u_{is}\right| \\
&\leq \sup_{1\leq i\leq n}\frac{T^{\gamma}}{-c_i^0}\left|\sum_{s=1}^{T}u_{is}\right| + \sup_{1\leq i\leq n}\frac{\rho_i^T T^{\gamma}}{-c_i^0}\left|\sum_{s=1}^{T}\rho_i^{-s}u_{is}\right|
\end{aligned}$$

It is obvious that

$$\sup_{1\leq i\leq n}\frac{T^{\gamma}}{-c_i^0}\left|\sum_{s=1}^{T}u_{is}\right| = O_p\left(n^{\frac{1}{4}}T^{\frac{1}{2}+\gamma}\sqrt{\log(n)}\right).$$

Note the fact that $\left\{\rho_i^{-s}u_{is}\right\}_{s=1}^{T}$ is a martingale difference sequence with $\mathcal{F}_{i,s-1} := \sigma\left\{u_{i,s-1}, u_{i,s-2}, ...\right\}$. We set a truncation rate $\breve{o}_{i,nT} = n^{\frac{1}{4}}T^{\frac{1}{4}}\rho_i^{-T}$. We define $\breve{x}_{it} := \rho_i^{-t}u_{it}$ and make the following decomposition $\sum_{t=1}^{T}\breve{x}_{it} = \sum_{t=1}^{T}\breve{x}_{1it} + \sum_{t=1}^{T}\breve{x}_{2it} - \sum_{t=1}^{T}\mathbb{E}\left[\breve{x}_{2it}|\mathcal{F}_{i,t-1}\right]$. Define $\breve{x}_{1it} := \breve{x}_{it}\mathbf{1}_{it} - \mathbb{E}\left[\breve{x}_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\right]$ and $\breve{x}_{2it} := \breve{x}_{it}\overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it} := \mathbf{1}\left\{|\breve{x}_{it}| \leq \breve{o}_{i,nt}\right\}$

and $\overline{\mathbf{1}}_{it} = 1 - \mathbf{1}_{it}$. We need to derive the uniform upper bound from the following negligible conditions. It suffices to show find $\check{f}_{nT}$ to ensure that

$$\begin{aligned}
n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} \check{x}_{1it}\right| \geq \frac{-c_i^0 \check{f}_{nT}}{T^\gamma}\right) &= o(1), \\
n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} \check{x}_{2it}\right| \geq \frac{-c_i^0 \check{f}_{nT}}{T^\gamma}\right) &= o(1), \\
n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} \mathbb{E}\left[\check{x}_{2it}|\mathcal{F}_{i,t-1}\right]\right| \geq \frac{-c_i^0 \check{f}_{nT}}{T^\gamma}\right) &= o(1). \qquad \text{(A.2.22)}
\end{aligned}$$

The second and third arguments of (A.2.22) share identical derivations, and without losing generality we only focus on the second term. We define $\check{W}_{iT} := \sum_{t=1}^{T} \mathbb{E}\left[\check{x}_{1it}^2|\mathcal{F}_{i,t-1}\right]$, and $\check{w}_{i,nT} = n^{\frac{1}{2}} T^{\frac{1}{2}+\frac{\gamma}{2}} \rho_i^{-2T}$ is a truncation rate for $\check{W}_{iT}$. With some constant $\check{C} > 0$, we have

$$\begin{aligned}
\mathbb{E}\left[\check{W}_{iT}^2\right] &= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[\check{x}_{1it}^2|\mathcal{F}_{i,t-1}\right]\right]^2 \leq T \sum_{t=1}^{T} \mathbb{E}\left[\check{x}_{1it}^4\right] \leq 16T \sum_{t=1}^{T} \mathbb{E}\left[\check{x}_{it}^4\right] \\
&\leq \check{C} T \sum_{t=1}^{T} \rho_i^{-4t} = O_p\left(T^{1+\gamma} \rho_i^{-4T}\right).
\end{aligned}$$

By Proposition 2.1 in Freedman (1975),

$$\begin{aligned}
& n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} \check{x}_{1it}\right| \geq \frac{-c_i^0 \check{f}_{nT}}{T^\gamma}\right) \\
\leq\ & n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} \check{x}_{1it}\right| \geq \frac{-c_i^0 \check{f}_{nT}}{T^\gamma}, \check{W}_{iT} \leq \check{w}_{i,nT}\right) + n \sup_{1\leq i\leq n} \Pr\left(\check{W}_{iT} > \check{w}_{i,nT}\right) \\
\leq\ & \sup_{1\leq i\leq n} \exp\left(\frac{\left(c_i^0\right)^2 \check{f}_{nT}^2/T^{2\gamma} + 2\check{w}_{i,nT} \log(n) + (-c_i^0) 4 \check{f}_{nT} \check{o}_{i,nt} \log(n)/T^\gamma}{2\check{w}_{i,nT} + (-c_i^0) 4 \check{f}_{nT} \check{o}_{i,nt}/T^\gamma}\right) \\
& + \sup_{1\leq i\leq n} o\left(nT^{1+\gamma} \rho_i^{-4T} \check{w}_{i,nT}^{-2}\right) \\
=\ & o(1). \qquad \text{(A.2.23)}
\end{aligned}$$

To show asymptotic negligibility of (A.2.23), we need $\check{f}_{nT} \succ n^{\frac{1}{4}} T^{\frac{5\gamma+1}{4}} \rho_i^{-T} \sqrt{\log(n)}$ and $\check{f}_{nT} \succ n^{\frac{1}{4}} T^{\frac{4\gamma+1}{4}} \rho_i^{-T} \log(n)$. Our assumption for $\check{w}_{i,nT}$ ensures $\sup_{1\leq i\leq n} o\left(nT^{1+\gamma} \check{w}_{i,nT}^{-2}\right) =$

$o(1)$. For the second term of (A.2.22), we have

$$
\begin{aligned}
& n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \check{z}_{2it}\right| \ge \frac{(-c_i^0)\check{f}_{nT}}{T^{\gamma}}\right) \\
\le \; & n \sup_{1\le i\le n} \Pr\left(\max_{1\le t\le T} |\check{x}_{it}| \ge \check{o}_{i,nt}\right) \le \frac{nT}{\check{o}_{i,nT}^4} \sup_{1\le i\le n} \max_{1\le t\le T} \mathbb{E}\left[|\check{x}_{it}|^4 1\left\{|\check{x}_{it}| > \check{o}_{i,nt}\right\}\right] \\
= \; & \sup_{1\le i\le n} \max_{1\le t\le T} o\left(\frac{nT\rho_i^{-4T}}{\check{o}_{i,nT}^4}\right) = o\left(1\right),
\end{aligned}
$$

which is guaranteed by our assumption for $\check{o}_{i,nt}$. Since $n^{\frac{1}{4}}T^{\frac{5\gamma+1}{4}}\rho_i^{-T}\sqrt{\log\left(n\right)} \succ n^{\frac{1}{4}}T^{\frac{4\gamma+1}{4}}\rho_i^{-T}\log\left(n\right)$, we need $\check{f}_{nT} \succ n^{\frac{1}{4}}T^{\frac{5\gamma+1}{4}}\rho_i^{-T}\sqrt{\log\left(n\right)}$. The abovementioned derivation shows

$$
\sup_{1\le i\le n} \left|\overline{y}_{i,-1}\right| = O_p\left(n^{\frac{1}{4}}T^{-\frac{1}{2}+\gamma}\log n\right).
$$

For any $\epsilon > 0$, this shows

$$
\sup_{1\le i\le n} \left|T\overline{y}_{i,-1}\overline{u}_i\right| \le T \sup_{1\le i\le n} \left|\overline{y}_{i,-1}\right| \sup_{1\le i\le n} \left|\overline{u}_i\right| = O_p\left(\left(\log n\right) n^{\frac{1}{2}}T^{\gamma+\epsilon}\right).
$$

Therefore, if $T^{2-2\gamma} \succ n\left(\log(n)\right)^2$,

$$
\sup_{1\le i\le n} \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| = O_p\left(n^{\frac{1}{4}}T^{\frac{1+\gamma}{2}+\epsilon}\sqrt{\log\left(n\right)}\right),
$$

and

$$
\sup_{1\le i\le n, c_i^0<0} \left|\widehat{c}_i - c_i^0\right| = O_p\left(\frac{n^{\frac{1}{4}+\epsilon}\sqrt{\log(n)}}{T^{\frac{1+\gamma}{2}}}\right),
$$

with arbitrary $\epsilon > 0$. ■

Lemma A.2.10 *If Assumptions 1 and 2 hold,*

$$
\sup_{1\le i\le n, c_i^0=0} \left|\widehat{c}_i - c_i^0\right| = O_p\left(\frac{n^{\frac{1}{2}+\epsilon}\sqrt{\log(n)}\log_2\left(T\right)}{T}\right),
$$

with arbitrary $\epsilon > 0$. *The notation* $\log_2(\cdot) := \log\log(\cdot)$.

Proof. Similarly, we demonstrate the uniform upper bounds and uniform lower bound of several terms as

$$\underbrace{\inf_{1\leq i\leq n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}}_{\text{(DD.1)}},\ \underbrace{\sup_{1\leq i\leq n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|}_{\text{(DD.2)}}.$$

For the term of (DD.1), by Lemma A.2 of Huang et al. (2019)

$$\inf_{1\leq i\leq n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}=O_p\left(\frac{T^2}{\log_2(T)}\right),$$

due to the law of the iterated logarithm. For the term of (DD.2), we accommodates the following decomposition as,

$$\begin{aligned}
&\Pr\left(\sup_{1\leq i\leq n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\geq \dot{M}_{nT}\right)\\
\leq\ &\Pr\left(\sup_{1\leq i\leq n}\left|\sum_{t=1}^{T}y_{i,t-1}u_{it}\right|\geq \dot{M}_{nT}\right)+\Pr\left(\sup_{1\leq i\leq n}\left|T\overline{y}_{i,-1}\overline{u}_{i}\right|\geq \dot{M}_{nT}\right)\\
\leq\ &\underbrace{n\sup_{1\leq i\leq n}\Pr\left(\left|\sum_{t=1}^{T}y_{i,t-1}u_{it}\right|\geq \dot{M}_{nT}\right)}_{\text{(DD.2.1)}}+\underbrace{n\sup_{1\leq i\leq n}\Pr\left(\left|T\overline{y}_{i,-1}\overline{u}_{i}\right|\geq \dot{M}_{nT}\right)}_{\text{(DD.2.2)}},
\end{aligned}$$

where the innovations dominate the fixed effect uniformly. For (DD.2.1) we apply the exponential inequality of Freedman (1975). The process $\{y_{i,t-1}u_{it}\}_{t=1}^{T}$ is a martingale difference sequence as $\mathbb{E}\left(y_{i,t-1}u_{it}|\mathcal{F}_{i,t-1}\right)=0$ with $\mathcal{F}_{i,t-1}:=\sigma\left\{u_{i,t-1},u_{i,t-2},...\right\}$. We set a truncation rate $\dot{d}_{i,nT}=n^{\frac{1}{4}}T^{\frac{3}{4}}$. We define $\dot{z}_{it}:=y_{i,t-1}u_{it}$ and make the following decomposition $\sum_{t=1}^{T}\dot{z}_{it}=\sum_{t=1}^{T}\dot{z}_{1it}+\sum_{t=1}^{T}\dot{z}_{2it}-\sum_{t=1}^{T}\mathbb{E}\left[\dot{z}_{2it}|\mathcal{F}_{i,t-1}\right]$. Define $\dot{z}_{1it}:=\dot{z}_{it}\mathbf{1}_{it}-\mathbb{E}\left[\dot{z}_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\right]$ and $\dot{z}_{2it}:=\dot{z}_{it}\overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it}:=\mathbf{1}\left\{|\dot{z}_{it}|\leq\dot{d}_{i,nT}\right\}$ and $\overline{\mathbf{1}}_{it}=1-\mathbf{1}_{it}$. We need to derive the uniform upper bound from the following negligible conditions. It suffices to find $\dot{M}_{nT}$ to ensure that

$$n\sup_{1\leq i\leq n}\Pr\left(\left|\sum_{t=1}^{T}\dot{z}_{1it}\right|\geq \dot{M}_{nT}\right)\ =\ o(1),$$

$$
\begin{aligned}
n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \dot{z}_{2it}\right| \ge \dot{M}_{nT}\right) &= o\left(1\right), \\
n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \mathbb{E}\left[\dot{z}_{2it}|\mathcal{F}_{i,t-1}\right]\right| \ge \dot{M}_{nT}\right) &= o\left(1\right). \qquad \text{(A.2.24)}
\end{aligned}
$$

The second and third arguments of (A.2.24) share identical derivations, and without losing generality we only focus on the second term. We define $\dot{V}_{iT} := \sum_{t=1}^{T} \mathbb{E}\left[\dot{z}_{1it}^{2}|\mathcal{F}_{i,t-1}\right]$, and $\dot{v}_{i,nT} = n^{\frac{1}{2}}T^{2}$ is a truncation rate for $\dot{V}_{iT}$.

$$
\begin{aligned}
\mathbb{E}\left[\dot{V}_{iT}^{2}\right] &= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[\dot{z}_{1it}^{2}|\mathcal{F}_{i,t-1}\right]\right]^{2} \le T\sum_{t=1}^{T}\mathbb{E}\left[\dot{z}_{1it}^{4}\right] \le 16T\sum_{t=1}^{T}\mathbb{E}\left[\dot{z}_{it}^{4}\right] \\
&\le CT^{4} = O_{p}\left(T^{4}\right),
\end{aligned}
$$

where $C > 0$ is some constant value. By Proposition 2.1 in Freedman (1975), we have

$$
\begin{aligned}
& n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \dot{z}_{1it}\right| \ge \dot{M}_{nT}\right) \\
\le\ & n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \dot{z}_{1it}\right| \ge \dot{M}_{nT}, \dot{V}_{iT} \le \dot{v}_{i,nT}\right) + n \sup_{1\le i\le n} \Pr\left(\dot{V}_{iT} > \dot{v}_{i,nT}\right) \\
\le\ & \sup_{1\le i\le n} \exp\left(\frac{-\dot{M}_{nT}^{2} + 2\dot{v}_{i,nT}\log\left(n\right) + 4\dot{M}_{nT}\dot{d}_{i,nT}\log\left(n\right)}{2\dot{v}_{i,nT} + 4\dot{M}_{nT}\dot{d}_{i,nT}}\right) + \sup_{1\le i\le n} o\left(nT^{4}\dot{v}_{i,nT}^{-2}\right) \\
=\ & o(1). \qquad \text{(A.2.25)}
\end{aligned}
$$

To show asymptotic negligibility of (A.2.25), we need $\dot{M}_{nT} \succ n^{\frac{1}{4}}T\sqrt{\log\left(n\right)}$ and $\dot{M}_{nT} \succ n^{\frac{1}{4}}T^{\frac{3}{4}}\log\left(n\right)$. Our assumption for $v_{i,nT}$ ensures $\sup_{1\le i\le n} o\left(nT^{4}\dot{v}_{i,nT}^{-2}\right) = o(1)$. For the second term of (A.2.24), we have

$$
\begin{aligned}
& n \sup_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^{T} \dot{z}_{2it}\right| \ge \dot{M}_{nT}\right) \\
\le\ & n \sup_{1\le i\le n} \Pr\left(\max_{1\le t\le T} \left|\dot{z}_{it}\right| \ge \dot{d}_{i,nT}\right) \le \frac{nT}{\dot{d}_{i,nT}^{4}} \sup_{1\le i\le n}\max_{1\le t\le T} \mathbb{E}\left[\left|\dot{z}_{it}\right|^{4} 1\left\{\left|\dot{z}_{it}\right| > \dot{d}_{i,nT}\right\}\right] \\
=\ & \sup_{1\le i\le n}\max_{1\le t\le T} o\left(\frac{nT^{3}}{\dot{d}_{i,nT}^{4}}\right) = o\left(1\right),
\end{aligned}
$$

which is guaranteed by our assumption for $\dot{d}_{i,nT}$. Since $n^{\frac{1}{4}}T\sqrt{\log(n)} \succ n^{\frac{1}{4}}T^{\frac{3}{4}}\log(n)$, we need $\dot{M}_{nT} \succ n^{\frac{1}{4}}T\sqrt{\log(n)}$. Therefore we have, for any $\epsilon > 0$,

$$\sup_{1\leq i\leq n}\left|\sum_{t=1}^{T} y_{i,t-1}u_{it}\right| = O_p\left(n^{\frac{1}{4}}T^{1+\epsilon}\sqrt{\log(n)}\right).$$

For (DD.2.2) term, we have

$$\sup_{1\leq i\leq n}\left|T\overline{y}_{i,-1}\overline{u}_i\right| \leq T\sup_{1\leq i\leq n}\left|\overline{y}_{i,-1}\right|\sup_{1\leq i\leq n}\left|\overline{u}_i\right|.$$

By applying exponential inequality, we have $\sup_{1\leq i\leq n}\left|T\overline{u}_i\right| = O_p\left(\sqrt{\log n}n^{\frac{1}{4}}T^{\frac{1}{2}}\right)$. Similarly, by applying exponential inequality, we can show that $T\sup_{1\leq i\leq n}\left|\overline{y}_{i,-1}\right| = O_p\left(\sqrt{\log n}n^{\frac{1}{4}}T^{\frac{3}{2}}\right)$. Therefore we have

$$\sup_{1\leq i\leq n}\left|T\overline{y}_{i,-1}\overline{u}_i\right| \leq T\sup_{1\leq i\leq n}\left|\overline{y}_{i,-1}\right|\sup_{1\leq i\leq n}\left|\overline{u}_i\right| = O_p\left((\log n)\, n^{\frac{1}{2}}T\right).$$

and

$$\sup_{1\leq i\leq n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| = O_p\left(n^{\frac{1}{2}+\epsilon}T\log(n)\right).$$

Therefore, the uniform upper bound for individual least squares is

$$\sup_{1\leq i\leq n, c_i^0=0}\left|\widehat{c}_i - c_i^0\right| = O_p\left(\frac{n^{\frac{1}{2}+\epsilon}\log(n)\log_2(T)}{T}\right),$$

with arbitrary $\epsilon > 0$. ■

Lemma A.2.11 *Suppose Assumption 1 and 2 hold. Then,*

$$\sup_{1\leq i\leq n}\left|\widehat{c}_i - c_i^0\right| = O_p\left(\frac{n^{\frac{1}{4}+\epsilon}\log(n)}{T^{\frac{1+\gamma}{2}}}\right) =: O_p(\delta_{nT}),$$

for arbitrary $\epsilon > 0$.

Proof. The uniform convergence rate of individual estimator follows

$$\sup_{1\leq i\leq n}\left|\widehat{c}_i - c_i^0\right|$$

$$
\begin{aligned}
&= \max\left\{\sup_{1\le i\le n, c_i^0>0}\left|\widehat{c}_i - c_i^0\right|, \sup_{1\le i\le n, c_i^0<0}\left|\widehat{c}_i - c_i^0\right|, \sup_{1\le i\le n, c_i^0=0}\left|\widehat{c}_i - c_i^0\right|\right\} \\
&= \max\left\{O_p\left(\frac{1}{T^{\frac{\gamma}{2}}\left(\rho_u\right)^T}\right), O_p\left(\frac{n^{\frac{1}{4}+\epsilon}\sqrt{\log\left(n\right)}}{T^{\frac{1+\gamma}{2}}}\right), O_p\left(\frac{n^{\frac{1}{2}+\epsilon}\log\left(n\right)\log_2(T)}{T}\right)\right\} \\
&= O_p\left(\frac{n^{\frac{1}{4}+\epsilon}\sqrt{\log\left(n\right)}}{T^{\frac{1+\gamma}{2}}}\right) \\
&= O_p(\delta_{nT}),
\end{aligned}
$$

by the rate restriction $T^{2\gamma-2}n\left(\log(n)\right)^2\left(\log_2 T\right)^4 \to 0$, Lemma A.2.8, Lemma A.2.9, Lemma A.2.10 and the dominance of exponential rates. We conclude our proof. ■

Lemma A.2.12 *If Assumption 1 and 2 hold,*

$$
d_H\left(\overline{c}^0, \widehat{\alpha}\right) = O_p\left(\delta_{nT}^{\frac{1}{2}}\right),
$$

where δ_{nT} is defined in Lemma A.2.11.

When n grows more slowly than T, the Hausdorff distance between $\overline{c}^0$ and $\widehat{\alpha}$ is asymptotically diminishing. This lemma shows the individual and uniform consistency of the modified k-means algorithm.

Proof. (This proof follows derivations in Su et al. (2019)). Let $Q_n\left(\alpha\right) = \sum_{g=1}^{K^0}\min_{1\le l\le K}\left(\overline{c}_g^0 - \alpha_l\right)^2\pi_g$.

