5-2019

Three essays on nonstationary time-series analysis and network dynamics

Yubo TAO
Singapore Management University, yubo.tao.2014@phdecons.smu.edu.sg

THREE ESSAYS ON NONSTATIONARY TIME-SERIES ANALYSIS AND NETWORK DYNAMICS

YUBO TAO

SINGAPORE MANAGEMENT UNIVERSITY

2019

Three Essays on Nonstationary Time-series Analysis and Network Dynamics

by
Yubo Tao

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 (Supervisor/Co-Chair)
Sterling Professor of Economics and Statistics
Yale University
Distinguished Term Professor of Economics
Singapore Management University

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

Yichong Zhang
Assistant Professor of Economics
Singapore Management University

Anastasios Magdalinos
Professor of Econometrics
University of Southampton

Singapore Management University
2019

Abstract

My dissertation consists of three essays which contribute new theoretical results to nonstationary time-series analysis and network dynamics.

Chapter 2 examines the limit properties of information criteria (such as AIC, BIC, HQIC) for distinguishing between the unit root model and the various kinds of explosive models. The explosive models include the local-to-unit-root model, the mildly explosive model and the regular explosive model. Initial conditions with different order of magnitude are considered. Both the OLS estimator and the indirect inference estimator are studied. It is found that BIC and HQIC, but not AIC, consistently select the unit root model when data come from the unit root model. When data come from the local-to-unit-root model, both BIC and HQIC select the wrong model with probability approaching 1 while AIC has a positive probability of selecting the right model in the limit. When data come from the regular explosive model or from the mildly explosive model in the form of $1 + n^{\alpha}/n$ with $\alpha \in (0, 1)$, all three information criteria consistently select the true model. Indirect inference estimation can increase or decrease the probability for information criteria to select the right model asymptotically relative to OLS, depending on the information criteria and the true model. Simulation results confirm our asymptotic results in finite sample.

Chapter 3 studies a continuous time dynamic system with a random persistence parameter. The exact discrete time representation is obtained and related to several discrete time random coefficient models currently in the literature. The model distinguishes various forms of unstable and explosive behavior according to specific

regions of the parameter space that open up the potential for testing these forms of extreme behavior. A two-stage approach that employs realized volatility is proposed for the continuous system estimation, asymptotic theory is developed, and test statistics to identify the different forms of extreme sample path behavior are proposed. Simulations show that the proposed estimators work well in empirically realistic settings and that the tests have good size and power properties in discriminating characteristics in the data that differ from typical unit root behavior. The theory is extended to cover models where the random persistence parameter is endogenously determined. An empirical application based on daily real S&P 500 index data over 1928-2018 reveals strong evidence against parameter constancy over the whole sample period leading to a long duration of what the model characterizes as extreme behavior in real stock prices.

Chapter 4 develops a dynamic covariate-assisted spectral clustering method to uniformly estimate the latent group membership of cryptocurrencies consistently. We show that return inter-predictability and crypto characteristics, including hashing algorithms and proof types, jointly determine the crypto market segmentation. Based on this classification result, it is natural to employ eigenvector centrality to identify a cryptocurrency's idiosyncratic risk. An asset pricing analysis finds that a cross-sectional portfolio with a higher centrality earns a higher risk premium. Further tests confirm that centrality serves as a risk factor well and delivers valuable information content on cryptocurrency markets.

Table of Contents

Acknowledgements

I would like to express my deepest gratitude to my advisor, Professor Jun Yu, for his excellent guidance, caring, patience, and providing me with an excellent atmosphere for doing research. I appreciate all his contributions of time and ideas to make my Ph.D. experience productive and stimulating. The joy and enthusiasm he has for his research was contagious and motivational for me, even during tough times in the Ph.D. pursuit. Professor Yu led me into the door of academic research and set a role model for me.

I own a special debt to Professor Peter C.B. Phillips for his wonderful course on Econometrics, valuable feedback on my research. I hope that I could be as lively, enthusiastic, and energetic as him and to someday be able to command an audience as well as he can. I would like to thank Professor Yichong Zhang for his encouragement, insightful comments, and hard questions. My sincere thanks also go to Professor Liangjun Su, Professor Dashan Huang, Professor Jun Tu and Professor Anthony Tay for their kind helps during my stay at SMU.

I would like to give thanks to my friend Li Guo for the stimulating discussions, for the sleepless nights we were working together and for his devotion to our joint works during our PhD journey. I have learnt a lot from him about the empirical finance and FinTech.

Last but not least, I wish to thank my wife Yi Peng for her love and encouragement. She was always there cheering me up and stood by me through the good time and bad. My thanks also goes to my mother and my parents-in-law for their longstanding support and regretless giving.

To my father, Chunhua Tao, who had watched me through the PhD journey from heaven.

Chapter 1 Introduction

Nonstationary time-series and network data are most common data structures that practitioners come across in the real world and have attracted constant attentions from researcher in the past decade. This dissertation comprises three papers that solves three different problems in nonstationary times-series and network analysis.

In the first two essays, we focus on the time-series data and study the model selection and estimation problems for near-explosive time-series. Specifically, in Chapter 2, co-authored with Jun Yu[1], we study the model selection problem for time-series exhibit explosive or near-explosive feature. We develop an asymptotic theory for distinguishing explosive models with the unit root model using information criteria. Both OLS estimator and indirect inference estimator are studied in all three types of explosive models. We find that information criteria may consistently choose the unit root model under cases where the autoregressive coefficient depends on the sample size. Moreover, we prove that the validity of using information criteria for explosive model selection against unit root depends crucially on both the speed of auto-regressive coefficient converging to unity and the speed of information criteria specific penalty converging to zero as sample size goes to infinity.

In Chapter 3, co-authored with Peter Phillips[2] and Jun Yu, we study the estimation and inference problem in a continuous time dynamic system with a random persistence parameter proposed by Föllmer and Schweizer (1993). The model distinguishes various forms of unstable and explosive behaviour according to specific regions of the parameter space that open up the potential for testing these forms of

[1]School of Economics, Singapore Management University. Email: yujun@smu.edu.sg.
[2]Cowles Foundation for Research in Economics, Yale University. Email: peter.phillips@yale.edu.

extreme behaviour. We employ a two-stage realized volatility approach proposed by Phillips and Yu (2009) to estimate the model and construct test statistics to identify the different forms of extreme sample path behaviour. An empirical application based on daily real S&P 500 index data over 1964-2015 reveals strong evidence against parameter constancy after early 1980, which strengthens after July 1997, leading to a long duration of what the model characterizes as extreme behaviour in real stock prices.

In the last essay, my focus turns to analyzing network data and seeking its usefulness in financial application. In Chapter 4, co-authored with Li Guo[3] and Wolfgang Härdle[4], we develop a dynamic covariate-assisted spectral clustering method to provide uniformly consistent estimates of the latent group membership in a dynamic network setting. By constructing a dynamic network structure with cryptocurrencies' return inter-predictability relationship and their characteristics, we classify the cryptocurrencies into 4 groups over time. Based on this classification result, we employ eigenvector centrality to identify a cryptocurrency's idiosyncratic risk. The asset pricing analysis finds that a cross-sectional portfolio with a higher centrality earns a higher risk premium. Further tests confirm that centrality serves as a risk factor well and delivers valuable information content on cryptocurrency markets.

[3]Lee Kong Chian School of Business, Singapore Management University. Email: li.guo.2014@pbs.smu.edu.sg.

[4]School of Business and Economics, Humboldt-Universitt zu Berlin. Email: haerdle@hu-berlin.de.

Chapter 2 Model Selection for Explosive Models

2.1 Introduction

Information criteria have found a wide range of practical applications in empirical work. Examples include choosing explanatory variables in regression models and selecting lag lengths in time series models. Frequently used information criteria are AIC of Akaike (1969, 1974), BIC of Schwarz (1978), HQIC of Hannan and Quinn (1979). A major nice feature in these information criteria is that the penalty term is trivial to compute and hence the implementation of them is straightforward and can be made automatic.

With a growing interest in nonstationarity in time series analysis, researchers have examined the properties of information criteria in the context of nonstationary models with the unit root behavior. An important form of nonstationarity in time series involves explosive roots. Recent global financial crisis has motivated researchers to study explosive behavior in economic and financial time series; see, for example, Phillips and Yu (2011), Phillips et al. (2011) and Phillips et al. (2015a,b).

In this paper, we study the limit properties of information criteria for distinguishing between the unit root model and the explosive models. The information criteria considered in this paper have a general form and include AIC, BIC and HQIC as the special cases. The impact of the initial condition on the limit properties is examined by allowing for an initial condition of three different orders of magnitude. Moreover, both the OLS estimator and the indirect inference estimator are studied when

investigating the limit properties of information criteria. The motivation for the use of indirect inference estimator comes from the existence of finite sample bias in the OLS estimator and the ability that the indirect inference method can reduce the bias.

It is found that information criteria consistently choose the unit root model against the explosive alternatives when data comes from the unit root model. Second, we prove that the probability for information criteria to correctly select the explosive model models against the unit root model depends crucially on both the degree of explosiveness and the size of the penalty term in information criteria. Finally and surprisingly, we show that indirect inference estimation can increase or decrease the probability for information criteria to select the right model asymptotically relative to OLS, depending on the information criteria and the true model.

The rest of this paper is organized as follows. Section 2.2 introduces the models and information criteria, and briefly reviews the literature. Section 2.3 gives the limit properties of information criteria for distinguishing models with an explosive root from the unit root model when the OLS estimator is used. Section 2.4 gives the limit properties of information criteria when the indirect inference estimator is used. Section 2.5 provides Monte Carlo evidence to support the theoretical results. Section 2.6 concludes. All the detailed proofs are provided in the appendix. To compress notation, we denote $\int_0^1 BdB$ and $\int_0^1 B^2$ in short for $\int_0^1 B(r)dB(r)$ and $\int_0^1 B(r)^2dr$ respectively throughout the paper, and $\Rightarrow$ denotes weak convergence.

2.2 Models, Information Criteria and A Literature Review

The model considered in the present paper is of the form:

$$X_t = \rho_n X_{t-1} + u_t,\ t = 1, \cdots, n, \tag{2.2.1}$$

where $u_t \overset{iid}{\sim} (0, \sigma^2)$ and the model is initialized at $t = 0$ with some X_0. The autoregressive (AR) coefficient ρ_n is the crucial parameter that determines the dynamic behavior of X_t. When $\rho_n = \rho$ and $|\rho| < 1$, X_t is stationary. When $\rho_n = 1$, X_t has a unit root (UR hereafter). When $\rho_n = 1 - c_n/n = 1 - c/n$ for $c > 0$, X_t is near-stationary and has a root that is local-to-unity (LTUS hereafter) (Phillips, 1987b; Chan and Wei, 1987). When $\rho_n = \rho$ and $|\rho| > 1$, X_t has an explosive root (EX hereafter). When $\rho_n = 1 + c_n/n = 1 + c/n$ for $c > 0$, X_t is near-explosive and also has a root that is local-to-unity (LTUE hereafter). When $\rho_n = 1 - c_n/n$ for $c_n \to \infty$ but $c_n/n \searrow 0$, the root represents moderate deviations from unity and X_t is near-stationary (Phillips and Magdalinos, 2007). When $\rho_n = 1 + c_n/n$ for $c_n \to \infty$ but $c_n/n \searrow 0$, X_t is mildly explosive (hereafter ME).

The asymptotic properties of the OLS estimator of the AR coefficient in the stationary AR(1) model is well known. The rate of convergence is $\sqrt{n}$ and the limiting distribution is Gaussian. Phillips (1987a) provided the limiting theory for the OLS estimator in the UR model and the rate of convergence is n. Phillips (1987b) and Chan and Wei (1987) established the asymptotic theory for the LTUS and LTUE models. The asymptotic theory is similar to that in the UR model and the rate of convergence is also n. In the cases of UR and LTU, u_t can be weakly dependent stationary. Anderson (1959) studied the limiting distribution of the OLS estimator in the EX model under the condition that $u_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ and $X_0 = 0$. The limiting distribution is Cauchy and the rate of convergence is ρ^n. However, no invariance principle applies. Assuming $X_0 = o_p(\sqrt{n/c_n})$, Phillips and Magdalinos (2007) developed the asymptotic theory for the model with $\rho_n = 1 - c_n/n$ for $c_n \to \infty$ but $c_n/n \searrow 0$ and showed that the asymptotic distribution is invariant to the error distribution. The rate of convergence is $n/\sqrt{c_n}$. If $c_n = n^\alpha$ with $\alpha \in (0, 1)$, this rate of convergence bridges that of UR/LTU models and that of the stationary process. Phillips and Magdalinos (2007) also developed the asymptotic theory for the ME model. The rate of convergence is $n\rho_n^n/c_n$. The limiting distribution is Cauchy which is the same as in the EX model. Interestingly, in the ME case, the asymptotic

theory is independent of the initial condition as long as $X_0 = o_p(\sqrt{n/c_n})$.

It is known that the OLS estimator of ρ_n is biased downward when $\rho_n = 1$ or when ρ_n is in the vicinity of unity. In this case, the indirect inference estimation is effective in reducing the bias. Phillips (2012) derives the asymptotic theory of the indirect inference estimator when the model is UR or LTU and $u_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. The rate of convergence remains unchanged while the limiting distribution is different from that of the OLS estimator.

Information criteria for model selection have been proposed by Akaike (1969, 1974), Schwarz (1978), Hannan and Quinn (1979), among many others. The general form of these criteria is

$$IC_k = \log \widehat{\sigma}_k^2 + \frac{kp_n}{n},$$

where k is the number of parameters to be estimated, $\widehat{\sigma}_k^2$ is the estimated σ^2 when k parameters are estimated. In general, IC_k trades off the term that measures the goodness-of-fit (i.e. $\log \widehat{\sigma}_k^2$) and the penalty term that measures the complexity of the model (i.e. kp_n/n). Coefficient $p_n = 2, \log n, 2 \log \log n$ corresponds to AIC of Akaike (1974), BIC of Schwarz (1978) and HQIC of Hannan and Quinn (1979). Other forms of p_n are possible.

In the time series literature, information criteria have been widely used to select the lag length both in the family of stationary models and in the family of nonstationary models; see for example, Ng and Perron (1995) and Ploberger and Phillips (2003). The information criteria can also be used to evaluate whether $\rho_n = 1$ (i.e. $k = 0$) or $\rho_n \neq 1$ (i.e. $k = 1$) in Model (2.2.1). For example, Phillips (2008) obtained limit properties of IC_k for distinguishing between the unit root model and the stationary model. Phillips and Lee (2015) show that BIC can successfully distinguish the UR model from the ME model. This is a surprising result as it is well known that BIC cannot consistently distinguish between the UR model and the LTU model; see Ploberger and Phillips (2003).

In this paper we focus our attention to distinguishability between the unit root

model and the three explosive models (i.e., LTUE, ME and EX) after the candidate models are estimated by OLS or by the indirect inference method. As a result, we make contributions in two strands of literature, explosive time series and indirect inference.

To visually understand the difference between the UR model, the LTU model and the ME model, we simulate a sample path of different length ($n = 100, 200, 500, 1000$) with $y_0 = 0$, based on the same realizations of the error process, iid $\mathcal{N}(0, 1)$, from the following four models, $\rho_n = 1$ (UR), $\rho_n = 1 + 1/n$ (LTUE), $\rho_n = 1 + n^{0.1}/n$ (ME1), and $\rho_n = 1 + n^{0.5}/n$ (ME2). Figures 1-3 give the time series plot of UR against LTU, UR against ME1, UR against ME2, respectively. It can be seen from Figure 1 that it is very difficult to distinguish between the UR process and the LTU process, even when the sample size is as large as 1,000. When the sample size increases, the gap between the UR process and the two ME processes becomes larger and larger, as apparent in Figure 2 and more so in Figure 3.

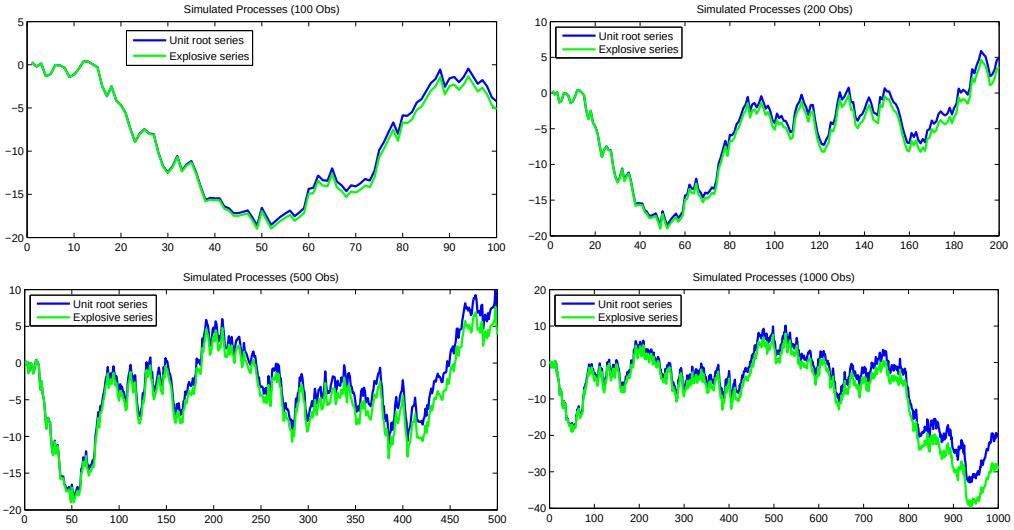


Figure 2.1: A realization of the UR model and the LTU model with $\rho_n = 1 + 1/n$.

2.3 Limit Properties Based on the OLS Estimator

When the data generating process (DGP) is the UR model, since $\rho_n = 1$, we set the parameter count to $k = 0$. For the LTU model, the ME model and the EX

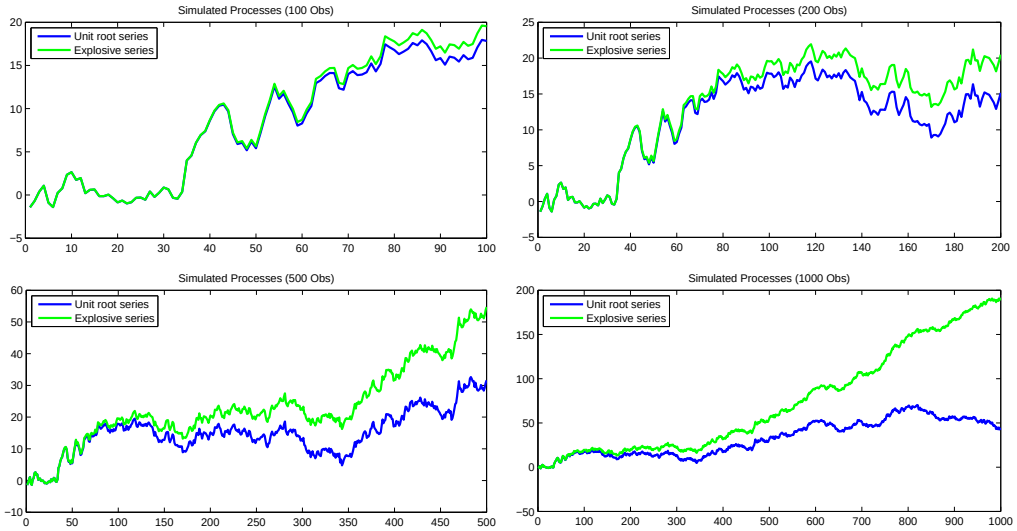


Figure 2.2: A realization of the UR model and the ME process with $\rho_n = 1+n^{0.1}/n$ (ME1).