(i) Firstly we derive the convergence rate of $\widehat{Q}_n\left(\alpha\right) - Q_n\left(\alpha\right)$ uniformly over $\alpha \in \mathcal{M} := \left\{(\alpha_1, ..., \alpha_{K^0}) : \sup_{1\le g\le K^0}|\alpha_g| \le 2c_u\right\}$ for the defined upper bound c_u. Let $R_n = \sup_{1\le i\le n}\left|\widehat{c}_i - \overline{c}_{g_i^0}^0\right|$. By Assumptions 1 and 2,

$$
R_n = \sup_{1\le i\le n}\left|\widehat{c}_i - \overline{c}_{g_i^0}^0\right| \preceq O_p(\delta_{nT}) \le c_u \text{ a.s.},
$$

by our previous derivations. In addition, we have,

$$
\left(\widehat{c}_i - \alpha_l\right)^2 \;\ge\; \left(\overline{c}_{g_i^0}^0 - \alpha_l\right)^2 - 2\left|\left(\overline{c}_{g_i^0}^0 - \widehat{c}_i\right)\left(\overline{c}_{g_i^0}^0 - \alpha_l\right)\right| - \left(\overline{c}_{g_i^0}^0 - \widehat{c}_i\right)^2
$$

$$\geq \left(\overline{c}^0_{g^0_i} - \alpha_l\right)^2 - 2\left|\left(\overline{c}^0_{g^0_i} - \widehat{c}_i\right)\left(\overline{c}^0_{g^0_i} - \alpha_l\right)\right| - \left(\sup_{1\leq i\leq n}\left(\overline{c}^0_{g^0_i} - \widehat{c}_i\right)\right)^2$$
$$\geq \left(\overline{c}^0_{g^0_i} - \alpha_l\right)^2 - 2\left(\sup_{1\leq i\leq n}\left|\overline{c}^0_{g^0_i} - \widehat{c}_i\right|\right)\left(\left|\overline{c}^0_{g^0_i} - \alpha_l\right|\right) - \left(\sup_{1\leq i\leq n}\left(\overline{c}^0_{g^0_i} - \widehat{c}_i\right)\right)^2$$
$$\geq \left(\overline{c}^0_{g^0_i} - \alpha_l\right)^2 - 2\left(\sup_{1\leq i\leq n}\left|\overline{c}^0_{g^0_i} - \widehat{c}_i\right|\right)\left(\left|\overline{c}^0_{g^0_i}\right| + |\alpha_l|\right) - \left(\sup_{1\leq i\leq n}\left(\overline{c}^0_{g^0_i} - \widehat{c}_i\right)\right)^2.$$

Taking $\min_{1\leq l\leq K^0}$ on both sides and take average over i, we have

$$\frac{1}{n}\sum_{i=1}^{n}\min_{1\leq l\leq K^0}\left(\widehat{c}_i - \alpha_l\right)^2 \geq \frac{1}{n}\sum_{i=1}^{n}\min_{1\leq l\leq K^0}\left(\overline{c}^0_{g^0_i} - \alpha_l\right)^2 - 2\delta_{nT}\left(2c_u + c_u\right) - c_u\delta_{nT},$$

where $\left(\sup_{1\leq i\leq n}\left(\overline{c}^0_{g^0_i} - \widehat{c}_i\right)\right)^2 \leq c_u\delta_{nT}$ is of lower order term. Therefore we have $\widehat{Q}_n(\alpha) \geq Q_n(\alpha) - 7c_u\delta_{nT}$. Similarly we have

$$\left(\widehat{c}_i - \alpha_l\right)^2 \leq \left(\overline{c}^0_{g^0_i} - \alpha_l\right)^2 + 2\left(\sup_{1\leq i\leq n}\left|\overline{c}^0_{g^0_i} - \widehat{c}_i\right|\right)\left(\left|\overline{c}^0_{g^0_i}\right| + |\alpha_l|\right) + \left(\sup_{1\leq i\leq n}\left(\overline{c}^0_{g^0_i} - \widehat{c}_i\right)\right)^2.$$

Hence we derive $\widehat{Q}_n(\alpha) \leq Q(\alpha) + 7c_u\delta_{nT}$ and $\sup_{\alpha\in\mathcal{M}}\left|\widehat{Q}_n(\alpha) - Q_n(\alpha)\right| \leq 7c_u\delta_{nT}$.

(ii) Secondly, we show that $\widehat{\alpha} \in \mathcal{M}$. Denote $\widehat{\alpha} := \{\widehat{\alpha}_1, \widehat{\alpha}_1, ..., \widehat{\alpha}_{K^0}\}$. By our assumption, we have

$$\sup_{1\leq i\leq n}|\widehat{c}_i| \leq \sup_{1\leq i\leq n}\left|\widehat{c}_i - \overline{c}^0_{g^0_i}\right| + \sup_{1\leq i\leq n}\left|\overline{c}^0_{g^0_i}\right| \leq 2c_u.$$

Denote $I_n(g) := \{i : g = \arg\min_{1\leq g\leq K}|\widehat{c}_i - \widehat{\alpha}_l|\}$ for some $g \leq K^0$. Following Su et al. (2019), we use contradictions to demonstrate our results:

(ii.1) If $|\widehat{\alpha}_g| > 2c_u$ and $I_n(g) = \varnothing$, then we choose $\widehat{\alpha}' := \left\{\widehat{\alpha}_1, ..., \widehat{\alpha}_{g-1}, \widehat{\alpha}'_g, \widehat{\alpha}_{g+1}..., \widehat{\alpha}_{K^0}\right\}$, with $\widehat{\alpha}'_g = \widehat{c}_i$ for each $i \in \{1, 2, ..., n\}$. Therefore we can get $\left|\widehat{\alpha}'_g\right| \leq 2c_u < |\widehat{\alpha}_g|$ and $\widehat{Q}_n\left(\widehat{\alpha}'\right) < \widehat{Q}_n(\widehat{\alpha})$. This demonstrates a contradiction.

(ii.2) If $|\widehat{\alpha}_g| > 2c_u$ and $I_n(g) \neq \varnothing$, then we choose $\widehat{\alpha}' := \left\{\widehat{\alpha}_1, ..., \widehat{\alpha}_{g-1}, \widehat{\alpha}'_g, \widehat{\alpha}_{g+1}..., \widehat{\alpha}_K\right\}$, with $\widehat{\alpha}'_g = \frac{1}{|I_n(g)|}\sum_{i\in I_n(g)}\widehat{c}_i$ for any $i \in \{1, 2, ..., n\}$. Here $|I_n(g)|$ denotes the cardinality of $I_n(g)$. This shows $\left|\widehat{\alpha}'_g\right| \leq 2c_u < \left|\widehat{\alpha}_g\right|$ and $\widehat{Q}_n\left(\widehat{\alpha}'\right) < \widehat{Q}_n(\widehat{\alpha})$. This is a contradiction too.

Based on (ii.1) and (ii.2), $\left|\widehat{\alpha}_g\right| \le 2c_u$ for each $g \in \{1, 2, ..., K^0\}$.

(iii) We show for any $\eta > 0$ $\inf_{\alpha: d_H\left(\alpha, \overline{c}^0\right) > \eta} Q_n(\alpha) \ge \underline{M} \min\left\{\eta^2, (c^*)^2\right\}$ where $\overline{c}^0 := \left\{\overline{c}_1^0, \overline{c}_2^0, ..., \overline{c}_{K^0}^0\right\}$ and $\underline{M} \le \pi_g \le 1$ for $g = 1, 2, ..., K^0$. If there exists some $l_o \in \{1, 2, ..., K^0\}$ and two indexes g_1 and g_2 such that

$$l_o = \arg\min_{1 \le l \le K} \left|\overline{c}_{g_1}^0 - \alpha_{l_o}\right| = \arg\min_{1 \le l \le K} \left|\overline{c}_{g_2}^0 - \alpha_{l_o}\right|,$$

then we have

$$\begin{aligned} Q_n(\alpha) &\ge \pi_{g_1}\left(\overline{c}_{g_1}^0 - \alpha_{l_o}\right)^2 + \pi_{g_2}\left(\overline{c}_{g_2}^0 - \alpha_{l_o}\right)^2 \\ &\ge \underline{M}\left(\left|\overline{c}_{g_1}^0 - \alpha_{l_o}\right| + \left|\overline{c}_{g_2}^0 - \alpha_{l_o}\right|\right)^2 \ge \underline{M}\left(\overline{c}_{g_1}^0 - \overline{c}_{g_2}^0\right)^2 \\ &\ge \underline{M}(c^*)^2. \end{aligned}$$

Besides, if there does not exist such an l_o, then there is one-to-one mapping h: $\{1, 2, ..., K^0\} \to \{1, 2, ..., K^0\}$ such that

$$h(g) = \arg\min_{1 \le l \le K} \left|\overline{c}_g^0 - \alpha_l\right|.$$

Thus $Q_n(\alpha) = \sum_{g=1}^{K^0} \pi_g \left(\overline{c}_g^0 - \alpha_{h(g)}\right)^2 \ge \left(\inf_{1 \le g \le K^0} \pi_g\right) d_H^2\left(\alpha, \overline{c}^0\right) \ge \underline{M}\eta^2$. Then we show (iii).

(iv) Lastly, we show that $d_H\left(\widehat{\alpha}, \overline{c}^0\right) \le \frac{\sqrt{15\overline{c}\delta_{nT}}}{\sqrt{\underline{M}}}$ for some constant $D > 0$ and arbitrary $\epsilon > 0$. We have

$$\begin{aligned} &\Pr\left(d_H\left(\widehat{\alpha}, \overline{c}^0\right) \ge \frac{\sqrt{15\overline{c}\delta_{nT}}}{\sqrt{\underline{M}}}\right) \\ =\ &\Pr\left(d_H\left(\widehat{\alpha}, \overline{c}^0\right) \ge \frac{\sqrt{15\overline{c}\delta_{nT}}}{\sqrt{\underline{M}}}, Q_n(\widehat{\alpha}) \ge Q_n\left(\overline{c}^0\right) + \min\left\{\underline{M}(c^*)^2, 15c_u\delta_{nT}\right\}\right) \\ \le\ &\Pr\left(Q_n(\widehat{\alpha}) \ge Q_n\left(\overline{c}^0\right) + \min\left\{\underline{M}(c^*)^2, 15c_u\delta_{nT}\right\}\right) \\ \le\ &\Pr\left(\widehat{Q}_n(\widehat{\alpha}) + R_n \ge \widehat{Q}_n\left(\overline{c}^0\right) - R_n + \min\left\{\underline{M}(c^*)^2, 15c_u\delta_{nT}\right\}\right) \\ =\ &\Pr\left(2R_n \ge \widehat{Q}_n\left(\overline{c}^0\right) - \widehat{Q}_n(\widehat{\alpha}) + \min\left\{\underline{M}(c^*)^2, 15c_u\delta_{nT}\right\}\right) \\ \le\ &\Pr\left(2R_n \ge \min\left\{\underline{M}(c^*)^2, 15c_u\delta_{nT}\right\}\right) = o(1), \end{aligned}$$

since $\widehat{Q}_n(\overline{c}^0) - \widehat{Q}_n(\widehat{\alpha}) \geq 0$. Note the fact that for large enough n and T, we have

$$2R_n \leq 2 \cdot 7c_u\delta_{nT} < 15c_u\delta_{nT} < \underline{M}\left(c^*\right)^2.$$

We complete the whole proof. ■

A.2.5 Proof of Theorem 3.4.3

Proof. By Assumptions 1 and 2, under joint convergence framework $(n,T) \to \infty$, there is a one-to-one mapping $F_n : \{1,2,...,K^0\} \to \{1,2,...,K^0\}$, such that

$$\sup_g \left|\widehat{\alpha}_g - \overline{c}^0_{F_n(g)}\right| \preceq O_p\left(\delta_{nT}^{\frac{1}{2}}\right).$$

Without losing generality, we can assume that $F_n(g) = g$ such that

$$\overline{R}_n = \sup_g \left|\widehat{\alpha}_g - \overline{c}^0_{F_n(g)}\right| = \sup_g \left|\widehat{\alpha}_g - \overline{c}^0_g\right| \preceq O_p\left(\delta_{nT}^{\frac{1}{2}}\right).$$

If $\widehat{g}_i \neq g_i^0$, then $|\widehat{c}_i - \widehat{\alpha}_{\widehat{g}_i}| \leq \left|\widehat{c}_i - \widehat{\alpha}_{g_i^0}\right|$. This, in conjunction with triangle inequality, implies that

$$\left|\widehat{\alpha}_{\widehat{g}_i} - \widehat{\alpha}_{g_i^0}\right| - \left|\widehat{c}_i - \widehat{\alpha}_{g_i^0}\right| \leq \left|\widehat{\alpha}_{\widehat{g}_i} - \widehat{c}_i\right| \leq \left|\widehat{c}_i - \widehat{\alpha}_{g_i^0}\right|.$$

It follows that $\left|\widehat{c}_i - \widehat{\alpha}_{g_i^0}\right| \geq \frac{1}{2}\left|\widehat{\alpha}_{\widehat{g}_i} - \widehat{\alpha}_{g_i^0}\right|$. Therefore we have,

$$\begin{aligned}
\delta_{nT} + \overline{R}_n &\geq \left|\widehat{c}_i - \overline{c}^0_{g_i^0}\right| + \left|\overline{c}^0_{g_i^0} - \widehat{\alpha}_{g_i^0}\right| \geq \left|\widehat{c}_i - \widehat{\alpha}_{g_i^0}\right| \geq \frac{1}{2}\left|\widehat{\alpha}_{\widehat{g}_i} - \widehat{\alpha}_{g_i^0}\right| \\
&= \frac{1}{2}\left|\left(\overline{c}^0_{\widehat{g}_i} - \overline{c}^0_{g_i^0}\right) + \left(\widehat{\alpha}_{\widehat{g}_i} - \overline{c}^0_{\widehat{g}_i}\right) + \left(\overline{c}^0_{g_i^0} - \widehat{\alpha}_{g_i^0}\right)\right| \\
&\geq \frac{1}{2}\left|\left(\overline{c}^0_{\widehat{g}_i} - \overline{c}^0_{g_i^0}\right)\right| - \frac{1}{2}\left|\left(\widehat{\alpha}_{\widehat{g}_i} - \overline{c}^0_{\widehat{g}_i}\right)\right| - \frac{1}{2}\left|\left(\overline{c}^0_{g_i^0} - \widehat{\alpha}_{g_i^0}\right)\right| \\
&\geq \frac{1}{2}\left|\left(\overline{c}^0_{\widehat{g}_i} - \overline{c}^0_{g_i^0}\right)\right| - \sup_g \left|\left(\widehat{\alpha}_g - \overline{c}^0_g\right)\right| \\
&= \frac{1}{2}\left|\left(\overline{c}^0_{\widehat{g}_i} - \overline{c}^0_{g_i^0}\right)\right| - \overline{R}_n \geq \frac{c^*}{2} - \overline{R}_n.
\end{aligned}$$

This implies that $\mathbf{1}\left\{\widehat{g}_i \neq g_i^0\right\} \leq \mathbf{1}\left\{2\overline{R}_n + \delta_{nT} \geq \frac{c^*}{2}\right\}$. Noting that the right hand side of the above term is independent of i, we have

$$\begin{aligned}
\Pr\left\{\sup_{1\leq i\leq n} \mathbf{1}\left\{\widehat{g}_i \neq g_i^0\right\} > 0\right\} &\leq \Pr\left\{2\overline{R}_n + \delta_{nT} \geq \frac{c^*}{2}\right\} \\
&\leq \Pr\left\{\delta_{nT} + 2\delta_{nT}^{\frac{1}{2}} \geq \frac{c^*}{2}\right\} \xrightarrow{p} 0,
\end{aligned}$$

as $(n, T) \to \infty$. This concludes our proof. ■

A.2.6 Proof of Theorem 3.4.4

Proof. For mixed-root panel with latent groups, the difference between the post-classification estimator and the oracle estimator is asymptotically diminishing. For the explosive group, the proof here is identical to the case of recursive k-means algorithm. For the stationary and unit root group, derivations follow similar procedures. These results are also shown in Phillips (2014b). ■

A.2.7 Proof of Lemma 3.4.1

Proof. For any $g = 1, 2, ..., K^0$ with $\alpha_g^0 > 0$, we have

$$\begin{aligned}
\widehat{\widetilde{u}}_{it} &= \widetilde{y}_{it} - \widehat{\overline{\rho}}_{\widehat{g}}\widetilde{y}_{i,t-1}, \\
\widetilde{u}_{it} &= \widetilde{y}_{it} - \overline{\rho}_g^0\widetilde{y}_{i,t-1}.
\end{aligned}$$

Therefore we have

$$\begin{aligned}
\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g}\sum_{t=1}^{T}\widehat{\widetilde{u}}_{it}^2 &= \frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_{\widehat{g}}}\sum_{t=1}^{T}\widetilde{u}_{it}^2 + (\widehat{\overline{\rho}}_{\widehat{g}} - \overline{\rho}_g^0)^2\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \\
&\quad -(\widehat{\overline{\rho}}_{\widehat{g}} - \overline{\rho}_g^0)\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} \\
&= \frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{u}_{it}^2 + \frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g\backslash G_g^0}\sum_{t=1}^{T}\widetilde{u}_{it}^2 - \frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0\backslash\widehat{G}_g}\sum_{t=1}^{T}\widetilde{u}_{it}^2
\end{aligned}$$

$$
\begin{aligned}
&+(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)^2\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2+(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)^2\frac{1}{2n_{\widehat{g}}T}\sum_{i\in \widehat{G}_g\backslash G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\\
&-(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)^2\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0\backslash \widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2-(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\\
&-(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)\frac{1}{2n_{\widehat{g}}T}\sum_{i\in \widehat{G}_g\backslash G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\\
&+(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0\backslash \widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}. \qquad \text{(A.2.26)}
\end{aligned}
$$

To demonstrate the dominance of the first term in (A.2.26), we first show that $n_{\widehat{g}} - n_g = o_p(1)$. Note the fact that $\mathbf{1}\left\{i \in \widehat{G}_g\right\} - \mathbf{1}\left\{i \in G_g^0\right\} = \mathbf{1}\left\{i \in \widehat{G}_g\backslash G_g^0\right\} - \mathbf{1}\left\{i \in G_g^0\backslash \widehat{G}_g\right\}$. By Markov inequality, for any $\epsilon > 0$, we have

$$
\begin{aligned}
\Pr\left(|n_{\widehat{g}}-n_g|>2\epsilon\right) &\le \Pr\left(\sum_{i=1}^{n}\mathbf{1}\left\{i\in \widehat{G}_g\backslash G_g^0\right\}\ge\epsilon\right)+\Pr\left(\sum_{i=1}^{n}\mathbf{1}\left\{i\in G_g^0\backslash \widehat{G}_g\right\}\ge\epsilon\right)\\
&\le \frac{1}{\varepsilon}\sum_{i=1}^{n}\Pr\left(\widehat{F}_{g,i}\right)+\frac{1}{\varepsilon}\sum_{i=1}^{n}\Pr\left(\widehat{E}_{g,i}\right)\\
&= \frac{1}{\varepsilon}\sum_{g=1}^{K^0}\sum_{i\in G_g^0}\Pr\left(\widehat{F}_{g,i}\right)+\frac{1}{\varepsilon}\sum_{g=1}^{K^0}\sum_{i\in G_g^0}\Pr\left(\widehat{E}_{g,i}\right)\\
&= o(1).
\end{aligned}
$$

The asymptotic negligibility of $\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0\backslash\widehat{G}_g}\sum_{t=1}^{T}\widetilde{u}_{it}^2$, $(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)^2\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0\backslash\widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2$ and $(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)\frac{1}{2n_gT}\sum_{i\in G_g^0\backslash\widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}$ follows the same techniques as in (ii-i) of the Theorem 3.4.2. The asymptotic negligibility of $\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g\backslash G_g^0}\sum_{t=1}^{T}\widetilde{u}_{it}^2$, $(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)^2\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g\backslash G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2$ and $(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g\backslash G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}$ follows the same techniques as in (i) (ii) of the Theorem 3.4.2. The asymptotic negligibility of $(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)^2\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2$ and $(\widehat{\rho}_{\widehat{g}}-\overline{\rho}_g^0)\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}$ follows identically as the proof of the oracle estimator. Therefore, under the joint convergence $(n, T) \to \infty$, we have

$$
\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g}\sum_{t=1}^{T}\widehat{\widetilde{u}}_{it}^2=\frac{1}{2n_gT}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{u}_{it}^2+o_p(1).
$$

For the stationary and unit root group, the consistency of the variance estimate follows Hahn and Kuersteiner (2002), and Phillips and Moon (1999). ■

A.2.8 Proof of Theorem 3.4.5

Proof. Under the joint convergence $(n,T) \to \infty$ and the null hypothesis, the $\widetilde{t}_{\widehat{g}}$ statistics follows chi-square distribution. Under the alternative hypothesis of explosive roots, we have

$$\begin{aligned} \frac{\left(\widehat{\rho}_g - 1\right)\left(\sqrt{\sum_{i\in G_g^0}\sum_{t=1}^T \widetilde{y}_{i,t-1}^2}\right)}{\widetilde{\sigma}} &= \frac{\left(\widehat{\rho}_g - \overline{\rho}_g^0\right) - \left(1-\overline{\rho}_g^0\right)}{\widetilde{\sigma}}\sqrt{\sum_{i\in G_g^0}\sum_{t=1}^T \widetilde{y}_{i,t-1}^2} \\ = \frac{\left(\widehat{\rho}_g - \overline{\rho}_g^0\right)\left(\sqrt{\sum_{i\in G_g^0}\sum_{t=1}^T \widetilde{y}_{i,t-1}^2}\right)}{\widetilde{\sigma}} &- \frac{\left(1-\overline{\rho}_g^0\right)\left(\sqrt{\sum_{i\in G_g^0}\sum_{t=1}^T \widetilde{y}_{i,t-1}^2}\right)}{\widetilde{\sigma}} \\ = O_p(1) + O_p\left(\sqrt{n}\left(\overline{\rho}_g^0\right)^T\right). \end{aligned}$$

Therefore we conclude our proof. ■

A.2.9 Proof of Theorem 3.4.6

Proof. As $(n,T) \to \infty$, we have $\beta_{Tn} \to 0$ and $cv_{\beta_{Tn}} \to \infty$. Since $\widetilde{t}_g \xrightarrow{d} \mathcal{N}(0,1)$ under the null hypothesis of no bubble episode for $g = 1, 2, ...K^0$, we have

$$\lim_{(n,T)\to\infty} \Pr\left(\widetilde{t}_g(\cdot) > cv_{\beta_{Tn}}\right) = \Pr\left(\mathcal{N}(0,1) = \infty\right) = 0,$$

where $cv_{\beta_{Tn}} \to \infty$. Hence, with joint convergence $(n,T) \to \infty$, no origination point for an explosive model in the data will be detected under the null hypothesis.