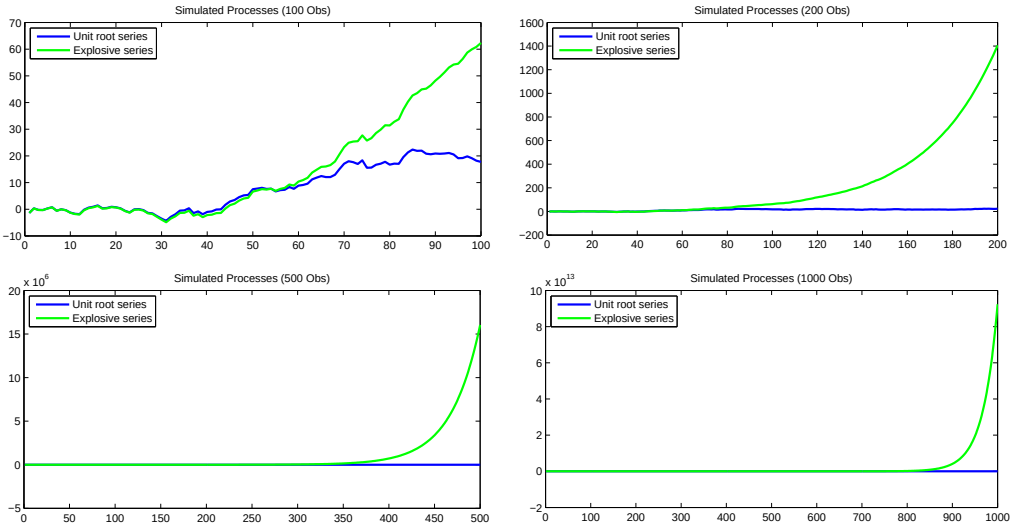


Figure 2.3: A realization of the UR model and the ME model with $\rho_n = 1+n^{0.5}/n$ (ME2).

model, we need to estimate the AR coefficient and hence set the parameter count to $k = 1$. Throughout the paper we denote $\widehat{\rho}$ the OLS estimator of ρ. $\widehat{k}_{IC} = 0$ or 1 means the information criterion of the UR model is smaller or larger than that of the competing model when ρ is estimated by OLS. We aim to find the limit of the following probabilities:

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0 | k = 0\right\}; \tag{2.3.1}$$

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1 | k = 0\right\}; \tag{2.3.2}$$

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0 | k = 1\right\}; \tag{2.3.3}$$

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1 | k = 1\right\}. \tag{2.3.4}$$

As shown in Phillips and Magdalinos (2009), the unit root asymptotic distribution is sensitive to initial conditions in the distant past. To understand how the initial condition affects the property of $\widehat{k}_{IC}$, we follow Phillips and Magdalinos (2009) by assuming alternative initial conditions.

Assumption 2.3.1 (IN) *The initial condition has the form*

$$X_0(n) = \sum_{j=0}^{\kappa_n} u_{-j}, \tag{2.3.5}$$

where κ_n is a sequence of integers satisfying $\kappa_n \to \infty$ and

$$\frac{\kappa_n}{n} \to \tau \in [0, \infty], \text{ as } n \to \infty. \tag{2.3.6}$$

The following cases are distinguished:

(i) If $\tau = 0$, $X_0(n)$ is said to be a recent past initialization.

(ii) If $\tau \in (0, \infty)$, $X_0(n)$ is said to be a distant past initialization.

(iii) If $\tau = \infty$, $X_0(n)$ is said to be an infinite past initialization.

Theorem 2.3.1 *Under Assumption 2.3.1 (i) or (ii) or (iii), we have*

(1) when $p_n \to \infty$ and $p_n/n \to 0$ as $n \to \infty$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0 | k = 0\right\} = \lim_{n\to\infty} P\left\{IC_0 - IC_1 \leq 0\right\} = 1,$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1 | k = 0\right\} = \lim_{n\to\infty} P\left\{IC_0 - IC_1 > 0\right\} = 0.$$

(2) when $p_n = 2$, the asymptotic distribution under the AIC criterion is

$$\lim_{n\to\infty} P\left\{\widehat{k}_{AIC} = 0 | k = 0\right\} = \lim_{n\to\infty} P\left\{AIC_0 - AIC_1 \leq 0\right\} = P\left(\xi^2 < 2\right),$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{AIC} = 1 | k = 0\right\} = \lim_{n\to\infty} P\left\{AIC_0 - AIC_1 > 0\right\} = 1 - P\left(\xi^2 < 2\right).$$

where

$$\xi^2 = \begin{cases} \dfrac{\left(\int_0^1 B dB\right)^2}{\int_0^1 B^2}, & \text{if } \tau = 0 \\ \dfrac{\left(\int_0^1 B_\tau dB\right)^2}{\int_0^1 B_\tau^2}, & \text{if } \tau \in (0,\infty) \\ B(1)^2, & \text{if } \tau = \infty \end{cases},$$

with $B(s)$ being a Brownian motion, and

$$B_\tau(s) = B(s) + \sqrt{\tau} B_0(1),$$

with $B_0(s)$ being an independent Brownian motion.

Remark 2.3.1 *Theorem 2.3.1 is the same as Theorem 1 in Phillips (2008) for distinguishing between the UR model and the stationary model. The condition that $p_n \to \infty$ and $p_n/n \to 0$ covers BIC and HQIC and hence, both BIC and HQIC can consistently select the UR model. The AIC criterion is inconsistent and its asymptotic distribution depends on ξ^2, the squared unit root t-statistic for the OLS estimator.*

Remark 2.3.2 *The validity of Theorem 2.3.1 does not require the iid assumption for the error term u_t. If we follow Phillips (2008) by denoting $F(L) = \sum_{j=0}^{\infty} F_j L^j$,*

with $F_0 = 1$ and $F(1) \neq 0$, and letting u_s have Wold representation

$$u_s = F(L)\varepsilon_s = \sum_{j=0}^{\infty} F_j \varepsilon_{s-j}, \text{ with } \sum_{j=0}^{\infty} j^{1/2} |F_j| < \infty, \tag{2.3.7}$$

where $\varepsilon_t \overset{iid}{\sim} (0, \sigma_\varepsilon^2)$, the results in Theorem 2.3.1 continue to hold. However, both B_0 and ξ^2 need to be modified to accommodate the dependence in u_t as in Phillips (2008).

Theorem 2.3.2 *Let Assumption 2.3.1 (i) or (ii) holds. Assume the true DGP is the LTUE model.*

(1) When $p_n \to \infty$ and $p_n/n \to 0$ as $n \to \infty$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0 | k = 1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}\left(IC_1 - IC_0\right) > 0\right\} = 1,$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1 | k = 1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}\left(IC_1 - IC_0\right) \leq 0\right\} = 0.$$

(2) When $p_n = 2$, the asymptotic distribution of the AIC criterion is

$$\lim_{n\to\infty} P\left\{\widehat{k}_{AIC} = 0 | k = 1\right\} = \lim_{n\to\infty} P\left\{n\left(AIC_1 - AIC_0\right) > 0\right\} = 1 - P\left(\zeta^2 > 2\right),$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{AIC} = 1 | k = 1\right\} = \lim_{n\to\infty} P\left\{n\left(AIC_1 - AIC_0\right) \leq 0\right\} = P\left(\zeta^2 > 2\right),$$

where

$$\zeta^2 = \frac{\left(\int_0^1 J_c dB\right)^2}{\int_0^1 J_c^2} + 2c\int_0^1 J_c dB + c^2 \int_0^1 J_c^2,$$

with

$$J_c(r) = \int_0^r \exp\left\{c(r-s)\right\} dB(s).$$

Remark 2.3.3 *Theorem 2.3.2 shows that all the information criteria are inconsistent in distinguishing between the LTUE model and the UR models when data comes from the LTUE model. AIC selects the wrong model with probability going to $1 - P(\zeta^2 > 2)$, which depends on the localization constant c. This problem*

worsens for BIC and HQIC as the probability of selecting the wrong model goes to one. Note that BIC is well known to be blind to local alternatives; see, for example, Ploberger and Phillips (2003).

Theorem 2.3.3 *Let Assumption 2.3.1 (i) or (ii) holds. Assume the true DGP is the ME model.*

(1) When $\lim_{n\to\infty} \frac{p_n}{\rho_n^{2n}} = 0$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}\left(IC_1 - IC_0\right) > 0\right\} = 0,$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}\left(IC_1 - IC_0\right) \leq 0\right\} = 1.$$

(2) When $\lim_{n\to\infty} \frac{p_n}{\rho_n^{2n}} = \pi \in (0, +\infty)$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}\left(IC_1 - IC_0\right) > 0\right\} = P\left(\chi^2(1) < 4\pi\right),$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}\left(IC_1 - IC_0\right) \leq 0\right\} = 1 - P\left(\chi^2(1) < 4\pi\right).$$

(3) When $\lim_{n\to\infty} \frac{p_n}{\rho_n^{2n}} \to +\infty$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}\left(IC_1 - IC_0\right) > 0\right\} = 1,$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}\left(IC_1 - IC_0\right) \leq 0\right\} = 0.$$

Remark 2.3.4 *Theorem 2.3.3 shows that the limit probability of selecting the correct model by information criteria under the ME model depends critically on two parameters,* c_n, p_n. *As expected, the larger* c_n, *the further the model away from the UR model and the higher probability for the information criteria to select the correct model. Interestingly, the smaller* p_n, *the higher probability for the information criteria to select the correct model. From Phillips and Magdalinos (2009), we know* $\rho_n^{-n} = o(c_n^{-1})$ *and hence* $\rho_n^n/c_n \to +\infty$. *In the special case where* $c_n = n^\alpha$,

for $\alpha \in (0,1)$, $\lim_{n\to\infty} p_n/\rho_n^{2n} = 0$ *no matter whether* $p_n = 2$ *or* $\log n$ *or* $2\log\log n$. *In this case, all the well-known information criteria can consistently select the true model.*

Theorem 2.3.4 *Let Assumption 2.3.1 (i) holds. Assume the true DGP is the EX model.*

(1) When $\lim_{n\to\infty} \frac{p_n}{\rho^{2n}} = 0$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0 | k = 1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho^{2n}}\left(IC_1 - IC_0\right) > 0\right\} = 0,$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1 | k = 1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho^{2n}}\left(IC_1 - IC_0\right) \leq 0\right\} = 1.$$

(2) When $\lim_{n\to\infty} \frac{p_n}{\rho^{2n}} = \pi \in (0, +\infty)$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0 | k = 1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho^{2n}}\left(IC_1 - IC_0\right) > 0\right\} = P\left(\chi^2(1) < (1+\rho)^2\pi\right),$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1 | k = 1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho^{2n}}\left(IC_1 - IC_0\right) \leq 0\right\} = 1 - P\left(\chi^2(1) < (1+\rho)^2\pi\right).$$

(3) When $\lim_{n\to\infty} \frac{p_n}{\rho^{2n}} \to +\infty$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0 | k = 1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}\left(IC_1 - IC_0\right) > 0\right\} = 1,$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1 | k = 1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}\left(IC_1 - IC_0\right) \leq 0\right\} = 0.$$

Remark 2.3.5 *Theorem 2.3.4 shows that the limit probability of selecting the correct model by information criteria under the EX model depends also critically on two parameters,* ρ, p_n. *As expected, the larger* ρ, *the higher probability for the information criteria to select the correct model. Interestingly, the smaller* p_n, *the higher probability for the information criteria to select the correct model. If* $p_n = 2$ *or* $\log n$ *or* $2\log\log n$, $\lim_{n\to\infty} p_n/\rho^{2n} = 0$ *and hence case (1) applies, suggesting that all the well-known information criteria can consistently select the true model.*

Results in Theorem 2.3.3 can be extended to cover the LTUE model and the ME model with weakly dependent errors. The following proposition establishes the results for the ME model.

Proposition 2.3.1 *Let Assumption 2.3.1 (i) or (ii) and the assumption specified in Equation (2.3.7) hold. Assume the true DGP is the ME model.*

(1) When $\lim_{n\to\infty} \frac{p_n}{\rho_n^{2n}} = 0$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}\left(IC_1 - IC_0\right) > 0\right\} = 0,$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}\left(IC_1 - IC_0\right) \leq 0\right\} = 1.$$

(2) When $\lim_{n\to\infty} \frac{p_n}{\rho_n^{2n}} = \pi \in (0, +\infty)$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}\left(IC_1 - IC_0\right) > 0\right\} = P\left(\chi^2(1) < \frac{4\pi}{\omega^2}\right),$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}\left(IC_1 - IC_0\right) \leq 0\right\} = 1 - P\left(\chi^2(1) < \frac{4\pi}{\omega^2}\right).$$

where $\omega^2 = \left(\sum_{j=0}^{\infty} F_j\right)^2$.

(3) When $\frac{p_n}{\rho_n^{2n}} \to +\infty$,

$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}\left(IC_1 - IC_0\right) > 0\right\} = 1,$$
$$\lim_{n\to\infty} P\left\{\widehat{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}\left(IC_1 - IC_0\right) \leq 0\right\} = 0.$$

2.4 Limit Properties Based on the Indirect Inference Estimator

The OLS estimator of ρ_n in Model (2.2.1) is known to be biased and the bias is acute when ρ_n is close to unity. To reduce the bias, the indirect inference method

of Smith (1993) and Gourieroux et al. (1993) can be used if Model (2.2.1) is fully specified. Phillips (2012) derives the asymptotic theory of the indirect inference estimator when the model is UR or LTU and $u_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Throughout the paper we denote $\breve{\rho}$ the indirect inference estimator of ρ. Let $h(c) = c + g(c)$ and $g(c) = g^-(c)1_{\{c\leq 0\}} + g^+(c)1_{\{c>0\}}$ with

$$\begin{aligned}
g^-(c) =& -\frac{3}{4}\int_0^\infty e^{-\frac{v}{4}} k^-(v;c)^{1/2} dv + \frac{1}{4}\int_0^\infty e^{-\frac{v}{4}} k^-(v;c)^{3/2} dv \\
& - \frac{e^{2c}}{8}\int_0^\infty e^{-\frac{5v}{4}} k^-(v;c)^{3/2} v dv, \\
g^+(c) =& \frac{3}{4}\int_0^\infty e^{\frac{w}{4}} k^+(w;c)^{1/2} dw - \frac{1}{4}\int_0^\infty e^{\frac{w}{4}} k^+(w;c)^{3/2} dw \\
& - \frac{e^{2c}}{8}\int_0^\infty e^{\frac{5w}{4}} k^+(w;c)^{3/2} w dw, \\
k^-(v;c) =& \frac{2v - 4c}{v + e^{2c} v e^{-v} - 4c}, \\
k^+(w;c) =& \frac{2w + 4c}{w + e^{2c} w e^{w} + 4c}.
\end{aligned}$$

Phillips (2012) shows that under the UR model,

$$n\left(\breve{\rho} - 1\right) \Rightarrow h^{-1}\left(\int_0^1 B dB / \int_0^1 B^2\right) \text{ as } n \to +\infty,$$

and under the LTUE model,

$$n\left(\breve{\rho} - \rho_n\right) \Rightarrow h^{-1}\left(\int_0^1 J_c dB / \int_0^1 J_c^2 + c\right) - c \text{ as } n \to +\infty.$$

Let $\breve{k}_{IC} = 0$ or 1 mean the information criterion of the UR model is smaller or larger than that of the competing model when the model is estimated by the indirect

inference method. We aim to find is the limit of the following probabilities:

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC}=0|k=1\right\}; \tag{2.4.1}$$

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC}=1|k=1\right\}; \tag{2.4.2}$$

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC}=0|k=0\right\}; \tag{2.4.3}$$

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC}=1|k=0\right\}. \tag{2.4.4}$$

Theorem 2.4.1 *Under Assumption 2.3.1(i) or (ii) or (iii), we have*

(1) when $p_n \to \infty$ and $p_n/n \to 0$ as $n \to \infty$,

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC}=0|k=0\right\} = \lim_{n\to\infty} P\left\{IC_0 - IC_1 \leq 0\right\} = 1,$$
$$\lim_{n\to\infty} P\left\{\breve{k}_{IC}=1|k=0\right\} = \lim_{n\to\infty} P\left\{IC_0 - IC_1 > 0\right\} = 0;$$

(2) when $p_n = 2$, the asymptotic distribution under the AIC criterion is

$$\lim_{n\to\infty} P\left\{\breve{k}_{AIC}=0|k=0\right\} = P\left(\varsigma^2 < 2\right),$$
$$\lim_{n\to\infty} P\left\{\breve{k}_{AIC}=1|k=0\right\} = 1 - P\left(\varsigma^2 < 2\right),$$

where

$$\varsigma^2 = \begin{cases} \int_0^1 B^2 \cdot h^{-1}\left(\left(\frac{\int_0^1 BdB}{\int_0^1 B^2}\right)^2\right) - 2\int_0^1 BdB \cdot h^{-1}\left(\frac{\int_0^1 BdB}{\int_0^1 B^2}\right), & \text{if } \tau = 0, \\ \int_0^1 B_\tau^2 \cdot h^{-1}\left(\left(\frac{\int_0^1 B_\tau dB}{\int_0^1 B_\tau^2}\right)^2\right) - 2\int_0^1 B_\tau dB \cdot h^{-1}\left(\frac{\int_0^1 B_\tau dB}{\int_0^1 B_\tau^2}\right), & \text{if } \tau \in (0,\infty), \\ h^{-1}\left(\mathcal{C}\right)^2 B_0^2(1) - 2h^{-1}\left(\mathcal{C}\right) B(1)B_0(1), & \text{if } \tau = \infty, \end{cases}$$

with $\mathcal{C}$ being a standard Cauchy variate.

Remark 2.4.1 *According to Theorem 2.4.1, as long as $p_n \to \infty$ and $p_n/n \to 0$, information criteria based on the indirect inference estimator is consistent in*

selecting the UR model. Hence, BIC and HQIC based on the indirect inference estimator can consistently select the UR model. Like the AIC criterion that is based on the OLS estimator, the AIC criterion based on the indirect inference estimator continues to be inconsistent. However, its asymptotic distribution depends on ς^2, the squared unit root t-statistic for the indirect inference estimator.

Remark 2.4.2 *As shown in Phillips (2012), the squared unit root t-statistic for the indirect inference estimator has a smaller variance than that of the squared unit root t-statistic for the OLS estimator. Consequently, $P(\varsigma^2 < 2) > P(\xi^2 < 2)$, suggesting that AIC based on the indirect inference estimator can select the true model (i.e. the UR model) with a larger probability than that based on the OLS estimator.*

Theorem 2.4.2 *Let Assumption 2.3.1 (i) or (ii) holds. Assume the true DGP is the LTUE model.*

(1) When $p_n \to \infty$ and $p_n/n \to 0$ as $n \to \infty$,

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 0 | k = 1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}(IC_1 - IC_0) > 0\right\} = 1,$$
$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 1 | k = 1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}(IC_1 - IC_0) \leq 0\right\} = 0.$$

(2) When $p_n = 2$, the asymptotic distribution under the AIC criterion is

$$\lim_{n\to\infty} P\left\{\breve{k}_{AIC} = 0 | k = 1\right\} = \lim_{n\to\infty} P\left\{n(AIC_1 - AIC_0) > 0\right\} = 1 - P\left(\vartheta^2 > 2\right),$$
$$\lim_{n\to\infty} P\left\{\breve{k}_{AIC} = 1 | k = 1\right\} = \lim_{n\to\infty} P\left\{n(AIC_1 - AIC_0) \leq 0\right\} = P\left(\vartheta^2 > 2\right),$$

where

$$\vartheta^2 \equiv 2h^{-1}\left(\frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c\right)\left(\int_0^1 J_c dB + c\int_0^1 J_c^2\right) - h^{-1}\left(\frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c\right)^2 \int_0^1 J_c^2.$$

Remark 2.4.3 *Theorem 2.4.2 shows that all the information criteria continue to be inconsistent in distinguishing between the LTUE model and the UR models when*

data come from the LTUE model even when the indirect inference estimation is employed. AIC selects the wrong model with probability going to $1 - P(\vartheta^2 > 2)$. *Since the variance of* ζ^2 *is bigger than that of* ϑ^2, *the tail probability of* ζ^2 *is larger than that of* ϑ^2, *suggesting that AIC based on OLS selects the true model (i.e. LTUE model) with a greater probability than AIC based on the indirect inference estimator. This is a rather surprising result and suggests that the superiority in estimation does not necessarily translate to the superiority in model selection.*

Theorem 2.4.3 *Let Assumption 2.3.1 (i) or (ii) holds. Assume the true DGP is the ME model.*

(1) When $\lim_{n\to\infty} \frac{p_n}{\rho_n^{2n}} = 0$,

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}(IC_1 - IC_0) > 0\right\} = 0,$$
$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}(IC_1 - IC_0) \leq 0\right\} = 1.$$

(2) When $\lim_{n\to\infty} \frac{p_n}{\rho_n^{2n}} = \pi \in (0, +\infty)$,

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}(IC_1 - IC_0) > 0\right\} = P\left(\chi^2(1) < 4\pi\right),$$
$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho_n^{2n}}(IC_1 - IC_0) \leq 0\right\} = 1 - P\left(\chi^2(1) < 4\pi\right).$$

(3) When $\frac{p_n}{\rho_n^{2n}} \to +\infty$,

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}(IC_1 - IC_0) > 0\right\} = 1,$$
$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}(IC_1 - IC_0) \leq 0\right\} = 0.$$

Remark 2.4.4 *The results in Theorem 2.4.3 are the same as those in Theorem 2.3.3, suggesting all the well-known information criteria can consistently select the true model (i.e. ME model) when* $c_n = n^\alpha$, *for* $\alpha \in (0, 1)$.