■

A.2.10 Proof of Theorem 3.4.7

Proof. We fix $g = 1, 2, ..., K^0$**.** Due to the uniform consistency of the classification algorithm, $\widehat{G}_g$ is equivalent to G_g^0, and $\widehat{g}_i$ is equivalent to g_i^0 asymptotically. Therefore, there is no difference between the true group and the estimated group. For all individual in G_g^0, the time series data is sampled from

$$y_{it} = \mu_i + \rho_{it} y_{i,t-1} + u_{it};\ i = 1, 2, ..., n; i = 1, 2, ..., n, \tag{A.2.27}$$

where $\rho_{it} = 1 + \frac{c_{it}}{T^\gamma}$, and $c_{it} = c_{1i}^0 \mathbf{1}\left\{t < \tau_{g_i^0}^e\right\} + c_{2i}^0 \mathbf{1}\left\{t \geq \tau_{g_i^0}^e\right\}$. For the g-th group, the true distance parameter of explosive episode is α_{2g}^0 (or $\overline{c}_{2g_i^0}$ and c_{2i}^0 equivalently). The true value of the slope in the g-th group is $\overline{\rho}_{2g}^0 \left(:= \exp\left(\frac{\alpha_{2g}^0}{T^\gamma}\right)\right)$, and we write as $\overline{\rho}_{2g}$ for simplicity. The true value of the explosive slope in the i-th individual is $\rho_{2i}^0 \left(:= \exp\left(\frac{c_{2i}^0}{T^\gamma}\right)\right)$, and we write as ρ_{2i} for simplicity. For any $i \in G_g^0$ (The estimated group $\widehat{G}_g$ is equivalent to the true group G_g^0, for $g = 1, 2, ..., K^0$), we have $g_i^0 = g$. For $r_g < r_g^e$, we have the convergence in distribution,

$$\widetilde{t}_g(r_g) \xrightarrow{d} \mathcal{N}(0, 1),$$

under the model of (A.2.27). Therefore, under the alternative hypothesis, we have for $r_g < r_g^e$, $\Pr\left(\widehat{r}_g^e < r_g^e\right) \to 0$ as $(n, T) \to \infty$.

Next suppose the data is sampled over $t = 1, 2, ..., \tau_{g_i^0} = \left[T r_{g_i^0}\right]$, where $r_{g_i^0} \geq r_{g_i^0}^e$ for each $i \in G_g^0$. In this case, the data $\left\{y_{it} : t = \tau_{g_i^0}^e, ..., \tau_{g_i^0}\right\}$ satisfies

$$y_{it} = \mu_i + \rho_{2i} y_{i,t-1} + u_{it} = \sum_{j=0}^{t-\tau_{g_i^0}^e} \rho_{2i}^j u_{i,t-j} + \sum_{j=0}^{t-\tau_{g_i^0}^e} \rho_{2i}^j \mu_i + \rho_{2i}^{t-\tau_{g_i^0}^e+1} y_{i,\tau_{g_i^0}^e-1}.$$

As $t - \tau_{g_i^0}^e \to \infty$, the following asymptotic theory holds,

$$\frac{1}{T^{\frac{\gamma}{2}}} \sum_{j=0}^{t-\tau_{g_i^0}^e} \rho_{2i}^{j-\left(t-\tau_{g_i^0}^e\right)} u_{i,t-j} \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{2c_{2i}^0}\right),$$

where $\frac{1}{\sqrt{T}} y_{i,\tau_{g_i^0}^e-1} \Rightarrow B_i\left(\tau_{g_i^0}^e\right)$ by the functional central limit theorem. Then as

$t-\tau^{e}_{g_i^0}\to\infty$, we have

$$\begin{aligned}\frac{\rho_{2i}^{-\left(t-\tau^{e}_{g_i^0}\right)}}{\sqrt{T}}y_{i,t} &= \frac{1}{T^{\frac{1-\gamma}{2}}}\frac{1}{\sqrt{T^{\gamma}}}\sum_{j=0}^{t-\tau^{e}_{g_i^0}}\rho_{2i}^{-\left(t-\tau^{e}_{g_i^0}-j\right)}u_{i,t-j}+\frac{\rho_{2i}}{\sqrt{T}}y_{i,\tau^{e}_{g_i^0}-1}\\ &\quad+\frac{1}{T^{\frac{1-\gamma}{2}}}\frac{1}{\sqrt{T^{\gamma}}}\sum_{j=0}^{t-\tau^{e}_{g_i^0}}\rho_{2i}^{-\left(t-\tau^{e}_{g_i^0}-j\right)}\mu_i\\ &\Rightarrow B_i\left(r^{e}_{g_i^0}\right).\end{aligned}$$

So that for each $i\in G_g^0$, we have $y_{it}\sim\sqrt{T}\rho_{2i}^{\left(t-\tau^{e}_{g_i^0}\right)}B_i\left(r^{e}_{g_i^0}\right)$, for all $t-\tau^{e}_{g_i^0}\to\infty$.

Now consider the centered quantities $\widetilde{y}_{it}=y_{it}-\frac{1}{\tau_{g_i^0}}\sum_{j=1}^{\tau_{g_i^0}}y_{i,j}$. Note the fact that $g_i^0=g$. For $\tau_g=[Tr_g]$ (or $\tau_{g_i^0}=\left[Tr_{g_i^0}\right]$) and $r_g^e<r_g$ (or $r^{e}_{g_i^0}<r_{g_i^0}$), we have

$$\frac{1}{\tau_{g_i^0}\sqrt{T}}\sum_{j=1}^{\tau_{g_i^0}}y_{ij}=\frac{1}{\tau_{g_i^0}\sqrt{T}}\sum_{j=\tau^{e}_{g_i^0}}^{\tau_{g_i^0}}y_{ij}+\frac{\tau^{e}_{g_i^0}}{\tau_{g_i^0}}\frac{1}{\tau^{e}_{g_i^0}}\sum_{j=1}^{\tau^{e}_{g_i^0}-1}\frac{y_{ij}}{\sqrt{T}}\sim\frac{1}{\tau_{g_i^0}\sqrt{T}}\sum_{j=\tau^{e}_{g_i^0}}^{\tau_{g_i^0}}y_{ij}+\frac{r^{e}_{g_i^0}}{r_{g_i^0}}\int_0^1B_i\left(s\right)ds,$$

and

$$\begin{aligned}\frac{1}{\tau_{g_i^0}}\sum_{j=\tau^{e}_{g_i^0}}^{\tau_{g_i^0}}y_{ij} &= \frac{1}{\tau_{g_i^0}}\sum_{j=\tau^{e}_{g_i^0}}^{\tau_{g_i^0}}\rho_{2i}^{(j-\tau^{e}_{g_i^0})}\left(\rho_{2i}^{-(j-\tau^{e}_{g_i^0})}y_{ij}\right)=\frac{y_{i,\tau^{e}_{g_i^0}}}{\tau_{g_i^0}}\sum_{k=0}^{\tau_{g_i^0}-\tau^{e}_{g_i^0}}\rho_{2i}^{k}\left(1+o(1)\right)\\ &= y_{i,\tau^{e}_{g_i^0}}\frac{\rho_{2i}^{\tau_{g_i^0}-\tau^{e}_{g_i^0}+1}-1}{\tau_{g_i^0}\left(\rho_{2i}-1\right)}\left(1+o\left(1\right)\right)\\ &= y_{i,\tau^{e}_{g_i^0}}\frac{T^{\gamma}\rho_{2i}^{\tau_{g_i^0}-\tau^{e}_{g_i^0}+1}}{\tau_{g_i^0}c_{2i}^{0}}\left(1+o\left(1\right)\right).\end{aligned}$$

It follows that

$$\widetilde{y}_{it}=y_{it}-\frac{1}{\tau_{g_i^0}}\sum_{j=1}^{\tau_{g_i^0}}y_{i,j}=\left[\rho_{2i}^{\left(t-\tau^{e}_{g_i^0}\right)}-\frac{T^{\gamma}\rho_{2i}^{\left(\tau_{g_i^0}-\tau^{e}_{g_i^0}\right)}}{\tau_{g_i^0}c_{2i}^{0}}\right]y_{i,\tau^{e}_{g_i^0}}\left\{1+o\left(1\right)\right\}.$$

Therefore we have the following asymptotics for the sample moment in the post-classification estimator as,

$$\sum_{i\in G_g^0}\sum_{j=1}^{\tau_g^e}\widetilde{y}_{it}^2=\frac{n_gT^{2\gamma}\tau_g^e\overline{\rho}_{2g}^{2\left(\tau_g-\tau_g^e\right)}}{\tau_g^2\left(\alpha_{2g}^0\right)^2}\mathbb{E}y_{i,\tau_g^e}^2\left\{1+o\left(1\right)\right\}.$$

Using these results in conjunction with the standard unit root limit theory, we have

$$\sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i^0}}\widetilde{y}_{i,j-1}^2=\sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i^0}^e-1}\widetilde{y}_{i,j-1}^2+\sum_{i\in G_g^0}\sum_{j=\tau_{g_i^0}^e}^{\tau_{g_i^0}}\widetilde{y}_{i,j-1}^2\sim\sum_{i\in G_g^0}\sum_{j=\tau_{g_i^0}^e}^{\tau_{g_i^0}}\widetilde{y}_{i,j-1}^2,\quad\text{(A.2.28)}$$

and

$$\sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i^0}}\widetilde{y}_{i,t-1}\left(y_{i,j}-\rho_{2i}y_{i,j-1}\right)=\sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i^0}^e-1}\widetilde{y}_{i,t-1}\left(u_{i,j}-\frac{c_{2i}^0}{T^{2\gamma}}y_{i,j-1}\right)+\sum_{i\in G_g^0}\sum_{j=\tau_{g_i^0}^e}^{\tau_{g_i^0}}\widetilde{y}_{i,t-1}u_{i,j}\left(1+o\left(1\right)\right).\quad\text{(A.2.29)}$$

Explicit probability limits of (A.2.28) and (A.2.29) are as,

$$\begin{aligned}\sum_{i\in G_g^0}\sum_{j=\tau_{g_i^0}^e}^{\tau_{g_i^0}}\widetilde{y}_{i,j-1}^2 &= \sum_{i\in G_g^0}\sum_{j=\tau_{g_i^0}^e}^{\tau_{g_i^0}}\rho_{2i}^{2\left(j-\tau_{g_i^0}^e\right)}y_{i,\tau_{g_i^0}^e}^2\left(1+o_p\left(1\right)\right)=\frac{n_g\overline{\rho}_{2g}^{2\left(\tau_g-\tau_g^e+1\right)}}{\overline{\rho}_{2g}^2-1}\mathbb{E}y_{i,\tau_g^e}^2\left\{1+o_p\left(1\right)\right\}\\ &= \frac{n_gT^{\gamma}\overline{\rho}_{2g}^{2\left(\tau_g-\tau_g^e\right)}}{2\alpha_{2g}^0}\mathbb{E}y_{i,\tau_g^e}^2\left\{1+o_p\left(1\right)\right\},\end{aligned}$$

which dominates $\sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i^0}^e}\widetilde{y}_{i,j-1}^2$. The above derivations rely on the uniform consistency of the classification algorithms. Besides, we have for each i,

$$\begin{aligned}\sum_{j=\tau_{g_i^0}^e}^{\tau_{g_i^0}}\widetilde{y}_{i,j-1}\widetilde{u}_{i,j} &= \sum_{j=\tau_{g_i^0}^e}^{\tau_{g_i^0}}\rho_{2i}^{j-1-\tau_{g_i^0}^e}y_{i,\tau_{g_i^0}^e}u_{i,j}\left(1+o_p\left(1\right)\right)\\ &= T^{\frac{\gamma}{2}}\rho_{2i}^{\tau_{g_i^0}-\tau_{g_i^0}^e}y_{i,\tau_{g_i^0}^e}\left[\frac{1}{T^{\frac{\gamma}{2}}}\sum_{j=\tau_{g_i^0}^e}^{\tau_{g_i^0}}\rho_{2i}^{-\left(\tau_{g_i^0}-j+1\right)}u_{i,j}\right]\left(1+o_p\left(1\right)\right)\\ &\sim T^{\frac{\gamma+1}{2}}\overline{\rho}_{2g}^{\tau_g-\tau_g^e}\cdot\mathcal{N}\left(0,\frac{\sigma^2}{2\alpha_{2g}^0}\right)\cdot\mathcal{N}\left(0,r_{g_i^0}^e\sigma^2\right).\end{aligned}$$

The terms $\frac{1}{\sqrt{T}} y_{i,\tau^e_{g^0_i}}$ and $\left[\frac{1}{T^{\frac{\gamma}{2}}}\sum_{j=\tau^e_{g^0_i}}^{\tau_{g^0_i}} \rho_{2i}^{-\left(\tau_{g^0_i}-j+1\right)} u_{i,j}\right]$ are joint Guassian and uncorrelated. Therefore the two limiting Guassian processes are independent. Under the joint convergence $(n,T) \to \infty$, we also have

$$\sum_{i\in G^0_g}\sum_{j=\tau^e_{g^0_i}}^{\tau_{g^0_i}} \widetilde{y}_{i,j-1}\widetilde{u}_{i,j} \sim \sqrt{n_g}T^{\frac{\gamma+1}{2}}\overline{\rho}_{2g}^{(\tau_g-\tau^e_g)} \cdot \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha^0_{2g}}\right) \cdot \mathcal{N}\left(0, r^e_g\sigma^2\right),$$

and

$$\sum_{i\in G^0_g}\sum_{j=\tau^e_{g^0_i}}^{\tau_{g^0_i}} \widetilde{y}^2_{i,j-1} \sim \frac{n_g T^{\gamma+1}\overline{\rho}_{2,g}^{2(\tau_g-\tau^e_g)}}{2\alpha^0_{2g}} r^e_g\sigma^2.$$

It follows that for $\tau_g = [Tr_g]$ and $r_g > r^e_g$,

$$\sqrt{n_g}T^{\frac{1+\gamma}{2}}\overline{\rho}_{2g}^{\tau_g-\tau^e_g}\left(\widehat{\overline{\rho}}_{2g} - \overline{\rho}_{2g}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\left(\alpha^0_{2g}\right)}{r^e_g}\right),$$

where under $(n,T) \to \infty$, we have

$$\frac{\overline{\rho}_{2g}^{-\left(\tau_g-\tau^e_g\right)}}{\sqrt{n_g}T^{\frac{1+\gamma}{2}}}\sum_{i\in G^0_g}\sum_{j=\tau^e_{g^0_i}}^{\tau_{g^0_i}} \widetilde{y}_{i,j-1}\widetilde{u}_{i,j} \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha^0_{2g}}\right) \cdot \mathcal{N}\left(0, r^e_g\sigma^2\right),$$

and

$$\frac{2\alpha^0_{2g}\overline{\rho}_{2g}^{-2\left(\tau_g-\tau^e_g\right)}}{n_g T^{1+\gamma}}\sum_{i\in G^0_g}\sum_{j=\tau^e_{g^0_i}}^{\tau_{g^0_i}} \widetilde{y}^2_{i,j-1} \xrightarrow{p} r^e_g\sigma^2.$$

The regression residuals variance estimate is as, for any $r_g \in [r_0, 1]$ and $g = 1, 2, ..., K^0$,

$$\begin{aligned}
\widetilde{\sigma}^2_g(r_g) &= \frac{1}{2n_g\tau_g}\sum_{i\in G^0_g}\sum_{j=1}^{\tau_g}\left(\widetilde{y}_{i,j} - \widehat{\rho}_{2i}(r_g)\widetilde{y}_{i,j-1}\right)^2 \\
&= \frac{1}{2n_g\tau_g}\sum_{i\in G^0_g}\sum_{j=1}^{\tau_g}\left(\widetilde{u}_{i,j} - \left(\widehat{\rho}_{2i}(r_g) - \rho_{2i}\right)\widetilde{y}_{i,j-1}\mathbf{1}\left\{j \geq \tau^e_{g^0_i}\right\} - \left(\widehat{\rho}_{2i}(r_g) - 1\right)\widetilde{y}_{i,j-1}\mathbf{1}\left\{j < \tau^e_{g^0_i}\right\}\right)^2 \\
&\sim \frac{\tau^e_g\overline{\rho}_{2g}^{2\left(\tau_g-\tau^e_g\right)}}{2\left(\alpha^0_{2g}\right)^2\tau^3_g}\mathbb{E}y^2_{i,\tau^e_g}\left(1 + o_p(1)\right),
\end{aligned}$$

due to the fact that $\frac{1}{2n_g\tau_g}\sum_{i\in G_g^0}\sum_{j=1}^{\tau_g}\widetilde{y}_{i,j-1}^2=\frac{\tau_g^e\overline{\rho}_{2g}^{2\left(\tau_g-\tau_g^e\right)}}{2\left(\alpha_{2g}^0\right)^2\tau_g^3}\mathbb{E}y_{i,\tau_g^e}^2\left(1+o_p\left(1\right)\right).$

Collecting all results, the $\widetilde{t}_g\left(r_g\right)$ statistics for any $r_g\in\left[r_g^e,1\right]$, is as

$$
\begin{aligned}
\widetilde{t}_g\left(r_g\right) &= \frac{\left(\widehat{\rho}_{2g}(r_g)-1\right)\sqrt{\sum_{i\in\widehat{G}_g}\sum_{j=1}^{\tau_g}\widetilde{y}_{i,j-1}^2}}{\widetilde{\sigma}_g\left(r_g\right)} \\
&= \frac{\left(\widehat{\rho}_{2g}(r_g)-1\right)\tau_g\sqrt{\frac{1}{\tau_g^2}\sum_{i\in G_g^0}\sum_{j=1}^{\tau_g}\widetilde{y}_{i,j-1}^2}}{\widetilde{\sigma}_g\left(r_g\right)} \\
&= \frac{T^{1-\gamma}\left(r_g\cdot\alpha_{2g}^0\right)\left(1+o_p\left(1\right)\right)\sqrt{\frac{1}{\tau_g^2}\sum_{i\in G_g^0}\sum_{j=1}^{\tau_g}\widetilde{y}_{i,j-1}^2}}{\widetilde{\sigma}_g\left(r_g\right)} \\
&= \frac{T^{1-\gamma}\left(r_g\cdot\alpha_{2g}^0\right)\left(1+o_p\left(1\right)\right)\sqrt{\frac{n_gT^\gamma\overline{\rho}_{2g}^{2\left(\tau_g-\tau_g^e\right)}}{2\tau_g^2\alpha_{2g}^0}\mathbb{E}y_{i,\tau_g^e}^2}}{\sqrt{\frac{r_g^e\overline{\rho}_{2g}^{2\left(\tau_g-\tau_g^e\right)}}{2r_g\left(\alpha_{2g}^0\right)^2\tau_g^2}\mathbb{E}y_{i,\tau_g^e}^2\left(1+o_p\left(1\right)\right)}} \\
&= \frac{\sqrt{n_gT^\gamma\alpha_{2g}^0r_g}\,T^{1-\gamma}\left(r_g\cdot\alpha_{2g}^0\right)\left(1+o_p\left(1\right)\right)}{\sqrt{r_g^e}}=O\left(\sqrt{n}T^{1-\frac{\gamma}{2}}\right)=O\left(P_{Tn}\right).
\end{aligned}
$$

This argument shows that the proposed test diverges at the speed of $O_p\left(P_{Tn}\right)$ under the alternative hypothesis. Therefore the consistency of the bubble origination estimate is also verified. ■

A.2.11 Proof of Theorem 3.4.8

Proof. Based on Theorem 3.4.4 and Lemma 3.4.1, we have

$$
\begin{aligned}
\text{BIC}(K^0) &= \left(\widetilde{\sigma}_{\widehat{G}_g(K^0)}^2\right)+\frac{K^0+n}{nT}\log(nT) \\
&= \left[\frac{1}{nT}\sum_{g=1}^{K^0}\sum_{i\in\widehat{G}_g(K^0)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widehat{\overline{\rho}}_{\widehat{g}_i^{(K^0)}}^{(K^0)}\widetilde{y}_{i,t-1}\right)^2\right]+o(1) \\
&\to \sigma^2.
\end{aligned}
$$

Moreover, under the under-fitted model with $K^*<K^0$, note that

$$
\widetilde{\sigma}_{\widehat{G}(K^*)}^2 = \left[\frac{1}{nT}\sum_{g=1}^{K^*}\sum_{i\in\widehat{G}_g(K^*)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widehat{\overline{\rho}}_{\widehat{g}_i^{(K^*)}}^{(K^*)}\widetilde{y}_{i,t-1}\right)^2\right]
$$

$$
\begin{aligned}
&\geq \min_{1\leq K^*<K^0} \inf_{\delta(K^*)\in\Delta_{K^*}} \frac{1}{nT}\sum_{g=1}^{K^*}\sum_{i\in G_g(K^*)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widehat{\overline{\rho}}^{(K^*)}_{\widehat{g}_i^{(K^*)}}\widetilde{y}_{i,t-1}\right)^2 \\
&= \min_{1\leq K^*<K^0} \inf_{\delta(K^*)\in\Delta_{K^*}} \widetilde{\sigma}^2_{G(K^*)}.
\end{aligned}
$$