Theorem 2.4.4 *Let Assumption 2.3.1 (i) holds. Assume the true DGP is the EX model.*

(1) When $\lim_{n\to\infty} \frac{p_n}{\rho^{2n}} = 0$,

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho^{2n}}\left(IC_1 - IC_0\right) > 0\right\} = 0,$$
$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho^{2n}}\left(IC_1 - IC_0\right) \leq 0\right\} = 1.$$

(2) When $\lim_{n\to\infty} \frac{p_n}{\rho^{2n}} = \pi \in (0, +\infty)$,

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho^{2n}}\left(IC_1 - IC_0\right) > 0\right\} = P\left(\chi^2(1) < (1+\rho)^2\pi\right),$$
$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{\rho^{2n}}\left(IC_1 - IC_0\right) \leq 0\right\} = 1 - P\left(\chi^2(1) < (1+\rho)^2\pi\right).$$

(3) When $\lim_{n\to\infty} \frac{p_n}{\rho^{2n}} \to +\infty$,

$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 0|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}\left(IC_1 - IC_0\right) > 0\right\} = 1,$$
$$\lim_{n\to\infty} P\left\{\breve{k}_{IC} = 1|k=1\right\} = \lim_{n\to\infty} P\left\{\frac{n}{p_n}\left(IC_1 - IC_0\right) \leq 0\right\} = 0.$$

Remark 2.4.5 *The results in Theorem 2.4.4 are the same as those in Theorem 2.3.4, suggesting that all the well-known information criteria can consistently select the true model (i.e. EX model).*

2.5 Monte Carlo Study

In this section, we examine the performance of alternative information criteria, namely, AIC, BIC and HQIC, in finite sample via simulated data and check the reliability of the asymptotic results developed in Section 3 and Section 4. In the simulation study, we use both OLS and the indirect inference method to estimate ρ_n from sample paths that are simulated from different DGPs. In total we design

four experiments. In the first experiment we simulate data from the UR model. In the second experiment we simulate data from the LTUE model with $c = 1$ (i.e. $\rho_n = 1 + 1/n$). In the third experiment we simulate data from two ME models with $c_n = n^{0.1}$, $n^{0.3}$, respectively. In the last experiment we simulate data from the EX model with $\rho = 1.01, 1.05$, respectively. In all experiments, we simulate 10,000 sample paths with initial value $X_0 = 0$ and four sample sizes are considered, $n = 100, 200, 500, 1000$. In each experiment, we report the fraction of the number of times in which the correct model is selected out of 10,000 replications.

Table 2.1 reports the results when the true DGP is UR. Several results can be found here. First, the probability for BIC and HQIC to select the true model grows as n grows. However, the probability for AIC to select the true model does not seem to increase or decrease as n grows. This observation is consistent with the asymptotic results reported in Theorem 2.3.1. Second, the probability for BIC to select the true model is larger than that in HQIC which is in turn larger than AIC in these four sample sizes. So we can conclude that the probability grows as p_n increases since $2 < 2\log\log n < \log n$ when $100 \leq n \leq 1000$. Third, the probability implied by AIC based on the indirect inference estimator is larger than that based on OLS. This finding is consistent with Theorem 2.4.1 and Remark 2.4.2.

Table 2.1: Probability of Selecting the Correct Model when Data Come from the UR Model

n	100			200		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
OLS	0.8160	0.9604	0.9020	0.8155	0.9751	0.9249
IIE	0.8731	0.9702	0.9292	0.8742	0.9810	0.9445
n	500			1000		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
OLS	0.8127	0.9849	0.9335	0.8195	0.9895	0.9402
IIE	0.8704	0.9881	0.9508	0.8759	0.9918	0.9566

Table 2.2 report the results when the true DGP is the LTUE model with $c_n = 1$. Also reported is the value of p_n/ρ_n^{2n}. Several results can be found here. First, the probability for BIC and HQIC to select the true model becomes smaller as n grows.

However, the probability for AIC to select the true model does not seem to increase or decrease as n grows. This observation is consistent with the asymptotic results in Theorem 2.3.2. Second, the probability implied by AIC based on the indirect inference estimator is smaller than that based on OLS. This finding is consistent with in Theorem 2.4.2 and Remark 2.4.3. Finally, it seems that AIC performs better than BIC and HQIC in all cases.

Table 2.2: Probability of Selecting the Correct Model when Data Come from the LTUE Model with $c_n = 1$

n	100			200		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	0.2734	0.6295	0.4175	0.2720	0.7206	0.4536
OLS	0.3516	0.1475	0.2420	0.3406	0.1305	0.2156
IIE	0.1485	0.0445	0.0922	0.1235	0.0269	0.0663
n	500			1000		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	0.2712	0.8427	0.4955	0.2709	0.9358	0.5236
OLS	0.3474	0.1019	0.1933	0.3416	0.0871	0.1823
IIE	0.1169	0.0134	0.0517	0.1089	0.0090	0.0394

Table 2.3 report the results when the true DGP is the ME model with $c_n = n^{0.1}, n^{0.3}$. Also reported is the value of p_n/ρ_n^{2n}. Several results can be found here. First, the probability for all three information criteria to select the true model grows as n increases. This observation is consistent with the asymptotic results reported in Theorem 2.3.3 and Remark 2.4.4. Second, comparing the results for $c_n = n^{0.1}$ and those for $c_n = n^{0.3}$, the probability for all three information criteria to select the true model increases when c_n is bigger. Third, the probability based on the indirect inference estimator is smaller than that based on OLS. Finally, it seems that AIC performs better than BIC and HQIC in all cases.

Table 2.4 report the results when the true DGP is the EX model with $\rho = 1.01, 1.05$. Also reported is the value of p_n/ρ^{2n}. Several results can be found here. First, when $\rho = 1.01$, which is larger than the unity by 1%, the probability for information criteria to select the correct model is small in all cases when the sample size

Table 2.3: Probability of Selecting the Correct Model when Data Come from the ME Model with $c_n = n^{0.1}$ and $c_n = n^{0.3}$

ME Model with $c_n = n^{0.1}$						
n	100			200		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	0.0861	0.1983	0.1316	0.0679	0.1799	0.1132
OLS	0.5183	0.3403	0.4349	0.5554	0.3638	0.4629
IIE	0.3071	0.1741	0.2406	0.3211	0.1624	0.2250
n	500			1000		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	0.0486	0.1512	0.0889	0.0371	0.1282	0.0718
OLS	0.6151	0.4083	0.5048	0.6469	0.4374	0.5494
IIE	0.3544	0.2008	0.2815	0.3925	0.2351	0.3129
ME Model with $c_n = n^{0.3}$						
n	100			200		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	0.0008	0.0019	0.0012	0.0001	0.0003	0.0002
OLS	0.9374	0.9066	0.9235	0.9749	0.9608	0.9683
IIE	0.9274	0.8979	0.9163	0.9716	0.9578	0.9648
n	500			1000		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ_n^{2n}	1.0e-06	1.0e-05	1.0e-06	1.0e-07	1.0e-07	1.0e-07
OLS	0.9948	0.9907	0.9938	0.9988	0.9985	0.9986
IIE	0.9938	0.9901	0.9933	0.9986	0.9985	0.9985

is small. However, it grows very quickly with the sample size. When $\rho = 1.05$, the probability for information criteria to select the correct model is almost 1 in all cases even when the sample size is small and increases with the sample size. Finally, it seems that AIC performs better than BIC and HQIC in all cases.

Table 2.4: Probability of Selecting the Correct Model when Data Come from the Regular Explosive Model with $\rho = 1.01, 1.05$.

Explosive Model with $\rho = 1.01$						
n	100			200		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ^{2n}	0.2734	0.6295	0.4175	0.0374	0.0990	0.0623
OLS	0.3516	0.1475	0.2420	0.6449	0.4820	0.5555
IIE	0.1485	0.0445	0.0922	0.4740	0.3059	0.3845
n	500			1000		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ^{2n}	1.0e-4	1.0e-4	1.0e-4	1.0e-9	1.0e-8	1.0e-9
OLS	0.9775	0.9599	0.9704	0.9998	0.9997	0.9998
IIE	0.9733	0.9563	0.9681	0.9998	0.9997	0.9998
Explosive Model with $\rho = 1.05$						
n	100			200		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ^{2n}	0.0001	0.0003	0.0002	1.0e-07	1.0e-07	1.0e-07
OLS	0.9741	0.9643	0.9681	0.9999	0.9998	0.9998
IIE	0.9703	0.9626	0.9655	0.9999	0.9998	0.9998
n	500			1000		
IC	AIC	BIC	HQIC	AIC	BIC	HQIC
p_n/ρ^{2n}	1.0e-20	1.0e-20	1.0e-20	1.0e-41	1.0e-41	1.0e-41
OLS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
IIE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

2.6 Conclusion

This paper studies the limit properties of information criteria for distinguishing between unit root model and three types of explosive models. Both the OLS estimator and the indirect inference estimator are employed to estimate the AR coefficient in the candidate model. This paper contributes to the literature in three aspects. First, our results extends results in the literature to the explosive side of the unit root, and

we find that information criteria consistently choose the unit root model when the unit root model is the true model. Second, we show that the limiting probabilities for information criteria to select the explosive model depends on both the distance of autoregressive coefficient from unity and the size of penalty term in the information criteria. When the penalty term is not too large and the root is not too close to unit root, all the information criteria consistently select the true model. It is known that the indirect inference method is effective in reducing the bias in OLS estimation in all cases as well as reducing the variance in OLS estimation in the UR model and in the LTU model. However, when information criteria are used in connection with the indirect inference estimation, the limiting probabilities for information criteria to select the correct model can go up or down relative to that with the OLS estimation, depending on the true DGP. When the true DGP is the UR model, the indirect inference estimation increases the probability. When the true DGP is the LTUE model or the ME model or the EX model, the indirect inference estimation decreases the probability. This rather surprising result suggests that the superiority in estimation does not necessarily translate to the superiority in model selection.

Chapter 3 Random Coefficient Continuous Systems: Testing for Extreme Sample Path Behaviour

3.1 Introduction

Many macroeconomic and financial time series are well described by autoregressive processes with roots that are close to unity but not necessarily constant over time. Motivated by this empirical characteristic, various strands of the literature have sought to extend pure unit root models to more flexible dynamic systems. One approach allows for structural breaks in which the autoregressive coefficient takes a constant value in each regime but changes value in different regimes (e.g. Chong, 2001; Pang et al., 2014; Jiang et al., 2017). Another assumes that the autoregressive coefficient is a continuous random variable or evolves according to a stochastic process (e.g. Granger and Swanson, 1997; Lieberman and Phillips, 2014, 2017c). Yet another allows for a time varying autoregressive parameter to capture evolution in the stochastic process, introduce flexibility, and enhance forecasting capability (Bykhovskaya and Phillips, 2018, 2019; Giraitis et al., 2014; Kristensen, 2012).

Complementary to this literature on autoregressive specification is a growing interest in modelling explosive behavior and collapse, particularly since the events leading up to and following the global financial crisis, where strong upward movements and subsequent major downturns in asset prices have occurred in various markets (Phillips and Yu, 2011). Empirical methods used to model these events

have made extensive use of the concepts of mildly explosive and mildly integrated autoregressive processes (see Phillips and Magdalinos, 2007). Thus, Phillips et al. (2011, PWY hereafter), Phillips and Yu (2011), Phillips et al. (2015a,b, PSY hereafter) assume data are generated according to unit root processes in one regime and as mildly explosive processes in another regime; and methods of date-stamping such regime changes have been developed (Phillips et al., 2011, 2015a,b) stimulating new empirical research and improvements in test methodology (e.g. Cavaliere et al., 2016; Phillips and Shi, 2017). Developments in random autoregressive coefficient approaches have also been pursued, with work by Aue (2008), who analyzed a near-integrated random coefficient autoregressive model, and by Banerjee et al. (2017) who studied a near-explosive random coefficient autoregressive model.[1]

The present paper contributes to this literature by working with a continuous time model in which the parameter that measures persistence is randomized. A novel advantage arising from this formulation is that extreme sample path behavior can be classified into distinct scenarios that represent various forms of instability and explosiveness. These scenarios are distinguished parametrically and corresponding hypotheses are formulated to facilitate empirical testing. Continuous time specification also enables the localizing coefficients that appear in mildly integrated and mildly explosive processes to be represented in terms of sampling frequency, which facilitates econometric estimation. These parameters are of great importance empirically because they control distance from martingale and unit root behavior in discrete time models (Banerjee et al., 2017). This advantage of continuous systems has been used in other recent work by Chen et al. (2017) and Wang and Yu (2016) in developing the discrete time methodology of Phillips and Magdalinos (2007).

Continuous system formulation and high frequency data open up the opportunity to employ methods such as realized volatility in estimating parameters that are identified in the quadratic variation process using in-fill asymptotic methods. The

[1]Considerable work has been done on discrete time random coefficient autoregressive models in the literature, including Aue et al. (2006), Berkes et al. (2009), Aue and Horváth (2011), and Horváth and Trapani (2016) among many others.

two-stage realized volatility approach employed here naturally accommodates heteroskedasticity in the process and allows for consistent estimation of the parameters in the diffusion function under both stationarity and explosiveness. The approach therefore offers potential for a unified in-fill limit theory of consistent parameter estimation in random coefficient autoregression.

A further well-known feature of continuous system formulations is that the effects of initial conditions are naturally incorporated by in-fill asymptotics (as in Phillips, 1987b) without having to specify orders of magnitude or use distant past representations (as in Phillips and Magdalinos, 2009) which involve additional unknown parameters. Moreover, continuous systems readily accommodate endogeneity by allowing for dependence between the random coefficient elements and system shocks. In this respect the present research relates to recent work on generalized random coefficient autoregressive models (Hwang and Basawa, 1998) and localized endogenous stochastic unit root models in (Lieberman and Phillips, 2017b). Initial condition effects appear directly in the asymptotic theory and, as is shown in the paper, the endogeneity parameter can be consistently estimated using realized volatility.

The continuous time model used in the present study is a special case of a financial market model developed in Föllmer and Schweizer (1993) obtained by applying an invariance principle to a discrete time market equilibrium model derived from first principles. In particular, Föllmer and Schweizer (1993) developed a microeconomic model of rational expectations equilibrium for a market that involves both information traders and noise traders. They showed that when the proportions of different types of traders fluctuates randomly the equilibrium outcome is a discrete time model with a random coefficient. The mapping from the theory model implies that noise traders contribute positively to the random coefficient whereas information traders contribute negatively. Correspondingly, the ratio of trader types affects the recurrence or transient properties of the resulting price process, thereby impacting the nature of the resulting price trajectories. Thus, the extent of randomness

in the coefficient reflects underlying market composition characteristics and is informative about the respective trader proportions. Econometric estimation of such models can therefore help to shed light on some of these properties and possibly also the changing nature of the market trader composition.

The remainder of the paper is organized as follows. Section 3.2 introduces a continuous system with randomized persistence and relates this system to several discrete time models already used in the literature. The multiple forms of behavior induced by this system are described and characterized parametrically. Section 3.3 proposes a novel two-stage approach to parameter estimation using realized volatility. Asymptotic theory is developed and test statistics for distinguishing different forms of explosive behavior are proposed in Section 3.4. Section 3.5 extends the methodology to the case of endogenous persistence. Section 3.6 gives the results of Monte Carlo simulations that explore the finite sample performance of the estimators and test statistics. Empirical applications of the model are reported in Section 3.7 using daily real S&P 500 index data from December 1927 to June 2018. Some empirical applications of the extended model using 5-minute real S&P 500 index data over the period from November 1, 1997 to October 31, 2013 are also discussed. Section 3.8 concludes. Proofs and other technical material are given in the Appendix. Additional simulation and empirical results can be found in an online supplement.

3.2 The Model

The model used here is a modified version of the Ornstein-Uhlenbeck process

$$dy(t) = y(t)\widetilde{\mu}dt + \sigma dB_{\varepsilon}(t), y(0) = y_0. \tag{3.2.1}$$

where B_{ε} is a standard Brownian motion and the sign of the drift parameter $\widetilde{\mu}$ determines stationary (< 0), nonstationary ($= 0$), and explosive (> 0) behavior in $y(t)$, the latter corresponding to a discrete time autoregression with a root that exceeds

unity and whose variance grows exponentially with t. In (3.2.1), the drift parameter $\widetilde{\mu}$ is taken as constant, an assumption that may not be well supported by data over extended periods of time.

The model considered in the present paper extends (3.2.1) by introducing random shocks to the drift component of (3.2.1) so that

$$dy(t) = y(t)\left[\widetilde{\mu}dt + \widetilde{\sigma}dB_u(t)\right] + \sigma dB_\varepsilon(t), y(0) = y_0, \tag{3.2.2}$$

where $B_u(t)$ and $B_\varepsilon(t)$ are both standard Brownian motions, and y_0 is independent of $B_u(t)$ and $B_\varepsilon(t)$. When $\widetilde{\sigma}^2 \neq 0$, model (3.2.2) may be viewed as an Ornstein-Uhlenbeck process with randomized drift or persistence. Initially, we focus on the case of independent noise processes $B_u(t)$ and $B_\varepsilon(t)$, and later consider the endogenous case where these processes are dependent.

Model (3.2.2) is a special case of a general model introduced by Föllmer and Schweizer (1993),

$$dy(t) = y(t)\left[\widetilde{\mu}(t)dt + \widetilde{\sigma}(t)dB_u(t)\right] + \mu(t)dt + \sigma(t)dB_\varepsilon(t), y(0) = y_0, \tag{3.2.3}$$

called an Ornstein-Uhlenbeck process in a random environment. Föllmer and Schweizer (1993) developed a discrete time version of this process in a market equilibrium setting that involved both information traders and noise traders and then derived its continuous-time limit given by the process in (3.2.3). Persistence in the dynamic model is determined by the relative proportions of the two types of traders, so random proportions lead to a randomized degree of persistence in the solution. Information traders contribute negatively to persistence while noise traders contribute positively.[2]

Föllmer and Schweizer (1993) derived the strong solution of (3.2.2) which takes

[2]Granger (1980) showed how simple cross section aggregation of random coefficient AR processes can – under certain conditions – generate long memory in the aggregated series. In Föllmer and Schweizer (1993) the aggregation is more complex with time varying weights; see Equation (3.6) in Föllmer and Schweizer (1993).

the explicit form

$$y(t) = \exp\left(\widetilde{\sigma} B_u(t) + \left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right) t\right) y(0) + K(t), \tag{3.2.4}$$

where

$$\begin{aligned} K(t) &= \sigma \int_0^t \exp\left(\widetilde{\sigma}\left(B_u(t) - B_u(s)\right) + \left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(t-s)\right) dB_\varepsilon(s) \\ &\sim \mathcal{MN}\left(0, \sigma^2 \int_0^t e^{2\widetilde{\sigma}(B_u(t)-B_u(s))+2\left(\widetilde{\mu}-\frac{1}{2}\widetilde{\sigma}^2\right)(t-s)} ds\right) \end{aligned} \tag{3.2.5}$$

under independence of B_u and B_ε and with

$$E\left\{K(t)^2\right\} = \sigma^2 E\left\{\int_0^t e^{2\widetilde{\sigma}(B_u(t)-B_u(s))+2\left(\widetilde{\mu}-\frac{1}{2}\widetilde{\sigma}^2\right)(t-s)} ds\right\} = \frac{\sigma^2}{2}\frac{e^{2\left(\widetilde{\mu}+\frac{1}{2}\widetilde{\sigma}^2\right)t} - 1}{\left(\widetilde{\mu}+\frac{1}{2}\widetilde{\sigma}^2\right)}. \tag{3.2.6}$$

Notably, $E\left\{K(t)^2\right\}$ diverges exponentially when $\widetilde{\mu} + \frac{1}{2}\widetilde{\sigma}^2 > 0$.

The exact discrete time model corresponding to (3.2.2) at the sampling interval Δ follows directly from the strong solution and has the explicit form

$$\begin{aligned} y_{t\Delta} = &\exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)\Delta + \widetilde{\sigma}\left[B_{u,t\Delta} - B_{u,(t-1)\Delta}\right]\right\} y_{(t-1)\Delta} \\ &+ \sigma \int_{(t-1)\Delta}^{t\Delta} \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(t\Delta - s) + \widetilde{\sigma}\left[B_{u,t\Delta} - B_u(s)\right]\right\} dB_\varepsilon(s), \end{aligned} \tag{3.2.7}$$

where $t = 1, ..., T/\Delta$, T is the time span, and we write discrete time data in subscripted form. This model is a random coefficient autoregression (RCAR) of the type considered by Nicholls and Quinn (1980) in which the autoregressive (AR) coefficient is

$$\rho_{t\Delta} = \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)\Delta + \widetilde{\sigma}\left[B_{u,t\Delta} - B_{u,(t-1)\Delta}\right]\right\}, \tag{3.2.8}$$

and is random when $\widetilde{\sigma}^2 > 0$.