Under Assumptions 1 and 2, we have

$$
\begin{aligned}
\min_{1\leq K^*<K^0} \text{BIC}\left(K^*\right) &\geq \min_{1\leq K^*<K} \inf_{\delta(K^*)\in\Delta_{K^*}} \left(\widetilde{\sigma}^2_{G(K^*)}\right)+\frac{K^*+n}{nT}\log\left(nT\right) \\
\xrightarrow{p} \left(\underline{\sigma}^2\right) &> 2\left(\sigma^2\right),
\end{aligned}
$$

and it follows that

$$
\Pr\left(\min_{1\leq K^*<K^0} \text{BIC}\left(K^*\right) > \text{BIC}\left(K^0\right)\right) \to 1.
$$

Lastly, under the overfitted moel, and in this case $K^0 < K^* \leq K_{\max}$,

$$
\begin{aligned}
& \Pr\left(\min_{K^0<K^*\leq K_{\max}} \text{BIC}\left(K^*\right) > \text{BIC}\left(K^0\right)\right) \\
= \; & \Pr\left(\min_{K^0<K^*\leq K_{\max}} nT\left(\widetilde{\sigma}^2_{G(K^*)}-\widetilde{\sigma}^2_{G(K^0)}\right)+\left(K^*-K^0\right)\ln\left(nT\right)>0\right) \\
& \to 1,
\end{aligned}
$$

as $(n,T)\to\infty$. ■

A.3 Proofs in Chapter 4

A.3.1 Technical lemmas in short-horizon predictive regression

Lemma A.3.1 *As $T\to\infty$, then*

$$
\begin{aligned}
\sup_{1\leq t\leq T} \mathbb{E}\left\|\eta^{(1)}_{T,t-1}\right\| &= O_p(T^{\gamma+\frac{1}{2}}), \\
\sup_{1\leq t\leq T} \mathbb{E}\left\|\eta^{(2)}_{T,t-1}\right\| &= O_p(T^{\frac{1+\gamma}{2}}) \\
\sup_{1\leq t\leq T} \mathbb{E}\left\|\eta^{(3)}_{T,t-1}\right\| &= O_p(T^{\gamma+\frac{1}{2}})
\end{aligned}
$$

Proof. Let $x_0 = 0$, then $x_{j-1} = \sum_{k=1}^{j-2}(\prod_{m=k+1}^{j-1} R_{Tm})u_{xk} + u_{x,j-1}$, $\breve{D}_{at} :=$ $\text{diag}\left\{a_1' u_{at}, a_2' u_{at}, ..., a_n' u_{at}\right\}$, $\breve{D}_{B_a}(r) := \text{diag}\left\{a_1' B_a(r), a_2' B_a(r), ..., a_n' B_a(r)\right\}$ and denote the autocovariance matrix of u_{xt} and u_{at} as $\Gamma_{ux}(h) := \mathbb{E}(u_{xt}u_{x,t-h}^{'})$, and $\Gamma_{ua}(h) := \mathbb{E}(u_{at}u_{a,t-h}^{'})$.

(i) For $\eta_{T,t-1}^{(2)}$, following the decomposition of Eqn (42) in Phillips and Magdalinos (2009),

$$
\begin{aligned}
\left\|\eta_{T,t-1}^{(1)}\right\|^2 &= tr\left\{\sum_{j=1}^{t-1} R_{Tz}^{t-j-1} R_{Tz}^{t-i-1}(x_{j-1}x_{i-1}^{'})\right\} \\
&\leq \sum_{i,j=1}^{t-1}\sum_{k=1}^{j}\sum_{l=1}^{i}\left\|R_{Tz}^{2t-2-j-i} R_{Tt}^{(j,k)} R_{Tt}^{(i,l)}\right\|_F \left\|u_{xk}u_{xl}^{'}\right\|_F \\
&\leq \sqrt{n}\sum_{i,j=1}^{t-1}\sum_{k=1}^{j}\sum_{l=1}^{i}\left\|R_{Tz}\right\|^{2t-2-j-i}\left\|R_{Tt}^{(j,k)}\right\|_F\left\|R_{Tt}^{(i,l)}\right\|_F\left\|u_{xk}u_{xl}^{'}\right\|_F,
\end{aligned}
$$

where $R_{Tt}^{(j,k)} := \prod_{m=k+1}^{j} R_{Tm}$ and $R_{Tt}^{(i,l)} := \prod_{m=l+1}^{i} R_{Tm}$. By Phillips and Magdalinos (2009), $\sup_{1\leq t\leq T}\left\|R_{Tz}\right\|^{t-j} = O(T^{\gamma})$. By the definition of $R_{Tt-1}^{(t,j)}$,

$$
\begin{aligned}
\frac{1}{T}\sup_{1\leq t\leq T}\sum_{j=1}^{t}\left\|R_{Tt}^{(t,j)}\right\|_F &= \sup_{0\leq r\leq 1}\frac{1}{T}\sum_{j=1}^{[Tr]}\left\|R_{Tt}^{([Tr],j)}\right\|_F \\
&= \frac{1}{T}\sup_{0\leq r\leq 1}\sum_{j=1}^{[Tr]}\left\|\exp\left(\frac{[Tr]-j-1}{T}C + \frac{\sum_{m=j+1}^{[Tr]}\breve{D}_{am}}{\sqrt{T}}\right)\right\| \\
&= \frac{1}{T}\sup_{0\leq r\leq 1}\sum_{j=1}^{[Tr]}\left\|\exp\left(\frac{[Tr]-j-1}{T}C + \frac{\sum_{m=1}^{j}\breve{D}_{am}}{\sqrt{T}}\right)\right\|_F \\
&\quad \cdot\left\|\exp\left(\frac{\sum_{m=1}^{[Tr]}\breve{D}_{am}}{\sqrt{T}}\right)\right\|_F \\
&= \left\|\exp(B_a(r))\right\|_F\cdot\int_0^r\left\|\exp((1-p)C - \breve{D}_{B_a}(p))\right\|_F dp + o_p(1) < +\infty,
\end{aligned}
$$

for LSTUR and STUR cases. By the assumption $\sum_{l=-\infty}^{\infty}\left\|\Gamma_{ux}(l)\right\|_F < \infty$,

$$
\sup_{1\leq t\leq T}\left\|\eta_{T,t-1}^{(1)}\right\|^2 \leq \sqrt{n}\left(\sup_{1\leq t\leq T}\sum_{i=1}^{t}\left\|R_{Tz}\right\|^{t-i}\right)^2\left(\sup_{1\leq t\leq T}\sum_{k=1}^{t}\left\|T_{Tt}^{(t,k)}\right\|_F\right)\left(\sum_{l=-\infty}^{\infty}\left\|\Gamma_{ux}(l)\right\|_F\right) = O_p(T^{1+2\gamma}).
$$

(ii) For $\eta^{(2)}_{T,t-1}$,

$$\begin{aligned}\left\|\eta^{(2)}_{T,t-1}\right\|^2 &\le tr\left\{\sum_{i,j=1}^{t-1} R_{Tz}^{2t-2-j-i}\breve{D}_{ai}R_{Tt}^{(i,l)}x_{i-1}x_{j-1}^{'}R_{Tt}^{(j-k)}\breve{D}_{aj}\right\}\\ &\le \sqrt{n}\sum_{i,j=1}^{t-1}\sum_{k=1}^{j}\sum_{l=1}^{i}\left\|R_{Tz}\right\|^{2t-2-j-i}\left\|R_{Tt}^{(i,l)}\right\|_F\left\|R_{Tt}^{(j,k)}\right\|_F\left\|\breve{D}_{ai}\breve{D}_{aj}u_{xk}u_{xl}^{'}\right\|_F,\end{aligned}$$

since $\sup_{1\le t\le T}\sum_{j=1}^{t}\left\|R_{Tz}\right\|^{t-j} = O(T^{\gamma})$, and $\sup_{1\le t\le T}\sum_{j=1}^{t-1}\left\|R_{Tt}^{(t,j)}\right\|_F = O_p(T)$. Moreover, $\sum_{i=-\infty}^{\infty}\left\|\Gamma_{ua}(i)\right\|_F<\infty$ and $\sum_{k=-\infty}^{\infty}\left\|\Gamma_{ux}(k)\right\|_F<\infty$, then

$$\begin{aligned}\sup_{1\le t\le T}\left\|\eta^{(2)}_{T,t-1}\right\|^2 &\le \sqrt{n}\left(\sup_{1\le t\le T}\sum_{i=1}^{t-1}\left\|R_{Tz}\right\|^{t-i}\right)\left(\sum_{j=-\infty}^{\infty}\left\|\Gamma_{ua}(j)\right\|_F\right)\\ &\quad\cdot\left(\sup_{1\le t\le T}\sum_{k=1}^{t-1}\left\|R_{Tt}^{(t,k)}\right\|_F\right)\cdot\left(\sum_{l=-\infty}^{\infty}\left\|\Gamma_{ux}(l)\right\|_F\right)\\ &= O(T^{1+\gamma}),\end{aligned}$$

due to the orthogonal assumptions imposed for innovations.

(iii) The proof of this case follows (i). ■

Lemma A.3.2 *As* $T\to\infty$,

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

(ii) $\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\widetilde{z}_{t-1}^{'} = \frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}z_{t-1}z_{t-1}^{'}+\frac{1}{T^{\frac{3}{2}+\gamma}}(\sum_{t=1}^{T}z_{t-1}(\eta^{(2)}_{T,t-1})^{'}+\sum_{t=1}^{T}\eta^{(2)}_{T,t-1}z_{t-1}^{'})+\frac{1}{T^{2+\gamma}}\sum_{t=1}^{T}\eta^{(2)}_{T,t-1}(\eta^{(2)}_{T,t-1})^{'}+o_p(1)$,

(iii) $\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\widetilde{z}_{t-1}x_{t-1}^{'} = \frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}z_{t-1}x_{t-1}^{'}+\frac{1}{T^{\frac{3}{2}+\gamma}}\sum_{t=1}^{T}\eta^{(2)}_{T,t-1}x_{t-1}^{'}+\frac{C}{T^{2+\gamma}}\sum_{t=1}^{T}\eta^{(1)}_{T,t-1}x_{t-1}^{'}+\frac{1}{T^{2+\gamma}}\sum_{t=1}^{T}\eta^{(3)}_{T,t-1}x_{t-1}^{'}$.

Proof. For part (i), note the fact that

$$\sum_{t=1}^{T}z_{t-1}u_{0t}=O_p(T^{\frac{1+\gamma}{2}}),\ \sum_{t=1}^{T}\eta^{(2)}_{t}u_{0t}=O_p(T^{\frac{2+\gamma}{2}}),\ \sum_{t=1}^{T}\eta^{(1)}_{t-1}u_{0t}=O_p(T^{1+\gamma}),\sum_{t=1}^{T}\eta^{(3)}_{t-1}u_{0t}=O_p(T^{1+\gamma}).$$

Therefore $\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}u_{0t}=\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}z_{t-1}u_{0t}+\frac{1}{T^{1+\frac{\gamma}{2}}}\sum_{t=1}^{T}\eta_{T,t-1}^{(2)}u_{0t}+o_p(1)$.

The proof of (ii) is a natural extension of (i). In (iii), define $R_T:=1+\frac{C}{T}$, then

$$\begin{aligned}x_tz_t^{'} &= R_Tx_{t-1}z_{t-1}^{'}R_{Tz}+R_Tx_{t-1}u_{xt}^{'}+u_{xt}z_{t-1}^{'}R_{Tz}+u_{xt}u_{xt}^{'}+\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}z_{t-1}^{'}R_{Tz}+\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}u_{xt}^{'}\\ &\quad +\frac{\breve{D}_{at}^2}{T}x_{t-1}z_{t-1}^{'}R_{Tz}+\frac{\breve{D}_{at}^2}{T}x_{t-1}u_{xt}^{'}.\end{aligned}$$

Note the following facts that $\sum_{t=1}^{T}\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}z_{t-1}^{'}\preceq O_p(T^{\frac{1+\gamma}{2}})$, $\sum_{t=1}^{T}\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}u_{xt}^{'}=O_p(T)$, $\sum_{t=1}^{T}u_{xt}z_{t-1}^{'}=O_p(T)$, and

$$\begin{aligned}&[I_{n\times n}-R_{Tz}\otimes R_T]\sum_{t=1}^{T}(x_{t-1}\otimes z_{t-1}) \qquad \text{(A.3.1)}\\ =\;& x_0\otimes z_0-x_T\otimes z_T+(I_{n\times n}\otimes R_T)\sum_{t=1}^{T}x_{t-1}\otimes(u_{xt})+(R_{Tz}\otimes I_n)\sum_{t=1}^{T}u_{xt}\otimes z_{t-1}\\ &+\sum_{t=1}^{T}u_{xt}\otimes u_{xt}+(R_{Tz}\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes z_{t-1}+(I_n\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes u_{xt}\\ &+(R_{Tz}\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}^2}{T}x_{t-1}\right)\otimes z_{t-1}+(I_n\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}^2}{T}x_{t-1}\right)\otimes u_{xt}.\end{aligned}$$

In (A.3.1), $\sup_{1\leq t\leq T}\|x_{t-1}\|=O_p(\sqrt{T})$, $\sum_{t=1}^{T}x_{t-1}u_{xt}^{'}=O_p(T)$ and $\sum_{t=1}^{T}z_{t-1}u_{xt}^{'}=O_p(T)$, $\sum_{t=1}^{T}u_{xt}u_{xt}^{'}=O_p(T)$, $\sum_{t=1}^{T}x_{t-1}\otimes(u_{xt})=O_p(T)$, $\sum_{t=1}^{T}u_{xt}\otimes z_{t-1}=O_p(T)$, $\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes z_{t-1}\preceq O_p(T^{\frac{1+\gamma}{2}})$, $\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes u_{xt}=O_p(T)$.
Based on Cauchy-Schwarz inequality, $\sum_{t=1}^{T}\frac{\breve{D}_{at}^2}{T}x_{t-1}\otimes z_{t-1}\preceq O_p(T^{\frac{1+\gamma}{2}})$, and $\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}^2}{T}x_{t-1}\right)\otimes u_{xt}\preceq O_p(1)$. When $\gamma\in(0,1)$,

$$I_{n\times n}-R_{Tz}\otimes R_T=\frac{1}{T^{\gamma}}\left(C_z\otimes I_n\right)\left[I_n+O_p\left(\frac{1}{T^{1-\gamma}}\right)\right].$$

Then we have $\sum_{t=1}^{T}x_{t-1}z_{t-1}^{'}=O_p(T^{1+\gamma})$.

In order to justify $\sum_{t=1}^{T}x_{t-1}(\eta_{T,t-1}^{(2)})^{'}=O_p(T^{\frac{3}{2}+\gamma})$, and define $R_T:=1+\frac{C}{T}$.
Note the fact that

$$x_t(\eta_{T,t}^{(2)})^{'} = R_Tx_{t-1}(\eta_{T,t-1}^{(2)})^{'}R_{Tz}+R_Tx_{t-1}(\breve{D}_{at}x_{t-1})^{'}+u_{xt}(\eta_{T,t-1}^{(2)})^{'}R_{Tz}+u_{xt}(\breve{D}_{at}x_{t-1})^{'}$$

$$+\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}(\eta^{(2)}_{T,t-1})'R_{Tz}+\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}(\breve{D}_{at}x_{t-1})'$$
$$+\frac{\breve{D}^2_{at}}{T}x_{t-1}(\eta^{(2)}_{T,t-1})'R_{Tz}+\frac{\breve{D}^2_{at}}{T}x_{t-1}(\breve{D}_{at}x_{t-1})'.$$

Therefore,

$$\begin{aligned}
& [I_{n\times n}-R_{Tz}\otimes R_T]\sum_{t=1}^{T}(x_{t-1}\otimes\eta^{(2)}_{T,t-1}) \qquad \text{(A.3.2)}\\
= \; & x_0\otimes\eta^{(2)}_{T,0}-x_T\otimes\eta^{(2)}_{T,T}+(I_n\otimes R_T)\sum_{t=1}^{T}x_{t-1}\otimes(\breve{D}_{at}x_{t-1})+(R_{Tz}\otimes I_n)\sum_{t=1}^{T}u_{xt}\otimes\eta^{(2)}_{T,t-1}\\
& +(I_n\otimes I_n)\sum_{t=1}^{T}u_{xt}\otimes(\breve{D}_{at}x_{t-1})+(R_{Tz}\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes\eta^{(2)}_{T,t-1}\\
& +(I_n\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes(\breve{D}_{at}x_{t-1})+(R_{Tz}\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}^2_{at}}{T}x_{t-1}\right)\otimes\eta^{(2)}_{T,t-1}\\
& +(I_n\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}^2_{at}}{T}x_{t-1}\right)\otimes(\breve{D}_{at}x_{t-1}).
\end{aligned}$$

Note the fact that $x_0(\eta^{(2)}_{T,0})'=O_p(1)$, $x_T(\eta^{(2)}_{T,T})'=O_p(T^{1+\frac{\gamma}{2}})$, $\sum_{t=1}^{T}x_{t-1}(\breve{D}_{at}x_{t-1})'=\sum_{t=1}^{T}x_{t-1}x'_{t-1}\breve{D}_{at}=O_p(T^{\frac{3}{2}})$, $\sum_{t=1}^{T}u_{xt}\otimes\eta^{(2)}_{T,t-1}=O_p(T^{\frac{2+\gamma}{2}})$, $\sum_{t=1}^{T}u_{xt}(\breve{D}_{at}x_{t-1})'=O_p(T^{\frac{3}{2}})$, $\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)(\eta^{(2)}_{T,t-1})'\preceq O_p(T^{1+\frac{\gamma}{2}})$, $\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)(\breve{D}_{at}x'_{t-1})'=\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)(x'_{t-1}\breve{D}_{at})=O_p(T^{\frac{3}{2}})$. By Cauchy-Schwarz inequality, $\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{T}x_{t-1}\right)(\eta^{(2)}_{T,t-1})'\preceq O_p(T^{\frac{1+\gamma}{2}})$ and $\sum_{t=1}^{T}\left(\frac{\breve{D}^2_{at}}{T}x_{t-1}\right)(\breve{D}_{at}x_{t-1})'\preceq O_p(T^{\frac{1}{2}})$. By combining above results,

$$I_{n\times n}-R_{Tz}\times R_T=\frac{1}{T^{\gamma}}\left[-C_z\otimes I_{n\times n}+O_p\left(\frac{1}{T^{1-\gamma}}\right)\right],$$

and

$$\begin{aligned}
& \frac{1}{T^{\frac{3}{2}+\gamma}}\sum_{t=1}^{T}\left[x_{t-1}\otimes\eta^{(2)}_{T,t-1}\right]\\
= \; & -(C_z^{-1}\otimes I_n)\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}x_{t-1}\otimes(\breve{D}_{at}x_{t-1})-(C_z^{-1}\otimes I_n)\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}u_{xt}\otimes(\breve{D}_{at}x_{t-1})\\
& -(C_z^{-1}\otimes I_n)\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes(\breve{D}_{at}x_{t-1})+o_p(1)\\
= \; & O_p(1).
\end{aligned}$$

In order to justify $\sum_{t=1}^{T} x_{t-1}(\eta_{T,t-1}^{(1)})' = O_p(T^{2+\gamma})$, define $R_T := 1 + \frac{C}{T}$. Note the following fact that

$$\begin{aligned}
x_t(\eta_{T,t}^{(1)})' &= R_T x_{t-1}(\eta_{T,t-1}^{(1)})' R_{Tz} + R_T x_{t-1}\left[x_{t-1}\right]' + u_{xt}(\eta_{T,t-1}^{(1)})' R_{Tz} + u_{xt}(x_{t-1})' \\
&+ \frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1}(\eta_{T,t-1}^{(1)})' R_{Tz} + \frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1}(x_{t-1})' \\
&+ \frac{\breve{D}_{at}^2}{T} x_{t-1}(\eta_{T,t-1}^{(1)})' R_{Tz} + \frac{\breve{D}_{at}^2}{T} x_{t-1}(x_{t-1})',
\end{aligned}$$

thus

$$\begin{aligned}
& [I_{n\times n} - R_{Tz} \otimes R_T] \sum_{t=1}^{T} (x_{t-1} \otimes \eta_{T,t-1}^{(1)}) \qquad \text{(A.3.3)} \\
= \; & x_0 \otimes \eta_{T,0}^{(1)} - x_T \otimes \eta_{T,T}^{(1)} + (I_n \otimes R_T) \sum_{t=1}^{T} x_{t-1} \otimes [x_{t-1}] + (R_{Tz} \otimes I_n) \sum_{t=1}^{T} u_{xt} \otimes \eta_{T,t-1}^{(1)} \\
& + (I_{n\times n} \otimes I_n) \sum_{t=1}^{T} u_{xt} \otimes (x_{t-1}) + (R_{Tz} \otimes I_n) \sum_{t=1}^{T} \left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1} \right) \otimes \eta_{T,t-1}^{(1)} \\
& + (I_n \otimes I_n) \sum_{t=1}^{T} \left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1} \right) \otimes (x_{t-1}) + (R_{Tz} \otimes I_n) \sum_{t=1}^{T} \left(\frac{\breve{D}_{at}^2}{T} x_{t-1} \right) \otimes \eta_{T,t-1}^{(1)} \\
& + (I_n \otimes I_n) \sum_{t=1}^{T} \left(\frac{\breve{D}_{at}^2}{T} x_{t-1} \right) \otimes (x_{t-1}).
\end{aligned}$$