For the ensuing development it will be helpful to fix the following simpler nota-

tions for the discrete system

$$\phi := \widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2, \quad \kappa := \widetilde{\mu} + \frac{1}{2}\widetilde{\sigma}^2, \;\; u_{t\Delta} := \frac{B_{u,t\Delta} - B_{u,(t-1)\Delta}}{\sqrt{\Delta}} \sim \mathcal{N}(0,1),$$
$$\rho_{t\Delta} := \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)\Delta + \widetilde{\sigma}\left[B_{u,t\Delta} - B_{u,(t-1)\Delta}\right]\right\} = \exp\left\{\phi\Delta + \widetilde{\sigma}\sqrt{\Delta}u_{t\Delta}\right\},$$
$$\eta_{t\Delta} := \int_{(t-1)\Delta}^{t\Delta} \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(t\Delta - s) + \widetilde{\sigma}\left[B_{u,t\Delta} - B_{u,s}\right]\right\} dB_{\varepsilon,s} \sim \mathcal{N}\left(0, \gamma_\Delta^2\right),$$

where $\gamma_\Delta = \sqrt{\left(e^{2\kappa\Delta} - 1\right)/2\kappa}$. Model (3.2.7) is then

$$y_{t\Delta} = \exp\left\{\phi\Delta + \widetilde{\sigma}\sqrt{\Delta}u_{t\Delta}\right\} y_{(t-1)\Delta} + \sigma\eta_{t\Delta} = \rho_{t\Delta}y_{(t-1)\Delta} + \sigma\eta_{t\Delta}, \qquad (3.2.9)$$

where y_t is initiated at y_0.

Importantly, when the driver Wiener processes B_u and B_ε are independent, data generated from (3.2.2) is observationally equivalent to data from the continuous system

$$dy(t) = y(t)\widetilde{\mu}dt + \sqrt{\widetilde{\sigma}^2 y^2(t) + \sigma^2}dB_v(t), \;\; y(0) = y_0, \qquad (3.2.10)$$

where $B_v(t)$ is another standard Brownian motion. In the same way, model (3.2.7), is observationally equivalent to the (approximate) discrete system[3]

$$y_{t\Delta} = \exp\left\{\widetilde{\mu}\Delta\right\} y_{(t-1)\Delta} + \sqrt{(\widetilde{\sigma}^2 y_{(t-1)\Delta}^2 + \sigma^2)\Delta} \cdot v_{t\Delta}, \qquad (3.2.11)$$

where $v_{t\Delta} \sim \mathcal{N}(0,1)$ and $y_{t\Delta}$ exhibits conditional heteroskedasticity. Notably, by standardization of $1/\sqrt{\Delta}$, the conditional variance of the process is $\widetilde{\sigma}^2 y_{(t-1)\Delta}^2 + \sigma^2$, so that large realizations of the process magnify its variability. This dependence has a substantial bearing on the properties of $y_{t\Delta}$ and the form of its trajectories. Moreover, $y_{t\Delta}$ has a submartingale property when $e^{\widetilde{\mu}\Delta} > 1$ and given $y_{(t-1)\Delta} > 0$ because in that case $E_{(t-1)\Delta}(y_{t\Delta}) = e^{\widetilde{\mu}\Delta}y_{(t-1)\Delta} > y_{(t-1)\Delta}$.

[3]In Appendix A.1, we derive the discretization error by directly calculating the difference between equation (3.2.9) and (3.2.11), and we show that this discretization error will not affect the limiting distribution of our estimators.

Assuming $\widetilde{\sigma}^2 > 0$, models (3.2.9) and (3.2.2) have the following properties: (1) $E(\rho_{t\Delta}) = e^{\widetilde{\mu}\Delta}$, which is unity if and only if $\widetilde{\mu} = 0$ and exceeds unity if and only if $\widetilde{\mu} > 0$; (2) $E(\rho_{t\Delta}^2) = \exp\left(2\widetilde{\mu}\Delta + \widetilde{\sigma}^2\Delta\right) = \exp(2\kappa\Delta)$, which exceeds unity if and only if $\kappa > 0$; (3) $Var(\rho_{t\Delta}) = e^{2\widetilde{\mu}\Delta}\left(e^{\widetilde{\sigma}^2\Delta} - 1\right) > 0$; (4) $E(\rho_{t\Delta}^k) = \exp\left(k\Delta\left[\widetilde{\mu} + \frac{1}{2}(k-1)\widetilde{\sigma}^2\right]\right) \to \infty$ when $k \to \infty$; (5) As shown in Föllmer and Schweizer (1993), when $\phi = \widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2 < 0$, the process is asymptotically stationary but may not have finite second moments.[4] To ensure the existence of second moments, we should impose a stronger condition that $\kappa = \widetilde{\mu} + \frac{1}{2}\widetilde{\sigma}^2 < 0$. From (3.2.5), when $\kappa < 0$, it is apparent that the variance of $K(t)$ exists and converges to $-0.5\sigma^2/\kappa < \infty$ as $t \to \infty$. It then follows that (3.2.2) is asymptotically covariance stationary; (6) If $\kappa = 0$, the variance of $K(t)$ equals to $\sigma^2 t$ that diverges as $t \to \infty$, which means (3.2.2) is not asymptotically covariance stationary. Since $\kappa = 0$ implies $\widetilde{\mu} < 0$ and $\phi < 0$, (3.2.2) is asymptotic stationarity; (7) If $\phi = \widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2 \geq 0$**,** $y(t)$ is no longer asymptotically stationary as shown in Föllmer and Schweizer (1993).

Table 3.1: Properties of Proposed Model Under Different Scenarios

Scenario	Asymptotically Stationary	Asym. Covariance Stationary	$E(\rho_{t\Delta})$	$E(\rho_{t\Delta}^2)$
$\widetilde{\mu} + \widetilde{\sigma}^2/2 < 0$	Yes	Yes	< 1	< 1
$\widetilde{\mu} + \widetilde{\sigma}^2/2 = 0$	Yes	No	< 1	$= 1$
$\widetilde{\mu} + \widetilde{\sigma}^2/2 > 0$ & $\widetilde{\mu} < 0$	Yes	No	< 1	> 1
$\widetilde{\mu} = 0$	Yes	No	$= 1$	> 1
$\widetilde{\mu} > 0$ & $\widetilde{\mu} - \widetilde{\sigma}^2/2 < 0$	Yes	No	> 1	> 1
$\widetilde{\mu} - \widetilde{\sigma}^2/2 \geq 0$	No	No	> 1	> 1

Table 3.1 summarizes the stationarity properties mentioned above and the respective values of $E(\rho_{t\Delta})$, and $E(\rho_{t\Delta}^2)$ under different regions of the parameter space depending on the values of $\widetilde{\mu}$ and $\widetilde{\sigma}^2$. When $\widetilde{\mu} + \widetilde{\sigma}^2/2 < 0$, the model is asymptotically covariance stationary with both $E(\rho_{t\Delta}) < 1$ and $E(\rho_{t\Delta}^2) < 1$. Figure 1(a) plots a simulated time series in this case with $\widetilde{\mu} = -5, \widetilde{\sigma}^2 = 0.5$ and $\widetilde{\mu} + \widetilde{\sigma}^2/2 = -4.75$ where stationary behavioral features of the data are apparent.

[4]When y_0 is fixed, the process is not stationary for finite t but is asymptotically stationary.

When $\widetilde{\mu} + \widetilde{\sigma}^2/2 = 0$, the model retains asymptotic stationarity but is no longer covariance stationary with $E(\rho_{t\Delta}) < 1$ and $E(\rho_{t\Delta}^2) = 1$. Figure 1(b) plots a simulated time series in this case with $\widetilde{\mu} = -2, \widetilde{\sigma}^2 = 4$ and $\widetilde{\mu} + \widetilde{\sigma}^2/2 = 0$ where stationarity is again apparent but with more evidence of persistence in the trajectory than in Figure 1(a). It was suggested in Granger and Swanson (1997) that the unit root hypothesis in a STUR random environment might be represented by the expectation $E(\rho_{t\Delta}^2) = 1$. However, the stationary properties of the time series in this case suggest stable and mean recursive trajectories that have greater persistence than when $E(\rho_{t\Delta}^2) < 1$.

When $\widetilde{\mu} + \widetilde{\sigma}^2/2 > 0$ and $\widetilde{\mu} < 0$, the model is asymptotically stationary but is not covariance stationary and $E(\rho_{t\Delta}^2) > 1$. Figure 1(c) plots a simulated time series in this case with $\widetilde{\mu} = -1, \widetilde{\sigma}^2 = 3.5$ and $\widetilde{\mu} + \widetilde{\sigma}^2/2 = 0.75$. Whereas the expectation $E(\rho_{t\Delta})$ is still less than unity, unstable behavior is evident in the simulated time series. In particular, the unstable subperiod of growth and collapse in the trajectory mimics bubble phenomena that are observed in actual data, such as that in Figure 3.5 in the empirical section of the present paper and in Figure 1 of PWY (2011). If $\widetilde{\mu} = 0$, the model continues to be asymptotically stationary but is not covariance stationary and $E(\rho_{t\Delta}) = 1$, so the model reduces to the stochastic unit root (STUR) model of Granger and Swanson (1997). Figure 1(d) plots a simulated time series in this case with $\widetilde{\mu} = 0, \widetilde{\sigma}^2 = 2$ and $\widetilde{\mu} + \widetilde{\sigma}^2/2 = 1$. Compared to the traditional (nonstochastic) unit root model, unstable behavior with bubble-like phenomenon in a subperiod of the simulated trajectory is now more evident.