Here we have $x_0(\eta_{T,0}^{(1)})' = O_p(1)$ and $x_T(\eta_{T,T}^{(1)})' = O_p(T^{1+\gamma})$, $\sum_{t=1}^{T} x_{t-1}(x_{t-1})' = O_p(T^2)$, $\sum_{t=1}^{T} u_{xt}(\eta_{T,t-1}^{(1)})' = O_p(T^{\frac{2+2\gamma}{2}})$, $\sum_{t=1}^{T} u_{xt}(x_{t-1})' = O_p(T)$, $\sum_{t=1}^{T} \left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1} \right) (\eta_{T,t-1}^{(1)})' \preceq O_p(T^{1+\gamma})$, $\sum_{t=1}^{T} \left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1} \right) (x_{t-1})' = \sum_{t=1}^{T} \left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1} \right) x_{t-1}' = O_p(T)$. By Cauchy-Schwarz inequality, then $\sum_{t=1}^{T} \left(\frac{\breve{D}_{at}^2}{T} x_{t-1} \right) (\eta_{T,t-1}^{(1)})' \preceq O_p(T^{1+\gamma})$ and $\sum_{t=1}^{T} \left(\frac{\breve{D}_{at}^2}{T} x_{t-1} \right) (x_{t-1})' \preceq O_p(T)$. We have

$$I_{n\times n} - R_{Tz} \times R_T = \frac{1}{T^{\gamma}} \left[-C_z \otimes I_{n\times n} + O_p\left(\frac{1}{T^{1-\gamma}} \right) \right].$$

By combining the above results,

$$\frac{1}{T^{2+\gamma}} \sum_{t=1}^{T} \left[x_{t-1} \otimes \eta_{T,t-1}^{(1)} \right] = -(C_z^{-1} \otimes I_{n\times n}) \frac{1}{T^2} \sum_{t=1}^{T} x_{t-1} \otimes x_{t-1} + o_p(1) = O_p(1).$$

Last, in order to justify $\sum_{t=1}^{T} x_{t-1}(\eta_{T,t-1}^{(3)})' = O_p(T^{2+\gamma})$, we apply the following decomposition as

$$\begin{aligned} x_t(\eta_{T,t}^{(3)})' &= R_T x_{t-1}(\eta_{T,t-1}^{(3)})' R_{Tz} + R_T x_{t-1}(\breve{D}_{at}^2 x_{t-1}') + u_{xt}(\eta_{T,t-1}^{(3)})' R_{Tz} + u_{xt}(x_{t-1})' \\ &\quad + \left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1}\right)(\eta_{T,t-1}^{(3)})' R_{Tz} + \left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1}\right)\left[\breve{D}_{at}^2 x_{t-1}\right]' \\ &\quad + \left(\frac{\breve{D}_{at}^2}{T} x_{t-1}\right)(\eta_{T,t-1}^{(3)})' R_{Tz} + \left(\frac{\breve{D}_{at}^2}{T} x_{t-1}\right)\left[\breve{D}_{at}^2 x_{t-1}\right]'. \end{aligned}$$

Therefore,

$$\begin{aligned} &[I_{n\times n} - R_{Tz} \otimes R_T] \sum_{t=1}^{T} (x_{t-1} \otimes \eta_{T,t-1}^{(3)}) \qquad \text{(A.3.4)} \\ &= x_0 \otimes \eta_{T,0}^{(3)} - x_T \otimes \eta_{T,T}^{(3)} + (I_n \otimes R_T)\sum_{t=1}^{T} x_{t-1} \otimes (\breve{D}_{at}^2 x_{t-1}) + (R_{Tz} \otimes I_n)\sum_{t=1}^{T} u_{xt} \otimes \eta_{T,t-1}^{(3)} \\ &\quad + (I_n \otimes I_n)\sum_{t=1}^{T} u_{xt} \otimes (\breve{D}_{at}^2 x_{t-1}) + (R_{Tz} \otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1}\right) \otimes \eta_{T,t-1}^{(3)} \\ &\quad + (I_n \otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1}\right) \otimes \left[\breve{D}_{at}^2 x_{t-1}\right] + (R_{Tz} \otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}^2}{T} x_{t-1}\right) \otimes \eta_{T,t-1}^{(3)} \\ &\quad + (I_n \otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}^2}{T} x_{t-1}\right) \otimes \left[\breve{D}_{at}^2 x_{t-1}\right]. \end{aligned}$$

Here $x_0(\eta_{T,0}^{(3)})' = O_p(1)$ and $x_T(\eta_{T,T}^{(3)})' = O_p(T^{1+\gamma})$, $\sum_{t=1}^{T} x_{t-1} \otimes (\breve{D}_{at}^2 x_{t-1}) = O_p(T^2)$, $\sum_{t=1}^{T} u_{xt} \otimes \eta_{T,t-1}^{(3)} = O_p(T^{\frac{2+2\gamma}{2}})$, $\sum_{t=1}^{T} u_{xt} \otimes (\breve{D}_{at}^2 x_{t-1}) = O_p(T)$, $\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1}\right)(\eta_{T,t-1}^{(1)})' \preceq O_p(T^{1+\gamma})$, $\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1}\right) \otimes (\breve{D}_{at}^2 x_{t-1}) = O_p(T)$. By Cauchy-Schwarz inequality, $\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}^2}{T} x_{t-1}\right) \otimes \eta_{T,t-1}^{(3)} \preceq O_p(T^{1+\gamma})$ and $\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}^2}{T} x_{t-1}\right) \otimes \left[\breve{D}_{at}^2 x_{t-1}\right] \preceq O_p(T)$.
Besides, we have

$$I_{n\times n} - R_{Tz} \times R_T = \frac{1}{T^{\gamma}}\left[-C_z \otimes I_{n\times n} + O_p\left(\frac{1}{T^{1-\gamma}}\right)\right].$$

By combining the above results,

$$\frac{1}{T^{2+\gamma}}\sum_{t=1}^{T}\left[x_{t-1} \otimes \eta_{T,t-1}^{(3)}\right] = -(C_z^{-1} \otimes I_{n\times n})\frac{1}{T^2}\sum_{t=1}^{T} x_{t-1} \otimes \left(\breve{D}_{at}^2 x_{t-1}\right) + o_p(1) = O_p(1).$$

■

Lemma A.3.3 *(i) As $T \to \infty$,*

$$
\begin{aligned}
\frac{1}{T^{1+\gamma}} \sum_{t=1}^{T} z_{t-1} z_{t-1}^{'} &\rightsquigarrow \int_0^{\infty} e^{C_z r} \Omega_{xx} e^{C_z r} dr, \\
\frac{1}{T^{2+\gamma}} \sum_{t=1}^{T} \eta_{T,t-1}^{(2)} (\eta_{T,t-1}^{(2)})^{'} &\rightsquigarrow \begin{cases} \int_0^{\infty} e^{C_z s} \left[\int_0^1 G_{a,c}(r) G_{a,c}^{'}(r) \overline{\Sigma}_{aa} dr \right] e^{C_z s} ds, & \text{under LSTUR,} \\ \int_0^{\infty} e^{C_z s} \left[\int_0^1 G_a(r) G_a^{'}(r) \overline{\Sigma}_{aa} dr \right] e^{C_z s} ds, & \text{under STUR,} \end{cases} \\
\frac{1}{T^{\frac{3}{2}+\gamma}} \sum_{t=1}^{T} z_{t-1} (\eta_{T,t-1}^{(2)})^{'} &\rightsquigarrow 0,
\end{aligned}
$$

(ii) As $T \to \infty$,

$$
\begin{aligned}
\frac{1}{T^{1+\gamma}} \sum_{t=1}^{T} x_{t-1} z_{t-1}^{'} &\rightsquigarrow \begin{Bmatrix} -(\int_0^1 G_{a,c}(r) dB_x^{'}(r) + \Omega_{xx}) \cdot C_z^{-1}, & \text{under LSTUR,} \\ -(\int_0^1 G_a(r) dB_x^{'}(r) + \Omega_{xx}) \cdot C_z^{-1}, & \text{under STUR,} \end{Bmatrix} \\
\frac{1}{T^{\frac{3}{2}+\gamma}} \sum_{t=1}^{T} x_{t-1} (\eta_{T,t-1}^{(2)})^{'} &\rightsquigarrow \begin{Bmatrix} \begin{matrix} -\int_0^1 G_{a,c}(r) G_{a,c}^{'}(r) d\breve{D}_{B_a}(r) \cdot C_z^{-1} \\ -\left[\int_0^1 G_{a,c}(r) G_{a,c}^{'}(r) \overline{\Omega}_{aa} dr \right] C_z^{-1}, \end{matrix} & \text{under LSTUR,} \\ \begin{matrix} -\int_0^1 G_a(r) G_a^{'}(r) d\breve{D}_{B_a}(r) \cdot C_z^{-1} \\ -\left[\int_0^1 G_a(r) G_a^{'}(r) \overline{\Omega}_{aa} dr \right] C_z^{-1}, \end{matrix} & \text{under STUR,} \end{Bmatrix} \\
\frac{1}{T^{2+\gamma}} \sum_{t=1}^{T} x_{t-1} (\eta_{T,t-1}^{(1)})^{'} &\rightsquigarrow \begin{Bmatrix} -\int_0^1 G_{a,c}(r) G_{a,c}^{'}(r) dr \cdot C_z^{-1}, & \text{under LSTUR,} \\ -\int_0^1 G_a(r) G_a^{'}(r) dr \cdot C_z^{-1}, & \text{under STUR,} \end{Bmatrix} \\
\frac{1}{T^{2+\gamma}} \sum_{t=1}^{T} x_{t-1} (\eta_{T,t-1}^{(3)})^{'} &\rightsquigarrow \begin{Bmatrix} -\int_0^1 G_{a,c}(r) G_{a,c}^{'}(r) \overline{\Sigma}_{aa} dr \cdot C_z^{-1}, & \text{under LSTUR,} \\ -\int_0^1 G_a(r) G_a^{'}(r) \overline{\Sigma}_{aa} dr \cdot C_z^{-1}, & \text{under STUR,} \end{Bmatrix}
\end{aligned}
$$

where $d\breve{D}_{B_a}(r) := \operatorname{diag}\{a_1^{'} dB_a(r), a_2^{'} dB_a(r), ..., a_n^{'} dB_a(r)\}$.

Proof. (i) For the term $\sum_{t=1}^{T} z_{t-1} z_{t-1}^{'}$ follows the decomposition as

$$
\begin{aligned}
[I_{n \times n} - R_{Tz} \otimes R_{Tz}] \sum_{t=1}^{T} z_{t-1} \otimes z_{t-1} &= z_0 \otimes z_0 - z_T \otimes z_T + (I_n \otimes R_{Tz}) \sum_{t=1}^{T} u_{xt} \otimes z_{t-1} \\
&\quad + (R_{Tz} \otimes I_n) \sum_{t=1}^{T} z_{t-1} \otimes u_{xt} + \sum_{t=1}^{T} u_{xt} \otimes u_{xt},
\end{aligned}
$$

where $z_0 \otimes z_0 = O_p(1)$, $z_T \otimes z_T = O_p(T^{\gamma})$, $\sum_{t=1}^{T} z_{t-1} \otimes u_{xt} = O_p(T)$, $\sum_{t=1}^{T} u_{xt} \otimes z_{t-1} = O_p(T)$. Besides, $\sum_{t=1}^{T} u_{xt} \otimes u_{xt} = O_p(T)$, thus

$$\begin{aligned} & -\left(\frac{C_z}{T^{\gamma}} \otimes I_n + I_n \otimes \frac{C_z}{T^{\gamma}} + O\left(\frac{1}{T^{2\gamma}}\right)\right) \frac{1}{T} \sum_{t=1}^{T} z_{t-1} \otimes z_{t-1} \\ = & \ \frac{1}{T} \sum_{t=1}^{T} \left(u_{xt} \otimes u_{xt} + z_{t-1} \otimes u_{xt} + (u_{xt} \otimes z_{t-1})\right) + o_p(1). \end{aligned}$$

The above equation is a Lyapunov equation ($(I_n \otimes A + A^{'} \otimes I_n) vec(X) = -vec(Q)$: If A is stable, the solution of X is given by $X = \int_0^{\infty} e^{A\tau} Q e^{A^{'}\tau} d\tau$). Note the fact that

$$\frac{1}{T} \sum_{t=1}^{T} z_{t-1} u_{xt} \rightsquigarrow \Lambda_{xx},$$

where $\Lambda_{xx} := \sum_{h=1}^{T} \mathbb{E}\left(z_t z_{t-h}^{'}\right)$ and $\Omega_{xx} = \Sigma_{xx} + \Lambda_{xx} + \Lambda_{xx}^{'}$. Hence

$$\frac{1}{T^{1+\gamma}} \sum_{t=1}^{T} z_{t-1} z_{t-1}^{'} \rightsquigarrow \int_0^{\infty} e^{C_z r} \Omega_{xx} e^{C_z r} dr.$$

For the term $\sum_{t=1}^{T} \eta_{T,t-1}^{(2)} (\eta_{T,t-1}^{(2)})^{'}$, the following decomposition applies as

$$\begin{aligned} \eta_{T,t}^{(2)} (\eta_{T,t}^{(2)})^{'} = & \ R_{Tz} \eta_{T,t-1}^{(2)} (\eta_{T,t-1}^{(2)})^{'} R_{Tz} + R_{Tz} \eta_{T,t-1}^{(2)} (\breve{D}_{at} x_{t-1})^{'} \\ & + (\breve{D}_{at} x_{t-1}) (\eta_{T,t-1}^{(2)})^{'} R_{Tz} + (\breve{D}_{at} x_{t-1}) x_{t-1}^{'} \breve{D}_{at}. \end{aligned}$$

Therefore

$$\begin{aligned} & [I_{n \times n} - R_{Tz} \otimes R_{Tz}] \sum_{t=1}^{T} \eta_{T,t-1}^{(2)} \otimes \eta_{T,t-1}^{(2)} \\ = & \ \eta_{T,0}^{(2)} \otimes \eta_{T,0}^{(2)} - \eta_{T,T}^{(2)} \otimes \eta_{T,T}^{(2)} + (I_n \otimes R_{Tz}) \sum_{t=1}^{T} \eta_{T,t-1}^{(2)} \otimes (\breve{D}_{at} x_{t-1}) \\ & + (R_{Tz} \otimes I_n) \sum_{t=1}^{T} (\breve{D}_{at} x_{t-1}) \otimes \eta_{T,t-1}^{(2)} + \sum_{t=1}^{T} (\breve{D}_{at} x_{t-1}) \otimes (\breve{D}_{at} x_{t-1}), \end{aligned}$$

where $\eta_{T,0}^{(2)} \otimes \eta_{T,0}^{(2)} = O_p(1)$, $\eta_{T,T}^{(2)} \otimes \eta_{T,T}^{(2)} = O_p(T^{1+\gamma})$, $\sum_{t=1}^{T} \eta_{T,t-1}^{(2)} \otimes (\breve{D}_{at} x_{t-1}) \preceq O_p(T^{\frac{3+\gamma}{2}})$, $\sum_{t=1}^{T} (\breve{D}_{at} x_{t-1}) \otimes \eta_{T,t-1}^{(2)} \preceq O_p(T^{\frac{3+\gamma}{2}})$, $\sum_{t=1}^{T} (\breve{D}_{at} x_{t-1}) \otimes (\breve{D}_{at} x_{t-1}) =$

$O_p(T^2)$. Therefore,

$$-\left(\frac{C_z}{T^\gamma}\otimes I_n + I_n\otimes\frac{C_z}{T^\gamma}+O\left(\frac{1}{T^{2\gamma}}\right)\right)\frac{1}{T^2}\sum_{t=1}^{T}\eta_{t-1}^{(2)}\otimes\eta_{T,t-1}^{(2)}=\frac{1}{T^2}\sum_{t=1}^{T}(\breve{D}_{at}x_{t-1})\otimes(\breve{D}_{at}x_{t-1})+o_p(1),$$

and

$$\frac{1}{T^{2+\gamma}}\sum_{t=1}^{T}\eta_{T,t-1}^{(2)}(\eta_{T,t-1}^{(2)})' \rightsquigarrow \begin{cases}\int_0^\infty e^{C_z s}\left[\int_0^1 G_{a,c}(r)G_{a,c}'(r)\overline{\Sigma}_{aa}dr\right]e^{C_z s}ds & \text{under LSTUR,}\\ \int_0^\infty e^{C_z s}\left[\int_0^1 G_{a}(r)G_{a}'(r)\overline{\Sigma}_{aa}dr\right]e^{C_z s}ds & \text{under STUR.}\end{cases}$$

For the term $\sum_{t=1}^{T} z_{t-1}(\eta_{T,t-1}^{(2)})'$, the following decomposition applies as,

$$z_t(\eta_{T,t}^{(2)})' = R_{Tz}z_{t-1}(\eta_{T,t-1}^{(2)})'R_{Tz}+R_{Tz}z_{t-1}(\breve{D}_{at}x_{t-1})'+u_{xt}(\eta_{T,t-1}^{(2)})'R_{Tz}+u_{xt}x_{t-1}'\breve{D}_{at}.$$

Therefore

$$\begin{aligned}&[I_{n\times n}-R_{Tz}\otimes R_{Tz}]\sum_{t=1}^{T}z_{t-1}\otimes\eta_{T,t-1}^{(2)}\\ = \ & z_0\otimes\eta_{T,0}^{(2)} - z_T\otimes\eta_{T,T}^{(2)} + (I_n\otimes R_{Tz})\sum_{t=1}^{T}z_{t-1}\otimes(\breve{D}_{at}x_{t-1})\\ &+(R_{Tz}\otimes I_n)\sum_{t=1}^{T}u_{xt}\otimes\eta_{T,t-1}^{(2)}+\sum_{t=1}^{T}u_{xt}\otimes(\breve{D}_{at}x_{t-1}),\end{aligned}$$

where $z_0\otimes\eta_{T,0}^{(2)} = O_p(1)$, $z_T\otimes\eta_{T,T}^{(2)} = O_p(T^{\frac{1+2\gamma}{2}})$, $\sum_{t=1}^{T}z_{t-1}\otimes(\breve{D}_{at}x_{t-1}) \preceq O_p(T^{\frac{2+\gamma}{2}})$, $\sum_{t=1}^{T}u_{xt}\otimes\eta_{T,t-1}^{(2)} \preceq O_p(T^{\frac{2+\gamma}{2}})$. Besides, $\sum_{t=1}^{T}u_{xt}\otimes(\breve{D}_{at}x_{t-1}) = O_p(T^{\frac{3}{2}})$. Therefore,

$$-\left(\frac{C_z}{T^\gamma}\otimes I_n + I_n\otimes\frac{C_z}{T^\gamma}+O\left(\frac{1}{T^{2\gamma}}\right)\right)\frac{1}{T^2}\sum_{t=1}^{T}z_{t-1}\otimes\eta_{T,t-1}^{(2)}=\frac{1}{T^2}\sum_{t=1}^{T}u_{xt}\otimes(x_{t-1}\breve{D}_{at})+o_p(1),$$

and under strict exogeneity condition where $\Sigma_{ax}^* := \mathbb{E}(\breve{D}_{at}u_{xt}) = 0$,

$$\frac{1}{T^{\frac{3}{2}+\gamma}}\sum_{t=1}^{T}z_{t-1}(\eta_{T,t-1}^{(2)})' \rightsquigarrow 0.$$

(ii) For $\sum_{t=1}^{T} x_{t-1}z_{t-1}^{'}$ term, based on (A.3.1) , we have the following decomposition.

$$
\begin{aligned}
& [I_{n\times n} - R_{Tz}\otimes R_T]\,\frac{1}{T}\sum_{t=1}^{T}(x_{t-1}\otimes z_{t-1}) && \text{(A.3.5)}\\
= \;& \frac{1}{T}x_0\otimes z_0 - \frac{1}{T}x_T\otimes z_T + (I_{n\times n}\otimes R_T)\frac{1}{T}\sum_{t=1}^{T}x_{t-1}\otimes u_{xt} + (R_{Tz}\otimes I_n)\frac{1}{T}\sum_{t=1}^{T}u_{xt}\otimes z_{t-1}\\
& +\frac{1}{T}\sum_{t=1}^{T}u_{xt}\otimes u_{xt}^{'} + (R_{Tz}\otimes I_n)\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes z_{t-1} + (I_n\otimes I_n)\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes u_{xt}\\
& +(R_{Tz}\otimes I_n)\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}^2}{T}x_{t-1}\right)\otimes z_{t-1} + (I_n\otimes I_n)\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}^2}{T}x_{t-1}\right)\otimes u_{xt} && \text{(A.3.6)}\\
= \;& \frac{1}{T}\sum_{t=1}^{T}x_{t-1}\otimes u_{xt} + (R_{Tz}\otimes I_n)\,\frac{1}{T}\sum_{t=1}^{T}u_{xt}\otimes z_{t-1} + \frac{1}{T}\sum_{t=1}^{T}u_{xt}\otimes u_{xt} && \text{(A.3.7)}\\
& +\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes u_{xt}^{'} + o_p(1).
\end{aligned}
$$

Therefore,

$$
\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t^{'} + \frac{1}{T}\sum_{t=1}^{T}z_{t-1}u_{xt}^{'} + \frac{1}{T}\sum_{t=1}^{T}u_{xt}u_{xt}^{'} \rightsquigarrow \int_0^1 G_{a,c}(r)dB_x^{'}(r) + \Omega_{xx}.
$$

By the exogeneity condition $\Sigma_{ax}^{*} := \mathbb{E}(\breve{D}_{at}u_{xt}) = 0$, then $\frac{1}{T}\sum_{t=1}^{T}\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}u_t^{'} \rightsquigarrow 0$. When $\gamma\in(0,1)$ case, $I_{n\times n} - R_{Tz}\otimes R_T = -\frac{1}{T^{\gamma}}(C_z\otimes I_{n\times n})\left[I_{n\times n} + O_p(\frac{1}{T^{1-\gamma}})\right]$. By combining the above intermediate results, we can derive the desirable results.

For $\sum_{t=1}^{T} x_{t-1}(\eta_{T,t-1}^{(2)})^{'}$ term, based on (A.3.2), we have the following decomposition:

$$
\begin{aligned}
& [I_{n\times n} - R_{Tz}\otimes R_T]\sum_{t=1}^{T}(x_{t-1}\otimes \eta_{T,t-1}^{(2)}) && \text{(A.3.8)}\\
= \;& x_0\otimes\eta_{T,0}^{(2)} - x_T\otimes\eta_{T,T}^{(2)} + (I_n\otimes R_T)\sum_{t=1}^{T}x_{t-1}\otimes(\breve{D}_{at}x_{t-1}) + (R_{Tz}\otimes I_n)\sum_{t=1}^{T}u_{xt}\otimes\eta_{T,t-1}^{(2)}\\
& +(I_n\otimes I_n)\sum_{t=1}^{T}u_{xt}\otimes(\breve{D}_{at}x_{t-1}) + (R_{Tz}\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes\eta_{T,t-1}^{(2)}\\
& +(I_n\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}}{\sqrt{T}}x_{t-1}\right)\otimes(\breve{D}_{at}x_{t-1}) + (R_{Tz}\otimes I_n)\sum_{t=1}^{T}\left(\frac{\breve{D}_{at}^2}{T}x_{t-1}\right)\otimes\eta_{T,t-1}^{(2)}
\end{aligned}
$$

$$+(I_n \otimes I_n) \sum_{t=1}^{T} \left(\frac{\breve{D}_{at}^2}{T} x_{t-1} \right) \otimes (\breve{D}_{at} x_{t-1}).$$

Here the three leading terms are $\sum_{t=1}^{T} x_{t-1} \otimes (\breve{D}_{at} x_{t-1})$, $(I_n \otimes I_n) \sum_{t=1}^{T} u_{xt} \otimes (\breve{D}_{at} x_{t-1})$ and $(I_n \otimes I_n) \sum_{t=1}^{T} \left(\frac{\breve{D}_{at}}{\sqrt{T}} x_{t-1} \right) \otimes (\breve{D}_{at} x_{t-1})$. By Lieberman and Phillips (2020), the following functional law applies here as

$$\frac{1}{n^{\frac{3}{2}}} \sum_t u_t Y_{t-1}^2 \rightsquigarrow \int_0^1 G_{a,c}^2(r) dB_u(r) + 2 \left(\Lambda_{uu} a \int_0^1 G_{a,c}^2(r) dr + \Lambda_{u\epsilon} G_{a,c}(r) dr \right), \tag{A.3.9}$$

where u_t and ϵ_t are the notations of Lieberman and Phillips (2017). By applying (A.3.9) and exdogeneity condition $\Sigma_{ax}^* := \mathbb{E}(\breve{D}_{at} u_{xt}) = 0$,

$$\frac{1}{T^{\frac{3}{2}+\gamma}} \sum_{t=1}^{T} x_{t-1} (\eta_{T,t-1}^{(2)})' \rightsquigarrow \left\{ \begin{array}{ll} -\int_0^1 G_{a,c}(r) G_{a,c}'(r) d\breve{D}_{B_a}(r) C_z^{-1} & \text{under LSTUR,} \\ -\left[\int_0^1 G_{a,c}(r) G_{a,c}'(r) \overline{\Omega}_{aa} dr \right] C_z^{-1} & \\ -\int_0^1 G_a(r) G_a'(r) d\breve{D}_{B_a}(r) C_z^{-1} & \text{under STUR.} \\ -\left[\int_0^1 G_a(r) G_a'(r) dr \overline{\Omega}_{aa} \right] C_z^{-1} & \end{array} \right\}$$

For the term of $\sum_{t=1}^{T} x_{t-1} (\eta_{T,t-1}^{(1)})'$, we have $I_{n\times n} - R_{Tz} \otimes R_T = \frac{1}{T^\gamma} \left[-C_z \otimes I_n + O_p(\frac{1}{T^{1-\gamma}}) \right]$. By combining above results, the following derivations are justified as,

$$\begin{aligned} \frac{1}{T^{2+\gamma}} \sum_{t=1}^{T} x_{t-1} (\eta_{T,t-1}^{(1)})' C &= -\frac{1}{T^2} \sum_{t=1}^{T} x_{t-1} x_{t-1}' \cdot C C_z^{-1} + o_p(1) \\ &\rightsquigarrow \begin{cases} -\int_0^1 G_{a,c}(r) G_{a,c}'(r) dr \cdot C C_z^{-1} & \text{under LSTUR,} \\ -\int_0^1 G_a(r) G_a'(r) dr \cdot C C_z^{-1} & \text{under STUR.} \end{cases} \end{aligned}$$

Similarly, as $\eta_{T,t-1}^{(1)}$ and $\eta_{T,t-1}^{(3)}$ share the identical stochastic order,

$$\begin{aligned} \frac{1}{T^{2+\gamma}} \sum_{t=1}^{T} x_{t-1} (\eta_{T,t-1}^{(3)})' &= -\frac{1}{T^2} \sum_{t=1}^{T} x_{t-1} x_{t-1}' \breve{D}_{at}^2 \cdot C_z^{-1} + o_p(1) \\ &\rightsquigarrow \begin{cases} -\int_0^1 G_{a,c}(r) G_{a,c}'(r) dr \overline{\Sigma}_{aa} C_z^{-1} & \text{under LSTUR,} \\ -\int_0^1 G_a(r) G_a'(r) dr \overline{\Sigma}_{aa} C_z^{-1} & \text{under STUR.} \end{cases} \end{aligned}$$

■

Lemma A.3.4 *(i) As $T \to \infty$,*

$$\begin{pmatrix} \frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} z_{t-1} u_{0t} \\ \frac{1}{T^{1+\frac{\gamma}{2}}} \sum_{t=1}^{T} \eta_{T,t-1}^{(2)} u_{0t} \end{pmatrix} \rightsquigarrow \mathcal{MN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \Omega_{00} \begin{pmatrix} V_{zz} & 0 \\ 0 & V_{\eta\eta}^{(2)} \end{pmatrix} \right),$$

where $V_{zz} := \int_0^1 e^{rC_z} \Omega_{xx} e^{rC_z} dr$,

$$V_{\eta\eta}^{(2)} := \begin{Bmatrix} \int_0^\infty e^{C_z s} \left[\int_0^1 G_{a,c}(r) G_{a,c}'(r) \overline{\Sigma}_{aa} dr \right] e^{C_z s} ds, & \text{under} & \text{LSTUR}, \\ \int_0^\infty e^{C_z s} \left[\int_0^1 G_{a,c}(r) G_{a,c}'(r) \overline{\Sigma}_{aa} dr \right] e^{C_z s} ds, & \text{under} & \text{STUR}, \end{Bmatrix}.$$

(ii) As $T \to \infty$, with $d\breve{D}_{B_a}(r) := \operatorname{diag}\{a_1' dB_a(r), a_2' dB_a(r), ..., a_n' dB_a(r)\}$,

$$\frac{1}{T^{1+\gamma}} \sum_{t=1}^{T} x_{t-1} \widetilde{z}_{t-1} \rightsquigarrow$$

$$V_{xz} := \begin{Bmatrix} \begin{array}{c} -(\Omega_{xx} + \int_0^1 G_{a,c}(r) dB_x'(r)) \cdot C_z^{-1} - \int_0^1 G_{a,c}(r) G_{a,c}'(r) d\breve{D}_{B_a}(r) \cdot C_z^{-1} \\ - \left[\int_0^1 G_{a,c}(r) G_{a,c}'(r) \overline{\Omega}_{aa} dr \right] C_z^{-1} - \int_0^1 G_{a,c}(r) G_{a,c}'(r) dr \cdot C_z^{-1} C \\ - \left[\int_0^1 G_{a,c}(r) G_{a,c}'(r) \overline{\Sigma}_{aa} dr \right] C_z^{-1} \end{array} & \text{under} \quad \text{LSTUR}, \\ \begin{array}{c} -(\Omega_{xx} + \int_0^1 G_a(r) dB_x'(r)) \cdot C_z^{-1} - \int_0^1 G_a(r) G_a'(r) d\breve{D}_{B_a}(r) \cdot C_z^{-1} \\ - \int_0^1 G_{a,c}(r) G_a'(r) \overline{\Omega}_{aa} dr C_z^{-1} - \left[\int_0^1 G_a(r) G_a'(r) \overline{\Sigma}_{aa} dr \right] C_z^{-1} \end{array} & \text{under} \quad \text{STUR}, \end{Bmatrix}$$

Proof. For (i), by Cramér-Wold device, in order to justify the joint convergence to normality, it is sufficient to show that

$$\alpha \left(\frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^{T} z_{t-1} u_{0t} \right) + \beta \left(\frac{1}{T^{\frac{2+\gamma}{2}}} \sum_{t=1}^{T} \eta_{T,t-1}^{(2)} u_{0t} \right) \rightsquigarrow \mathcal{MN}\left(0, \ \Omega_{00}\left(\alpha^2 V_{zz} + \beta^2 V_{\eta\eta}^{(2)}\right)\right),$$

for any real numbers α, and β. We rewrite $\alpha X_T + \beta Y_T = \sum_{t=1}^{T} \xi_{Tt}$, where the array is as $\xi_{Tt} := \frac{\alpha}{T^{\frac{1+\gamma}{2}}} z_{t-1} u_{0t} + \frac{\beta}{T^{\frac{2+\gamma}{2}}} \eta_{T,t-1}^{(2)} u_{0t}$. To satisfy the martingale CLT (White, 2014), the stability condition is shown by

$$\sum_{t=1}^{T}\mathbb{E}_{t-1}\left[\xi_{Tt}\xi_{Tt}^{'}\right]=\Omega_{00}\left[\frac{\alpha^2\sum_{t=1}^{T}z_{t-1}z_{t-1}^{'}}{T^{1+\gamma}}+\frac{\beta^2\sum_{t=1}^{T}\eta_{T,t-1}^{(2)}(\eta_{T,t-1}^{(2)})^{'}}{T^{2+\gamma}}\right]$$
$$+\Omega_{00}\alpha\beta\left[\sum_{t=1}^{T}z_{t-1}(\eta_{T,t-1}^{(2)})^{'}+\sum_{t=1}^{T}\eta_{T,t-1}^{(2)}(z_{t-1})^{'}\right]/T^{\frac{3+\gamma}{2}}$$
$$\rightsquigarrow\Omega_{00}\left[\alpha^2V_{zz}+\beta^2V_{\eta\eta}^{(2)}\right].$$

Secondly, the Lindeberg condition is also confirmed. The function $1(\cdot)$ is an indication function. For any $\epsilon>0$,

$$\sum_{t=1}^{T}\mathbb{E}_{t-1}\left[\|\xi_{Tt}\|^2 1(\|\xi_{Tt}\|>\epsilon)\right]\leq\sum_{t=1}^{T}\mathbb{E}_{t-1}\left[\|\xi_{Tt}\|^2 1\left(\left\|\frac{\alpha}{T^{\frac{1+\gamma}{2}}}z_{t-1}u_{0t}+\frac{\beta}{T^{1+\frac{\gamma}{2}}}\eta_{T,t-1}^{(2)}u_{0t}\right\|>\epsilon\right)\right]$$
$$\leq\sum_{t=1}^{T}\left[\left\|\frac{z_{t-1}z_{t-1}^{'}}{T^{1+\gamma}}+\frac{z_{t-1}\eta_{T,t-1}^{(2)'}+\eta_{T,t-1}^{(2)}z_{t-1}^{'}}{T^{\frac{3+\gamma}{2}}}+\frac{\eta_{T,t-1}^{(2)}\eta_{T,t-1}^{(2)'}}{T^{2+\gamma}}\right\|\right]$$
$$\cdot\mathbb{E}_{t-1}\left\{|u_{0t}|^2 1\left(\left\|\frac{\alpha z_{t-1}u_{0t}}{T^{\frac{1+\gamma}{2}}}+\frac{\beta\eta_{T,t-1}^{(2)}u_{0t}}{T^{1+\frac{\gamma}{2}}}\right\|>\epsilon\right)\right\}$$
$$\leq K\cdot\max_{1\leq t\leq T}\mathbb{E}_{t-1}\left\{|u_{ot}|^2\left[\begin{array}{c}1\left(\max_{1\leq t\leq T}\left\|\frac{\alpha^2}{T^{1+\gamma}}z_{t-1}z_{t-1}^{'}\right\||u_{0t}|^2>\frac{\epsilon^2}{2}\right)\\+1\left(\max_{1\leq t\leq T}\left\|\frac{\beta^2}{T^{2+\gamma}}\eta_{T,t-1}^{(2)}\eta_{T,t-1}^{(2)'}\right\||u_{0t}|^2>\frac{\epsilon^2}{2}\right)\end{array}\right]\right\}$$
$$\leq K\cdot\max_{1\leq t\leq T}\mathbb{E}_{t-1}\left\{|u_{ot}|^2\left[\begin{array}{c}1\left(\left(\max_{1\leq t\leq T}\left\|\frac{\alpha^2 z_{t-1}z_{t-1}^{'}}{T^{\gamma}}\right\|\right)|u_{0t}|^2>\frac{\epsilon^2T}{2}\right)\\+1\left(\left(\max_{1\leq t\leq T}\left\|\frac{\beta^2\eta_{T,t-1}^{(2)}\eta_{T,t-1}^{(2)'}}{T^{1+\gamma}}\right\|\right)|u_{0t}|^2>\frac{\epsilon^2T}{2}\right)\end{array}\right]\right\}$$
$$\leq K\cdot\max_{1\leq t\leq T}\mathbb{E}_{t-1}\left\{2|u_{0t}|^2\left[1\left(M\cdot|u_{ot}|^2>\frac{\epsilon^2T}{2}\right)\right]\right\}=o_p(1),$$

where M and K are two constants. Hence the Lindeberg condition is satisfied and martingale CLT shows the joint normality. ■

Lemma A.3.5 *(i) For the numerator of short-horizon IVX and QR-IVX, as $T\to\infty$,*

$$\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\Psi_\tau(u_{0t\tau})\rightsquigarrow M_\tau:=\mathcal{MN}\left(0,\tau(1-\tau)\left[V_{zz}+V_{\eta\eta}^{(2)}\right]\right),$$

for short-horizon QR-IVX, and

$$\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}u_{0t}\rightsquigarrow M:=\mathcal{MN}\left(0,\Omega_{00}\left[V_{zz}+V_{\eta\eta}^{(2)}\right]\right),$$

for short-horizon IVX, where $V_{zz}:=\int_0^1 e^{rC_z}\Omega_{xx}e^{rC_z}dr$*,*

$$V_{\eta\eta}^{(2)}:=\left\{\begin{array}{ll}\int_0^{\infty}e^{C_zs}\left[\int_0^1 G_{a,c}(r)G_{a,c}^{'}(r)\overline{\Sigma}_{aa}dr\right]e^{C_zs}ds & \text{under, LSTUR,}\\ \int_0^{\infty}e^{C_zs}\left[\int_0^1 G_{a,c}(r)G_{a,c}^{'}(r)\overline{\Sigma}_{aa}dr\right]e^{C_zs}ds, & \text{under STUR,}\end{array}\right\}.$$

(ii) As $T\rightarrow\infty$*,*

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\widetilde{z}_{t-1}(\widetilde{z}_{t-1})^{'}\rightsquigarrow(V_{zz}+V_{\eta\eta}^{(2)}),$$

where V_{zz}*,* $V_{z\eta}^{(2)}$ *are defined in (i).*

(iii) For the denominator,

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}p_{u0t\tau,t-1}(0)x_{t-1}\widetilde{z}_{t-1}^{'}$$

$$\rightsquigarrow\left\{\begin{array}{ll}\begin{array}{c}-p_{u0\tau}(0)(\Omega_{xx}+\int_0^1 G_{a,c}(r)dB_x^{'}(r))\cdot C_z^{-1}\\ -p_{u0\tau}(0)\int_0^1 G_{a,c}(r)G_{a,c}^{'}(r)dr\cdot C_z^{-1}C\\ -p_{u0\tau}(0)\int_0^1 G_{a,c}(r)G_{a,c}^{'}(r)d\check{D}_{B_a}(r)\cdot C_z^{-1}\\ -p_{u0\tau}(0)\int_0^1 G_{a,c}(r)G_{a,c}^{'}(r)\overline{\Omega}_{aa}drC_z^{-1}\\ -p_{u0\tau}(0)\int_0^1 G_{a,c}(r)G_{a,c}^{'}(r)\overline{\Sigma}_{aa}drC_z^{-1},\end{array} & \text{under LSTUR,}\\ \begin{array}{c}-p_{u0\tau}(0)(\Omega_{xx}+\int_0^1 G_a(r)dB_x^{'}(r))\cdot C_z^{-1}\\ -p_{u0\tau}(0)\int_0^1 G_a(r)G_a^{'}(r)d\check{D}_{B_a}(r)\cdot C_z^{-1}\\ -p_{u0\tau}(0)\int_0^1 G_a(r)G_a^{'}(r)\overline{\Omega}_{aa}drC_z^{-1}\\ -p_{u0\tau}(0)\int_0^1 G_a(r)G_a^{'}(r)\overline{\Sigma}_{aa}drC_z^{-1},\end{array} & \text{under STUR,}\end{array}\right\}$$

for short-horizon QR-IVX, and

$$\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}x_{t-1}\widetilde{z}_{t-1}^{'}$$

$$\rightsquigarrow \left\{ \begin{array}{ll} -(\Omega_{xx} + \int_0^1 G_{a,c}(r) dB_x'(r)) \cdot C_z^{-1} - \int_0^1 G_{a,c}(r) G_{a,c}'(r) d\breve{D}_{B_a}(r) \cdot C_z^{-1} & \\ \quad - \int_0^1 G_{a,c}(r) G_{a,c}'(r) dr \cdot C_z^{-1} C - \int_0^1 G_{a,c}(r) G_{a,c}'(r) \overline{\Omega}_{aa} dr C_z^{-1} & \text{under LSTUR,} \\ \quad - \int_0^1 G_{a,c}(r) G_{a,c}'(r) \overline{\Sigma}_{aa} dr C_z^{-1}, & \\ -(\Omega_{xx} + \int_0^1 G_a(r) dB_x'(r)) \cdot C_z^{-1} - \int_0^1 G_a(r) G_a'(r) d\breve{D}_{B_a}(r) \cdot C_z^{-1} & \text{under STUR,} \\ \quad - \int_0^1 G_a(r) G_a'(r) \overline{\Omega}_{aa} dr C_z^{-1} - \int_0^1 G_a(r) G_a'(r) \overline{\Sigma}_{aa} dr C_z^{-1}, & \end{array} \right\}$$

for short-horizon IVX, where $d\breve{D}_{B_a}(r) := \operatorname{diag}\{a_1' dB_a(r), a_2' dB_a(r), ..., a_n' dB_a(r)\}$.

Proof. For (i) under the QR-IVX case, use $\Psi_\tau(u_{0t\tau})$ to replace u_{0t}, and use Ω_{00} to replace $\tau(1-\tau)$, the way of proving the asymptotic normality of short-horizon IVX can be easily extended to the case of short-horizon QR-IVX. ■

Lemma A.3.6 *For a generic constant* $C > 0$*, thus*

$$\sup \left\{ \|H_T(\epsilon) - H_T(0)\| : \|\epsilon\| \leq T^{\frac{1+\gamma}{2}} C \right\} = o_p(1).$$

Proof. The results are identical to the statements of Lee (2016). ■

A.3.2 Proof of Theorem 4.3.1

Proof. Since the uniform convergence is confirmed in Lemma A.3.6, the standard result for the extreme estimation with non-smooth criterion function holds following the approach in Lee (2016). Let $\widehat{\beta}_{1,\tau} = \widehat{\beta}_{1,\tau}^{QRIVX}$ within the proof. Define $\widehat{\epsilon}_\tau = (\widehat{\beta}_{1,\tau} - \beta_{1,\tau})$, then

$$\widehat{\beta}_{1,\tau} \sim \arg\min \left(\sum_{t=1}^T m_t(\beta_1) \right)' \left(\sum_{t=1}^T m_t(\beta_1) \right),$$

where $m_t(\beta_1) = \widetilde{z}_{t-1} \Psi_\tau(u_{0t\tau}(\beta_1)) = \widetilde{z}_{t-1} \left(\tau - 1(y_{t\tau} \leq \beta_{1,\tau}' x_{t-1}) \right)$. Based on Lemma A.3.6, the first-order condition for the QR-IVX objective function is given as,

$$\begin{aligned} & o_p(1) \\ = \; & \frac{1}{T^{\frac{1+\gamma}{2}}} \sum_{t=1}^T \widetilde{z}_{t-1} \left\{ \Psi \left(u_{0t\tau} - \left(\widehat{\beta}_{1,\tau} - \beta_{1,\tau} \right)' x_{t-1} \right) \right\} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\left\{\Psi_\tau\left(u_{0t\tau}-\widehat{\epsilon}_\tau^{'}x_{t-1}\right)-\mathbb{E}_{t-1}\left(\Psi_\tau\left(u_{0t\tau}-\widehat{\epsilon}_\tau^{'}x_{t-1}\right)\right)-\Psi_\tau\left(u_{0t\tau}\right)+\mathbb{E}_{t-1}\left(\Psi_\tau\left(u_{0t\tau}\right)\right)\right\} \\
&\quad +\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\mathbb{E}_{t-1}\left(\Psi_\tau\left(u_{0t\tau}-\widehat{\epsilon}_\tau^{'}x_{t-1}\right)\right)+\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\left\{\Psi_\tau\left(u_{0t\tau}\right)\right\}+o_p(1) \\
&= \frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\mathbb{E}_{t-1}\left(\Psi_\tau\left(u_{0t\tau}-\widehat{\epsilon}_\tau^{'}x_{t-1}\right)\right)+\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}\left\{\Psi_\tau\left(u_{0t\tau}\right)\right\}+o_p(1).
\end{aligned}
$$

The term $\mathbb{E}_{t-1}\left(\Psi_\tau\left(u_{0t\tau}-\widehat{\epsilon}_\tau^{'}x_{t-1}\right)\right)$ can be expanded around $\epsilon_\tau=0$ as

$$
\begin{aligned}
&\mathbb{E}_{t-1}\left(\Psi_\tau\left(u_{0t\tau}-\widehat{\epsilon}_\tau^{'}x_{t-1}\right)\right) \\
&= \mathbb{E}_{t-1}\left[1\left(u_{0t\tau}-\epsilon_\tau^{'}x_{t-1}\right)\right]|_{\epsilon_\tau=0}+\frac{\partial\mathbb{E}_{t-1}\left(\Psi_\tau\left(u_{0t\tau}-\widehat{\epsilon}_\tau^{'}x_{t-1}\right)\right)}{\partial\epsilon_\tau^{'}}|_{\epsilon_\tau=0}\widehat{\epsilon}_\tau+o_p\left(\widehat{\epsilon}_\tau\right),
\end{aligned}
$$

where

$$
\mathbb{E}_{t-1}\left(\Psi_\tau\left(u_{0t\tau}-\epsilon_\tau^{'}x_{t-1}\right)\right)=\tau-\mathbb{E}_{t-1}\left(1\left(u_{0t\tau}<\epsilon_\tau^{'}x_{t-1}\right)\right)=\tau-\int_{-\infty}^{\epsilon_\tau^{'}x_{t-1}}p_{u0t\tau,t-1}(s)ds,
$$

and

$$
\frac{\partial\mathbb{E}_{t-1}\left[\xi_\tau(u_{ot\tau}-\epsilon_\tau^{'}x_{t-1})\right]}{\partial\epsilon_\tau^{'}}|_{\epsilon_\tau=0}=-x_{t-1}^{'}p_{u0t\tau,t-1}(0).
$$

So we have

$$
\mathbb{E}_{t-1}\left[\xi_\tau(u_{0t\tau}-\widehat{\epsilon}_\tau^{'}x_{t-1})\right]=-x_{t-1}^{'}p_{u0t\tau,t-1}(0)\widehat{\epsilon}_\tau+o_p(1).
$$

Therefore, the first order condition follows

$$
T^{\frac{1+\gamma}{2}}(\widehat{\beta}_1-\beta_1)=\left(\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\widetilde{z}_{t-1}x_{t-1}^{'}\right)^{-1}\left(\frac{1}{T^{\frac{1+\gamma}{2}}}\sum_{t=1}^{T}\widetilde{z}_{t-1}u_{0t}\right)+o_p(1).
$$

In all, we complete our proofs. ■

A.3.3 Proof of Theorem 4.3.2

Proof. Based on the asymptotic distribution results in Theorem 4.3.1, the proof for the Wald test is trivial and straightforward. ■

A.3.4 Technical lemmas in long-horizon predictive regression

Lemma A.3.7 *Under the rate condition* $\frac{\sqrt{T}}{T^{\gamma}}+\frac{T^{\gamma}}{k}+\frac{k}{T}\rightarrow 0$*, with* $\breve{D}_{B_a}(r) := \operatorname{diag}\{a_1' B_a(r), a_2' B_a(r), ..., a_n' B_a(r)\}$*, then*

(i) With $t=[Tr]$, $\frac{C_z}{kT^{\frac{1}{2}+\gamma}}\eta_{Tt}^{1,k} \rightsquigarrow \begin{cases} G_a(r), & \text{under LSTUR,} \\ G_{a,c}(r), & \text{under STUR.} \end{cases}$

(ii) $\sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{1,k})' = O_p(kT^{1+\gamma})$*, thus* $\frac{1}{\sqrt{k}T^{\frac{3}{2}+\gamma}}\sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{1,k})' = o_p(1)$.

(iii) $\frac{C_z}{\sqrt{k}T^{\gamma}}z_t^k \Rightarrow B_x(1) =_d \mathcal{N}(0,\Omega_{xx})$*, and* $\frac{1}{T}\sum_{t=1}^{T-k}(\frac{C_z}{\sqrt{k}T^{\gamma}}z_t^k)(\frac{C_z}{\sqrt{k}T^{\gamma}}z_t^k)' \rightsquigarrow \Omega_{xx}$.

(iv) $\frac{1}{T^{\frac{1}{2}+\gamma}\sqrt{k}}\sum_{t=1}^{T-k} u_{0,t+k}(z_t^k)' \rightsquigarrow \mathcal{N}(0, C_z^{-1}\Omega_{xx}C_z^{-1}\Omega_{00})$.

(v) With $t=[Tr]$, $\frac{C_z}{\sqrt{k}T^{\frac{1}{2}+\gamma}}\eta_{T,t}^{2,k} \rightsquigarrow \begin{cases} \breve{D}_{B_a}(1)\, G_{a,c}(r), & \text{under LSTUR,} \\ \breve{D}_{B_a}(1)\, G_a(r), & \text{under STUR.} \end{cases}$

(vi) $\frac{C_z}{\sqrt{k}T^{1+\gamma}}\sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{2,k})' \rightsquigarrow \begin{cases} \int_0^1 \breve{D}_{B_a}(1)\, G_{a,c}(r)dB_0(r), & \text{under LSTUR,} \\ \int_0^1 \breve{D}_{B_a}(1)\, G_a(r)dB_0(r), & \text{under STUR.} \end{cases}$

(vii)

$$\frac{1}{kT^{3+2\gamma}}\sum_{t=1}^{T-k}\eta_{T,t}^{2,k}(\eta_{T,t}^{2,k})' \rightsquigarrow \begin{cases} C_z^{-1}\cdot\int_0^1 \breve{D}_{B_a}(1)\, G_{a,c}(r)G_{a,c}'(r)\breve{D}_{B_a}(r)\, dr\cdot C_z^{-1}, & \text{under LSTUR,} \\ C_z^{-1}\cdot\int_0^1 \breve{D}_{B_a}(1)\, G_a(r)G_a'(r)\breve{D}_{B_a}(r)\, dr\cdot C_z^{-1}, & \text{under STUR.} \end{cases}$$

(viii) $\frac{1}{kT^{\frac{5}{2}+2\gamma}}\sum_{t=1}^{T-k} z_t^k(\eta_{T,t}^{2,k})' \rightsquigarrow \begin{cases} C_z^{-1}\cdot\int_0^1 B_x(1)\cdot G_{a,c}'(r)\breve{D}_{B_a}(r)\, dr\cdot C_z^{-1}, & \text{under LSTUR,} \\ C_z^{-1}\cdot\int_0^1 B_x(1)\cdot G_a'(r)\breve{D}_{B_a}(r)\, dr\cdot C_z^{-1}, & \text{under STUR.} \end{cases}$

(ix) With $t=[Tr]$, $\frac{C_z}{kT^{\frac{1}{2}+\gamma}}\eta_{Tt}^{3,k} \rightsquigarrow \begin{cases} \overline{\Sigma}_{aa}G_{a,c}(r), & \text{under LSTUR,} \\ \overline{\Sigma}_{aa}G_a(r), & \text{under STUR.} \end{cases}$

(x) $\sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{3,k})' = O_p(kT^{1+\gamma})$*, thus* $\frac{1}{\sqrt{k}T^{\frac{3}{2}+\gamma}}\sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{3,k})' = o_p(1)$.

Proof. (i)(ii) and (iii) are identical to those of Lemma A.1 in Phillips and Lee (2013). Just replace $J_c(r)$ by $G_a(r)$ and $G_{a,c}(r)$, since $\sup_{1\leq t\leq T}\mathbb{E}\left\|\eta_{T,t-1}^{(1)}\right\| = O_p(T^{1+2\gamma})$ under LP(2017) and LP(2020). This rate is the standard rate for the unit-root and local-to-unity case; (iv) is identical to the proof of Lemma A.2 in Phillips and Lee (2013);

(v) The recursive formula is satisfied as $\eta^{(2)}_{T,t} = R_{Tt}\eta^{(2)}_{T,t-1} + \breve{D}_{at}x_{t-1}$, and $\eta^{(2)}_{T,t+j-1} = R_{Tt}\eta^{(2)}_{T,t+j-2} + \breve{D}_{a,t+j-1}x_{t+j-2}$. Take summation up to k,

$$\begin{aligned}
\eta^{2,k}_{T,t} &= \sum_{j=1}^{k}\eta^{(2)}_{T,t+j-1} = R_{Tz}\sum_{j=1}^{k}\eta^{(2)}_{T,t+j-2} + \sum_{j=1}^{k}\breve{D}_{a,t+j-1}x_{t+j-2} \\
&\Rightarrow \; (I_n - R_{Tz})\sum_{j=1}^{k}\eta^{(2)}_{T,t+j-1} = \sum_{j=1}^{k}\breve{D}_{a,t+j-1}x_{t+j-2} - R_{Tz}\eta^{(2)}_{T,t+k-1} + R_{Tz}\eta^{(2)}_{T,t-1} \\
&\Rightarrow \; \frac{C_z}{\sqrt{k}T^{\frac{1+2\gamma}{2}}}\sum_{j=1}^{k}\eta^{(2)}_{T,t+j-1} = \frac{1}{\sqrt{k}T^{\frac{1}{2}}}\sum_{j=1}^{k}\breve{D}_{a,t+j-1}x_{t+j-2} + o_p(1),
\end{aligned}$$

where $[Tr] = t$ and $\frac{1}{l_T} + \frac{l_T}{k} \to 0$. Hence $\frac{1}{\sqrt{k}}\sum_{j=1}^{l_T}\breve{D}_{at+j-1}(x_{t+j-2}/\sqrt{T}) = o_p(1)$. By the techniques of summation splitting and residuals putting back, we have

$$\begin{aligned}
\frac{C_z}{\sqrt{k}T^{\frac{1+2\gamma}{2}}}\sum_{j=1}^{k}\eta^{(2)}_{T,t+j-1} &= \frac{1}{\sqrt{k}}\sum_{j=1}^{k}\breve{D}_{at+j-1}(x_{t+j-2}/\sqrt{T}) + o_p(1) \\
&= \frac{1}{\sqrt{k}}\sum_{j=l_T}^{k}\breve{D}_{at+j-1}(x_{t+j-2}/\sqrt{T}) + o_p(1) \\
&= \frac{1}{\sqrt{k}}\sum_{j=l_T}^{k}\breve{D}_{at+j-1}(G_{a,c}(r) + o_p(1)) + o_p(1) \\
&\rightsquigarrow \breve{D}_{B_a}(1)\, G_{a,c}(r),
\end{aligned}$$

for LSTUR, under $\frac{1}{\sqrt{kT}}\eta^{(2)}_{T,t} = O_p(\frac{T^{\frac{\gamma}{2}}}{\sqrt{k}}) = o_p(1)$. Let $C = 0_{n\times n}$, we then have the STUR casse.

(vi) For the LSTUR case,

$$\frac{C_z}{\sqrt{k}T^{1+\gamma}}\sum_{t=1}^{T-k}u_{0,t+k}(\eta^{2,k}_{T,t})' = \sum_{t=1}^{T-k}\left(\frac{u_{0,t+k}}{\sqrt{T}}\right)\left(\frac{\eta^{2,k}_{T,t}}{\sqrt{k}T^{\frac{1+2\gamma}{2}}}\right)' = \int_0^1 dB_0(r)G'_{a,c}(r)\breve{D}_{B_a}(1) + o_p(1),$$

and for STUR, $\frac{C_z}{\sqrt{k}T^{1+\gamma}}\sum_{t=1}^{T-k}u_{0,t+k}(\eta^{2,k}_{T,t})' = \int_0^1 dB_0(r)G'_a(r)\breve{D}_{B_a}(1) + o_p(1)$.

(vii) From (v) and FCLT (White, 2014),

$$\begin{aligned}
\frac{1}{kT^{\frac{3}{2}+2\gamma}}\sum_{t=1}^{T-k}z^k_t(\eta^{2,k}_{T,t})' &= C_z^{-1}\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{C_z}{\sqrt{k}T^{\gamma}}z^k_t\right)\left(\frac{C_z}{\sqrt{k}T^{\frac{1}{2}+\gamma}}\eta^{2,k}_{T,t}\right)'C_z^{-1} \\
&\rightsquigarrow C_z^{-1}\int_0^1 B_x(1)G'_{a,c}(r)\breve{D}_{B_a}(1)\, dr C_z^{-1}.
\end{aligned}$$

(viii) From (iii) (v) and the FCLT (White, 2014),

$$\begin{aligned}\frac{1}{kT^{2+2\gamma}}\sum_{t=1}^{T-k}\eta_{T,t}^{2,k}(\eta_{T,t}^{2,k})' &= C_z^{-1}\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{C_z}{\sqrt{k}T^{\frac{1}{2}+\gamma}}\eta_{T,t}^{2,k}\right)\left(\frac{C_z}{\sqrt{k}T^{\frac{1}{2}+\gamma}}\eta_{T,t}^{2,k}\right)'C_z^{-1}\\ &\rightsquigarrow C_z^{-1}\int_0^1 \breve{D}_{B_a}(1)\,G_{a,c}(r)G'_{a,c}(r)\breve{D}_{B_a}(1)\,drC_z^{-1}.\end{aligned}$$

(ix) The recursive formula applies for $\eta_{T,t}^{(3)} = R_{Tz}\eta_{T,t-1}^{(3)} + \breve{D}_{at}^2 x_{t-1}$, and take summation up to k,

$$\begin{aligned}\eta_{T,t}^{3,k} &= \sum_{j=1}^{k}\eta_{T,t+j-1}^{(3)} = R_{Tz}\sum_{j=1}^{k}\eta_{T,t+j-2}^{(3)} + \sum_{j=1}^{k}[\breve{D}_{at}^2]x_{t+j-2}\\ &\Rightarrow (I_n - R_{Tz})\sum_{j=1}^{k}\eta_{T,t+j-1}^{(3)} = \sum_{j=1}^{k}[\breve{D}_{at}^2]x_{t+j-2} - R_{Tz}\eta_{T,t+k-1}^{(3)} + R_{Tz}\eta_{T,t-1}^{(3)}\\ &\Rightarrow \frac{C_z}{kT^{\frac{1+2\gamma}{2}}}\sum_{j=1}^{k}\eta_{T,t+j-1}^{(3)} = \frac{1}{kT^{\frac{1}{2}}}\sum_{j=1}^{k}[\breve{D}_{at}^2]x_{t+j-2} + o_p(1),\end{aligned}$$

where $[Tr] = t$ and $\frac{1}{l_T} + \frac{l_T}{k} \to 0$. Hence $\frac{C_z}{\sqrt{k}}\sum_{j=1}^{l_T}(\breve{D}_{at+j-1}^2)(x_{t+j-2}/\sqrt{T}) = o_p(1)$. Therefore,

$$\begin{aligned}\frac{C_z}{kT^{\frac{1+2\gamma}{2}}}\sum_{j=1}^{k}\eta_{T,t+j-1}^{(3)} &= \frac{1}{k}\sum_{j=1}^{k}[\breve{D}_{a,t+j-1}^2](x_{t+j-2}/\sqrt{T}) + o_p(1)\\ &= \frac{1}{k}\sum_{j=l_T}^{k}[\breve{D}_{a,t+j-1}^2](x_{t+j-2}/\sqrt{T}) + o_p(1)\\ &= \frac{1}{k}\sum_{j=l_T}^{k}[\breve{D}_{a,t+j-1}^2](G_{a,c}(r) + o_p(1)) + o_p(1)\\ &\rightsquigarrow \overline{\Sigma}_{aa}G_{a,c}(r),\end{aligned}$$

for LSTUR. Let $C = 0_{n\times n}$, these results extend to STUR.

(x) follows Lemma A.1 of Phillips and Lee (2013). ■

Lemma A.3.8 *Under the rate condition that* $\frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0$,

$$\frac{1}{\sqrt{k}T^{\frac{1}{2}+\gamma}}\sum_{t=1}^{T-k}u_{0,t+k}(\tilde{z}_t^k)' \rightsquigarrow \mathcal{MN}(0,\ (V_{zz}^{LH} + V_{\eta\eta}^{(2),LH} + V_{z\eta}^{(2),LH} + (V_{z\eta}^{(2),LH})')\Omega_{00}), \tag{A.3.10}$$

where $V_{zz}^{LH} := C_z^{-1}\Omega_{xx}C_z^{-1}$,

$$V_{\eta\eta}^{(2),LH} := \begin{cases} C_z^{-1} \cdot \int_0^1 \breve{D}_{B_a}(1)\, G_{a,c}(r)G_{a,c}^{'}(r)\breve{D}_{B_a}(r)\, dr \cdot C_z^{-1}, & \text{under LSTUR}, \\ C_z^{-1} \cdot \int_0^1 \breve{D}_{B_a}(1)\, G_a(r)G_a^{'}(r)\breve{D}_{B_a}(r)\, dr \cdot C_z^{-1}, & \text{under STUR}, \end{cases}$$

and

$$V_{z\eta}^{(2),LH} := \begin{cases} C_z^{-1} \cdot \int_0^1 B_x(1)G_{a,c}^{'}(r)\breve{D}_{B_a}(1)\, dr \cdot C_z^{-1}, & \text{under LSTUR}, \\ C_z^{-1} \cdot \int_0^1 B_x(1)G_a^{'}(r)\breve{D}_{B_a}(1)\, dr \cdot C_z^{-1}, & \text{under STUR}. \end{cases}$$

Proof. It is sufficient to justify the joint asymptotical normality of two leading terms, as

$$\begin{pmatrix} \frac{1}{T^{\frac{1}{2}+\gamma}\sqrt{k}} \sum_{t=1}^{T-k} u_{0,t+k}(z_t^k)' \\ \frac{1}{T^{1+\gamma}\sqrt{k}} \sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{2,k})' \end{pmatrix} \rightsquigarrow \mathcal{MN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \;\; \Omega_{00} \begin{pmatrix} V_{zz}^{LH} & V_{z\eta}^{(2),LH} \\ (V_{z\eta}^{(2),LH})' & V_{\eta\eta}^{(2),LH} \end{pmatrix} \right).$$

By Cramér-Wold device, it is sufficient to show that for any constants α and β,

$$\begin{aligned} &\frac{\alpha}{T^{\frac{1}{2}+\gamma}\sqrt{k}} \sum_{t=1}^{T-k} u_{0,t+k}(z_t^k)' + \frac{\beta}{T^{1+\gamma}\sqrt{k}} \sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{2,k})' \\ &\rightsquigarrow \mathcal{MN}\left(0, \Omega_{00}\left[\alpha^2 V_{zz}^{LH} + \beta^2 V_{\eta\eta}^{(2),LH} + \alpha\beta V_{z\eta}^{(2),LH} + \alpha\beta \left(V_{z\eta}^{(2),LH}\right)'\right]\right), \end{aligned}$$

where V_{zz}^{LH}, $V_{z\eta}^{(2),LH}$, $V_{\eta\eta}^{(2),LH}$ are defined in Lemma A.3.7. Define $\frac{\alpha}{T^{\frac{1}{2}+\gamma}\sqrt{k}} \sum_{t=1}^{T-k} u_{0,t+k}(z_t^k)' + \frac{\beta}{T^{1+\gamma}\sqrt{k}} \sum_{t=1}^{T-k} u_{0,t+k}(\eta_{T,t}^{2,k})' = \sum_{t=1}^{T-k} \xi_{Tt}$, where $\xi_{Tt} := \frac{\alpha}{T^{\frac{1}{2}+\gamma}\sqrt{k}} u_{0,t+k}(z_t^k)' + \frac{\beta}{T^{1+\gamma}\sqrt{k}} u_{0,t+k}(\eta_{T,t}^{2,k})'$. Here ξ_{Tt} are martingale difference sequence for any $t \in \{1, 2, ..., T\}$. In order to apply the martingale CLT (White, 2014), the stability and Lindeberg conditions need to be justified.

First, with $\mathbb{E}_{t+k-1}(\cdot) := \mathbb{E}(\cdot|\mathcal{F}_{t+k-1})$, the stability condition is shown as,

$$\begin{aligned} &\sum_{t=1}^{T-k} \mathbb{E}_{t+k-1}\left[\xi_{Tt}\xi_{Tt}^{'}\right] \\ = \;\; &\Omega_{00}\left[\sum_{t=1}^{T-k} \frac{\alpha^2}{T^{1+2\gamma}} z_t^k \left(z_t^k\right)' + \sum_{t=1}^{T-k} \frac{\alpha\beta}{T^{\frac{3}{2}+2\gamma}} \left(z_t^k(\eta_{T,t}^{2,k})' + \eta_{T,t}^{2,k}\left(z_t^k\right)'\right) + \sum_{t=1}^{T-k} \frac{\beta^2}{T^{2+2\gamma}} \eta_{T,t}^{2,k}\left(\eta_{T,t}^{2,k}\right)'\right] \end{aligned}$$

$$\rightsquigarrow \quad \Omega_{00}\left[\alpha^2 V_{zz}^{LH} + \beta^2 V_{\eta\eta}^{(2),LH} + \alpha\beta\left(V_{z\eta}^{(2),LH} + \left(V_{z\eta}^{(2),LH}\right)'\right)\right],$$

since $\sup_{1\leq t\leq T}\mathbb{E}\left\|z_t^k\right\| = O_p(\sqrt{k}T^{\gamma})$, $\sup_{1\leq t\leq T}\mathbb{E}\left\|\eta_{T,t}^{2,k}\right\| = O_p(\sqrt{k}T^{\frac{1}{2}+\gamma})$, and Lemma A.3.7, and the following facts

$$\frac{1}{kT^{1+2\gamma}}\sum_{t=1}^{T-k}(z_t^k)(z_t^k)' \rightsquigarrow V_{zz}^{LH},$$
$$\frac{1}{kT^{2+2\gamma}}\sum_{t=1}^{T-k}\eta_{T,t}^{2,k}(\eta_{T,t}^{2,k})' \rightsquigarrow V_{\eta\eta}^{(2),LH},$$
$$\frac{1}{kT^{\frac{3}{2}+2\gamma}}\sum_{t=1}^{T-k}z_t^k(\eta_{T,t}^{2,k})' \rightsquigarrow V_{z\eta}^{(2),LH},$$

where V_{zz}^{LH}, $V_{z\eta}^{(2),LH}$, and $V_{\eta\eta}^{(2),LH}$ are defined in Lemma A.3.7.

Second, Lindeberg condition is shown: For any $\epsilon > 0$, with $\mathbb{E}_{t+k-1}(\cdot) := \mathbb{E}(\cdot|\mathcal{F}_{t+k-1})$

$$\begin{aligned}
&\sum_{t=1}^{T-k}\mathbb{E}_{t+k-1}\left[\left\|\eta_{T,t}^{2,k}\right\|^2 1\left(\left\|\eta_{T,t}^{2,k}\right\| > \epsilon\right)\right] \\
&\leq \sum_{t=1}^{T-k}\left[\frac{\alpha^2}{T^{1+2\gamma}k}z_t^k\left(z_t^k\right)' + \frac{\alpha\beta}{T^{\frac{3}{2}+2\gamma}k}\left(z_t^k\left(\eta_{T,t}^{2,k}\right)' + \eta_{T,t}^{2,k}\left(z_t^k\right)'\right) + \frac{\beta^2}{T^{2+2\gamma}}\eta_{T,t}^{2,k}\left(\eta_{T,t}^{2,k}\right)'\right]\cdot \\
&\max_{1\leq t+k\leq T}\mathbb{E}_{t+k-1}\left[u_{0,t+k}^2 1\left(\left\|\frac{\alpha}{T^{\frac{1}{2}+\gamma}\sqrt{k}}u_{0,t+k}\left(z_t^k\right)' + \frac{\beta}{T^{1+\gamma}\sqrt{k}}u_{0,t+k}\left(\eta_{T,t}^{2,k}\right)'\right\| > \epsilon\right)\right] \\
&\leq K\cdot\max_{1\leq t+k\leq T}\mathbb{E}_{t+k-1}\left[\begin{array}{c} u_{0,t+k}^2 1\left(\left\|\frac{\alpha^2}{T^{1+2\gamma}k}u_{0,t+k}^2 z_t^k(z_t^k)'\right\| > \frac{\epsilon^2}{2}\right) \\ +u_{0,t+k}^2\cdot 1\left(\left\|\frac{\beta^2}{T^{2+2\gamma}k}u_{0,t+k}^2\eta_{T,t}^{2,k}\left(\eta_{T,t}^{2,k}\right)'\right\| > \frac{\epsilon^2}{2}\right)\end{array}\right] \\
&\leq K\cdot\max_{1\leq t+k\leq T}\mathbb{E}_{t+k-1}\left[u_{0,t+k}^2 1\left(\left(\max_{1\leq t\leq T}\left\|\frac{z_t^k\left(z_t^k\right)'}{kT^{2\gamma}}\right\|\right)|u_{0,t+k}^2|^2 > \frac{\epsilon^2 T}{2}\right)\right] \\
&+ K\cdot\max_{1\leq t+k\leq T}\mathbb{E}_{t+k-1}\left[u_{0,t+k}^2 1\left(\left(\max_{1\leq t\leq T}\left\|\frac{\eta_{T,t}^{2,k}\left(\eta_{T,t}^{2,k}\right)'}{kT^{1+2\gamma}}\right\|\right)|u_{0,t+k}^2|^2 > \frac{\epsilon^2 T}{2}\right)\right] \\
&\leq K\cdot\max_{1\leq t+k\leq T}\mathbb{E}_{t+k-1}\left[u_{0,t+k}^2 1\left(M_1\cdot|u_{0,t+k}^2|^2 > \frac{\epsilon^2 T}{2}\right)\right] \\
&+ K\cdot\max_{1\leq t+k\leq T}\mathbb{E}_{t+k-1}\left[u_{0,t+k}^2 1\left(M_2\cdot|u_{0,t+k}^2|^2 > \frac{\epsilon^2 T}{2}\right)\right] \\
&= o_p(1),
\end{aligned}$$

where K, M_1 and M_2 are three constants. Then the martingale CLT (White, 2014) applies, and shows the asymptotics by letting $\alpha = \beta = 1$. ■

Lemma A.3.9 *Under the rate condition, $\frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0$, we have*

(i) $\frac{1}{T^{1+\gamma}k^2} \sum_{t=1}^{T-k} x_t^k (z_t^k)' \rightsquigarrow \frac{1}{2}\Omega_{xx} C_z^{-1}$.

(ii) $\frac{1}{T^{2+\gamma}k^2} \sum_{t=1}^{T-k} x_t^k \left(\eta_{T,t}^{1,k}\right)' \rightsquigarrow \begin{cases} \int_0^1 G_{a,c}(r) G_{a,c}'(r) dr \cdot C_z^{-1}, & \text{under LSTUR}, \\ \int_0^1 G_a(r) G_a'(r) dr \cdot C_z^{-1}, & \text{under STUR}. \end{cases}$

(iii) $\frac{1}{T^{\frac{3}{2}+\gamma}k^2} \sum_{t=1}^{T-k} x_t^k (\eta_{T,t}^{2,k})' \rightsquigarrow V_{ax} C_z^{-1}$*, where* $V_{ax} := \frac{1}{2}\overline{\Sigma}_{aa}\Omega_{xx}$.

(iv) $\frac{1}{T^{2+\gamma}k^2} \sum_{t=1}^{T-k} x_t^k (\eta_{T,t}^{3,k})' \rightsquigarrow \begin{cases} \int_0^1 G_{a,c}(r) G_{a,c}'(r) \overline{\Sigma}_{aa} dr \cdot C_z^{-1}, & \text{under LSTUR}, \\ \int_0^1 G_a(r) G_a'(r) \overline{\Sigma}_{aa} dr \cdot C_z^{-1}, & \text{under STUR}. \end{cases}$

Proof. (i) Here C is one positive definite matrix proportional to identity matrix. Besides, we employ $\frac{1}{C}$ and $\frac{1}{C_z}$ to denote C^{-1} and C_z^{-1}. The discussions are the same when a positive matrix C that is not proportional to identity matrix or a negative definite C case is considered. Define $R_{Tt}^{(t+j,1+s)} := \prod_{m=s+1}^{t+j-1} R_{Ts}$.

$$\begin{aligned} \sum_{t=1}^{T-k} x_t^k (z_t^k)' &= \sum_{t=1}^{T-k} \left(\sum_{j=1}^{k} x_{t+j-1}\right) \left(\sum_{i=1}^{k} z_{t+i-1}\right)' \\ &= \sum_{t=1}^{T-1} \left(\sum_{j=1}^{k} \sum_{s=1}^{t+j-1} R_{Tt}^{(t+j,1+s)} u_{xs}\right) \left(\sum_{i=1}^{k} \sum_{l=1}^{t+i-1} R_{Tz}^{t+i-1-l} u_{xl}\right)'. \end{aligned}$$

With L defined as the lower order term compared to the leading term, the magnitude of the term is derived:

$$\begin{aligned} \mathbb{E}\left[x_t^k (z_t^k)'\right] &= \mathbb{E}\left[\sum_{t=1}^{T-1} \left(\sum_{j=1}^{k} \sum_{s=1}^{t+j-1} R_{Tt}^{(t+j,1+s)} u_{xs}\right) \left(\sum_{i=1}^{k} \sum_{l=1}^{t+i-1} R_{Tz}^{t+i-1-l} u_{xl}\right)'\right] \\ &= \mathbb{E}\left[\sum_{t=1}^{T-1} \left(\sum_{j=1}^{k} \sum_{s=1}^{t+j-1} R_{T}^{t+j-1-s} u_{xs}\right) \left(\sum_{i=1}^{k} \sum_{l=1}^{t+i-1} R_{Tz}^{t+i-1-l} u_{xl}\right)'\right] \\ &\quad + \mathbb{E}\left[\sum_{t=1}^{T-1} \left(\sum_{j=1}^{k} \sum_{s=1}^{t+j-1} R_{T}^{t+j-1-s} \left(\sum_{m=s+1}^{t+j-1} \frac{\breve{D}_{am}}{\sqrt{T}}\right) u_{xs}\right) \left(\sum_{i=1}^{k} \sum_{l=1}^{t+i-1} R_{Tz}^{t+i-1-l} u_{xl}\right)'\right] \\ &\quad + \mathbb{E}\left[\sum_{t=1}^{T-1} \left(\sum_{j=1}^{k} \sum_{s=1}^{t+j-1} R_{T}^{t+j-1-s} \left(\sum_{m=s+1}^{t+j-1} \frac{\breve{D}_{am}^2}{T}\right) u_{xs}\right) \left(\sum_{i=1}^{k} \sum_{l=1}^{t+i-1} R_{Tz}^{t+i-1-l} u_{xl}\right)'\right] \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\sum_{t=1}^{T-1}\left(\sum_{j=1}^{k}\sum_{s=1}^{t+j-1} R_T^{t+j-1-s}u_{xs}\right)\left(\sum_{i=1}^{k}\sum_{l=1}^{t+i-1} R_{Tz}^{t+i-1-l}u_{xl}\right)'\right] + L \\
&= \Omega_{xx}\frac{\left(1-R_T^k\right)\left(1-R_{Tz}^k\right)}{\left(1-R_T\right)\left(1-R_{Tz}\right)}\left\{\left(R_TR_{Tz}\right)^0 + \left(R_TR_{Tz}\right)^1 + ... + \left(R_TR_{Tz}\right)^{t-1}\right\} \\
&\quad +\Omega_{xx}\frac{1}{\left(1-R_T\right)\left(1-R_{Tz}\right)}\left\{\begin{array}{c}\left(1-R_T\right)\left(1-R_{Tz}\right)+\left(1-R_T^2\right)\left(1-R_{Tz}^2\right)\\ +...+\left(1-R_T^{k-1}\right)\left(1-R_{Tz}^{k-1}\right)\end{array}\right\} + L \\
&=: AA + BB + L,
\end{aligned}
$$

since $R_{Tt}^{(t+j,1+s)} \sim R_T^{t+j-1-s}\prod_{m=s+1}^{t+j-1}\exp\left(\frac{\check{D}_{am}}{\sqrt{T}}\right) \sim R_T^{t+j-1-s}+R_T^{t+j-1-s}\sum_{m=s+1}^{t+j-1}\frac{\check{D}_{am}}{\sqrt{T}}+ R_T^{t+j-1-s}\sum_{m=s+1}^{t+j-1}\frac{\check{D}_{am}^2}{T}$ where $R_T := I_n - \frac{C}{T}$. First, for term A, since $\frac{\sqrt{T}}{k}+\frac{k}{T}+\frac{T^\gamma}{k} \to 0$, $R_{Tz}^k \to 0$ and $R_T^k \sim \exp\left(-C\frac{k}{T}\right)$, thus

$$
\begin{aligned}
AA &= \left(\frac{\Omega_{xx}}{CC_z}\right)T^{1+\gamma}\left(1-R_T^k\right)\left(1-R_{Tz}^k\right)\left\{\frac{1-\left(R_TR_{Tz}\right)^t}{1-R_TR_{Tz}}\right\} \\
&= \left(\frac{\Omega_{xx}}{CC_z}\right)T^{1+\gamma}\left(1-R_T^k\right)\left\{\frac{1-\left(R_TR_{Tz}\right)^t}{1-R_TR_{Tz}}\right\}(1+o(1)).
\end{aligned}
$$

Note that $T^\gamma(1-R_TR_{Tz}) = C_z + \frac{C}{T^{1-\gamma}} - \frac{CC_z}{T} = C_z + o(1)$ and $(R_TR_{Tz})^k \to 0$ where t is sufficiently large since $R_{Tz}^t \to 0$. Therefore

$$
AA = \left(\frac{\Omega_{xx}}{CC_z}\right)T^{1+2\gamma}(1-R_T^k)\left\{\frac{1+o(1)}{C_z+o(1)}\right\}(1+o(1)),
$$

with $1-R_T^k \sim \frac{k}{T}C$. As $T\to\infty$,

$$
\frac{AA}{kT^{2\gamma}} = \left(\frac{\Omega_{xx}}{CC_z}\right)(C+o(1))\left\{\frac{1+o(1)}{C_z+o(1)}\right\}(1+o(1)) \to \frac{\Omega_{xx}}{C_z^2},
$$

and $AA = O_p(kT^{2\gamma})$. The lower order term L is of the order $O_p(kT^{2\gamma-1})$. Second, for the term BB,

$$
\begin{aligned}
BB &= \left(\frac{\Omega_{xx}}{CC_z}\right)T^{1+\gamma}\left\{(k-1)-\frac{1-R_T^{k-1}}{1-R_T}R_T-\frac{1-R_{Tz}^{k-1}}{1-R_{Tz}}R_{Rz}+\frac{1-(R_TR_{Tz})^{k-1}}{1-R_TR_{Tz}}R_TR_{Tz}\right\} \\
&\sim \left(\frac{\Omega_{xx}}{CC_z}\right)T^{1+\gamma}\left\{k-\frac{T}{C}\left(1-\exp\left(-C\frac{k}{T}\right)\right)-\frac{T^\gamma}{C_z}\left(1-\exp\left(-C_z\frac{k}{T^\gamma}\right)\right)+T^\gamma\frac{1+o(1)}{(C_z+o(1))}\right\}.
\end{aligned}
$$

since $\frac{T}{k^2}\left\{k - \frac{T}{C}(1-\exp(-C\frac{k}{T})-1)\right\} = \frac{T}{C}\left\{-C\frac{k}{T} + \frac{C^2}{2}\frac{k^2}{T^2} - \frac{C^3}{6}\frac{k^3}{T^3} + ...\right\} = -k + \frac{C}{2}\frac{k^2}{T} - \frac{C^2}{6}\frac{k^3}{T^2} + ... = -k + \frac{C}{2}\frac{k^2}{T} + o(\frac{k^2}{T})$. If $\frac{T^{1+\gamma}}{k^2} \to 0$, thus

$$BB = \left(\frac{\Omega_{xx}}{CC_z}\right) T^{1+\gamma}\frac{k^2}{T}\left\{\frac{C}{2} + o(1) - \frac{T^{1+\gamma}}{k^2}O_p(1)\right\} \rightsquigarrow \frac{BB}{k^2T^\gamma} = \left(\frac{\Omega_{xx}}{2C_z}\right) + o_p(1).$$

Since the parameter γ is chosen by the researcher: If $\frac{T^{1+\gamma}}{k^2} \to 0$, the term BB dominates the term AA, and

$$\frac{1}{k^2T^{1+\gamma}}\sum_{t=1}^{T-k} x_t^k\left(z_t^k\right)' = \frac{1}{2}\Omega_{xx}C_z^{-1} + o_p(1).$$

(ii) By the FCLT (White, 2014), as $T \to \infty$

$$\begin{aligned}\frac{1}{k^2T^{2+\gamma}}\sum_{t=1}^{T-k} x_t^k(\eta_{T,t}^{1,t})' &= \frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{1}{k\sqrt{T}}x_t^k\right)\left(\frac{C_z}{kT^{\frac{1}{2}+\gamma}}\eta_{T,t}^{1,k}\right)'\frac{1}{T}C_z^{-1} \\ &\rightsquigarrow \begin{cases}\int_0^1 G_{a,c}(r)G_{a,c}'(r)dr\cdot C_z^{-1} & \text{under LSTUR,}\\ \int_0^1 G_a(r)G_a'(r)dr\cdot C_z^{-1} & \text{under STUR.}\end{cases}\end{aligned}$$

(iii) The proof follows the identical procedure of (i).

(iv) Similarly as (ii) and by FCLT (White, 2014), as $T \to \infty$,

$$\begin{aligned}\frac{1}{k^2T^{2+\gamma}}\sum_{t=1}^{T-k} x_t^k(\eta_{T,t}^{3,t})' &= \frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{1}{k\sqrt{T}}x_t^k\right)\left(\frac{C_z}{kT^{\frac{1}{2}+\gamma}}\eta_{T,t}^{3,k}\right)'C_z^{-1} \\ &\rightsquigarrow \begin{cases}\int_0^1 G_{a,c}(r)G_{a,c}'(r)\overline{\Sigma}_{aa}dr\cdot C_z^{-1} & \text{under LSTUR,}\\ \int_0^1 G_a(r)G_a'(r)\overline{\Sigma}_{aa}dr\cdot C_z^{-1} & \text{under STUR.}\end{cases}\end{aligned}$$

■

Lemma A.3.10 *Under the rate condition that* $\frac{\sqrt{T}}{T^\gamma} + \frac{T^\gamma}{k} + \frac{k}{T} \to 0$, *we have*

$$\frac{1}{T^{1+\gamma}k^2}\sum_{t=1}^{T-k} x_t^k(\tilde{z}_t^k)' \rightsquigarrow \Upsilon := \begin{cases}\frac{1}{2}\Omega_{xx}C_z^{-1} + V_{ax}C_z^{-1} + \int_0^1 G_{a,c}(r)G_{a,c}'(r)dr\cdot C_z^{-1}C \\ \quad + \int_0^1 G_{a,c}(r)G_{a,c}'(r)\overline{\Sigma}_{aa}dr\cdot C_z^{-1}, & \text{under LSTUR,}\\ \frac{1}{2}\Omega_{xx}C_z^{-1} + V_{ax}C_z^{-1} + \int_0^1 G_a(r)G_a'(r)\overline{\Sigma}_{aa}dr\cdot C_z^{-1} & \text{under STUR.}\end{cases}$$

where V_{ax} is defined in Lemma A.3.9.

Proof. The proof is easily derived based on the results of Lemma A.3.9. ■

A.3.5 Simulation methods

As shown in Lee (2016), the computational burden of original IVX method is heavy as the non-smooth objective function can generate multiple local optima. Instead, Chapter 4 follows the alternative computational method proposed in Lee (2016), where we test $\mathcal{H}_0 : \vartheta_\tau = 0$ in the alternative ordinary QR model:

$$\widehat{\vartheta}_\tau^{QRIVX} := \arg\min_{\vartheta} \sum_{t=1}^{T} \rho_\tau \left(y_{\tau t} - \vartheta_\tau' \widetilde{z}_{t-1} \right), \tag{A.3.11}$$

where $\rho_\tau(u) = u(\tau - 1(u < 0))$, $\tau \in (0, 1)$. As the self-generated instrument $\widetilde{z}_t$ is produced by the persistent regressor x_t, their "distance" can be asymptotically negligible. This argument justifies the validity and convenience of the conventional QR method based on the self-generated instrument $\widetilde{z}_t$. Following the above discussions of Chapter 4, we can show the limit distribution of $\widehat{\vartheta}_\tau$ is Gaussian for each τ for STUR/LSTUR regressors, while the cases of MER, MIR and LUR have already been proved in Lee (2016). Although the asymptotic normality is not free of nuisance parameter C and A, the self-normalized Wald test still follows pivotal distribution under the null hypothesis of no predictability: Under $\mathcal{H}_0 : \vartheta_\tau = 0$,

$$\frac{\widehat{p_{u0\tau}(0)}^2}{\tau(1-\tau)} \left(\widehat{\vartheta}_\tau^{QRIVX} - \vartheta_\tau \right)' \left(\sum_{t=1}^{T} \widetilde{z}_{t-1} z_{t-1}' \right) \left(\widehat{\vartheta}_\tau^{QRIVX} - \vartheta_\tau \right) \rightsquigarrow \chi^2(n),$$

with STUR/LSTUR regressors. The mixed-root case follow similar derivations. The Monte Carlo simulation shows the robustness of the alternative procedure in Lee (2016). The numerical simulation of this chapter employs the codes provided by Professor Ji Hyung Lee on his website: https://sites.google.com/site/jihyung412/.