When $\widetilde{\mu} > 0$, $E(\rho_{t\Delta}) > 1$ and $\Pr(\rho_{t\Delta} > 1) > 0.5$, giving greater probability to the realization of an explosive root than a unit or stationary root. However, unlike the traditional (nonstochastic) explosive AR(1) model which is nonstationary, this model is still asymptotically stationary although not covariance stationary. Figure 1(e) plots a simulated time series in this case with $\widetilde{\mu} = 0.5, \widetilde{\sigma}^2 = 2, \widetilde{\mu} + \widetilde{\sigma}^2/2 = 1.5$, $\widetilde{\mu} - \widetilde{\sigma}^2/2 = -0.5$. Although the trajectory in Figure 1(e) appears similar to those of Figure 1(c) and Figure 1(d), the process exhibits larger variation, as is apparent from

the vertical scale of the figure. When $\widetilde{\mu} - \widetilde{\sigma}^2/2 > 0$, the model is asymptotically nonstationary and both moments $E(\rho_{t\Delta})$ and $E(\rho_{t\Delta}^2)$ exceed unity. Figure 1(f) plots a simulated time series in this case with $\widetilde{\mu} = 1, \widetilde{\sigma}^2 = 0.5$ and $\widetilde{\mu} - \widetilde{\sigma}^2/2 = 1.25$. The explosive growth behavior is clearly evident in the plotted trajectory.

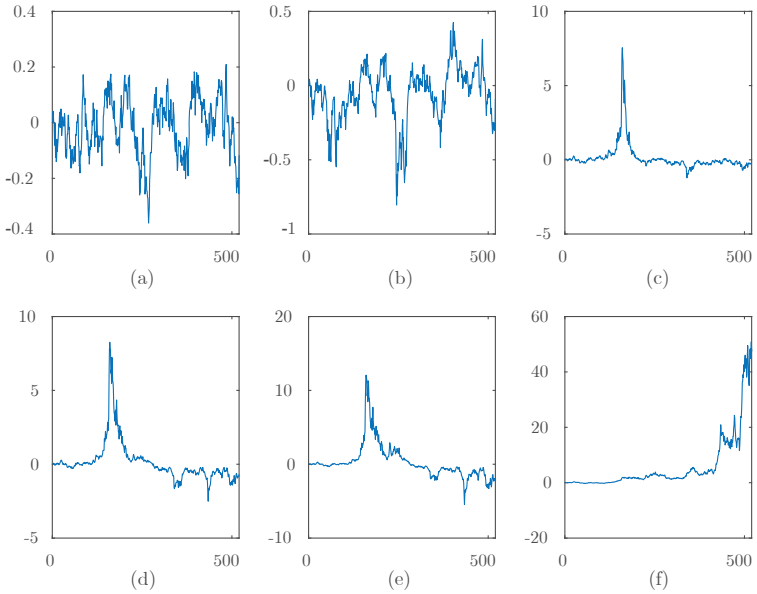


Figure 3.1: Simulated paths from the proposed model (3.2.2) when $\widetilde{\mu}$ and $\widetilde{\sigma}^2$ are in different regions.

The exact discrete time representation of our model is closely related to the near explosive random coefficient (NERC) model proposed recently in Banerjee et al. (2017) and to the multivariate local STUR model that is studied in Lieberman and Phillips (2017a) which combines deterministic local unit root (LUR) and random STUR component departures from unity. In particular, if Δ is chosen as $1/T^{\alpha}$ and $y_0 = 0$, then model (3.2.9) is the same as model (1) in Banerjee et al. (2017); and if Δ is chosen as $1/T$ and $y_0 = 0$, then our (3.2.9) has the same form as equation (4) in Lieberman and Phillips (2017a). As discussed in Phillips and Magdalinos (2007), the power rate α in the fraction $1/T^{\alpha}$ controls the degree of mild deviation from a unit root and is typically assumed to lie strictly between zero and unity, which assures that such deviations are localized to unity and exceed the usual local to unity departure of $O\left(T^{-1}\right)$.

In the standard discrete time modeling framework, the localizing rate parameter

α is difficult to estimate, although it is possible to do so at a slow rate (Phillips, 2012). Following the argument used in Wang and Yu (2016) with double asymptotics (i.e., both large span and infill schemes), the discrete time model (3.2.9), or equivalently (3.2.11), implies mild deviations from a unit root in which the localizing rate is determined by the sampling frequency Δ, and so there is no need to estimate a separate parameter α. This distinction implies an important advantage of the underlying continuous system framework when it is appropriate in practical work to employ this model using discrete time observations. A further useful difference is that the continuous system allows for flexible initial condition assumptions.

The model reduces to a simple autoregression with a time-invariant coefficient when $\widetilde{\sigma}^2 = 0$, in which case $\kappa = \phi = \widetilde{\mu}$ and then explosive behavior applies when $\phi > 0$. Conventional tests for a unit versus an explosive root therefore reduce to testing $\phi = 0$ against $\phi > 0$. This formulation explains the focus on right-tailed unit root testing (Diba and Grossman, 1988), including the recursive methodology used in PWY (2011), Phillips and Yu (2011), PSY (2015a, b) and related work.

In the extended model (3.2.9), a wider set of dynamic patterns are possible for studying various types of extreme behavior in realized sample trajectories. More specifically, we consider three cases distinguished by the following typology.

1. Unstable trajectory: $\kappa = \widetilde{\mu} + \widetilde{\sigma}^2/2 > 0$ which is equivalent to $E(\rho_{t\Delta}^2) > 1$. In this case, the model is covariance nonstationary asymptotically and is capable of generating trajectories with explosive and collapse behavior;

2. Locally Explosive trajectory: $\widetilde{\mu} > 0$ which is equivalent to $E(\rho_{t\Delta}) > 1$. In this case, there is greater probability for an explosive root to be realized in the sample than a unit or stationary root and the model is covariance nonstationary asymptotically. The model is capable of generating both explosive and collapsing behavior;

3. Explosive trajectory: $\phi = \widetilde{\mu} - \widetilde{\sigma}^2/2 > 0$. Here the model is nonstationary asymptotically and generates explosive behavior.

According to this terminology *explosiveness* implies *local explosiveness* which implies *instability*. We characterize all of these cases as various forms of extreme behavior. Figure 3.2 shows regions of the parameter space $(\widetilde{\mu}, \widetilde{\sigma}^2)$ that accord with these classifications of sample behavior.

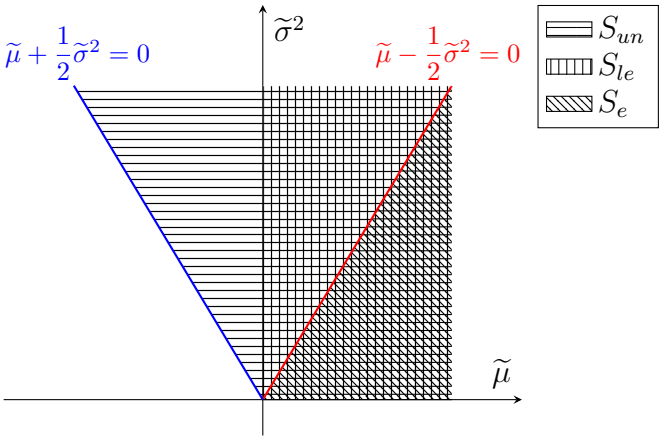


Figure 3.2: Various explosive regions of $\{y_t\}$ characterized by different parameter combinations of $(\widetilde{\mu}, \widetilde{\sigma}^2) \in \mathbb{R} \times \mathbb{R}_+$. S_{un} is the region for instability. S_{le} is the region for local explosiveness. S_e is the region for explosiveness.

It is helpful to link the above concepts of instability, local explosiveness and explosiveness to some well-known concepts in the stochastic process literature and to those used recently in Kim and Park (2016). Note first that the observational equivalent model (3.2.11) is a special case of generalized Höpfner and Kutoyants (GHK) diffusion (Höpfner and Kutoyants, 2003):[5]

$$dX_t = \frac{\widetilde{\mu} X_t}{\left(\widetilde{\sigma}^2 X_t^2 + \sigma^2\right)^{1-d}} dt + \left(\widetilde{\sigma}^2 X_t^2 + \sigma^2\right)^{d/2} dW_t$$

with $d = 1$. In this case, we can easily calculate the scale density $(s'(x))$ and the speed density $(m(x))$ of the model (3.2.11) as follows:

$$s'(x) = (\sigma^2 + \widetilde{\sigma}^2 x^2)^{-\widetilde{\mu}/\widetilde{\sigma}^2} \quad \text{and} \quad m(x) = (\sigma^2 + \widetilde{\sigma}^2 x^2)^{(\widetilde{\mu}/\widetilde{\sigma}^2 - 1)}. \tag{3.2.12}$$

Thus, the model (3.2.11) is recurrent if $\widetilde{\mu}/\widetilde{\sigma}^2 \leq 1/2$, i.e., $\phi \leq 0$. It is pos-

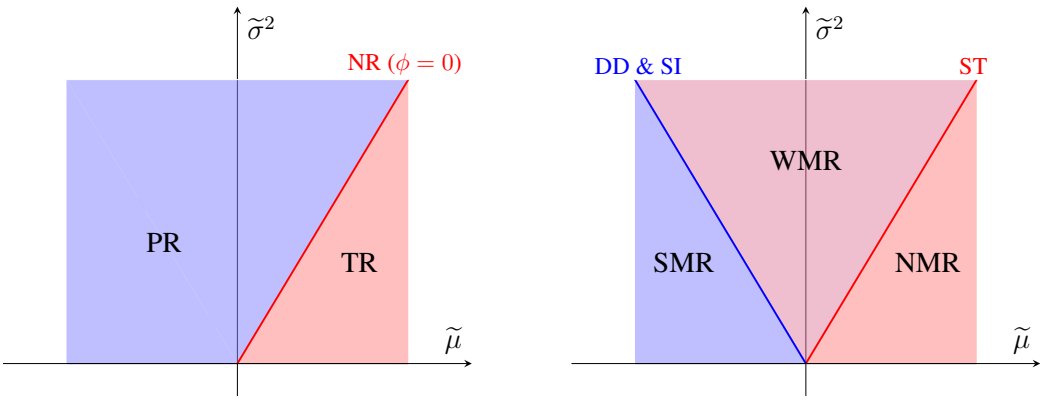
[5]The diffusion process studied here is a generalization of Example 2.1 in Kim and Park (2016) by adding a coefficient in front of X_t^2.

itive recurrent (PR) if $\widetilde{\mu}/\widetilde{\sigma}^2 < 1/2$, i.e., $\phi < 0$. Thus, it is null recurrent (NR) when $\phi = 0$ and transient (TR) when $\phi > 0$. Therefore, our definition of explosiveness corresponds to the transient property, which typically applies to processes that trend upwards or downwards and may be rendered recurrent after suitable detrending techniques as discussed by Kim and Park (2016) who considered various notions of mean reversion for financial time series. These authors related the mean-reversion property to the following three conditions:

(ST): the speed measure m is either integrable or barely nonintegrable[6];

(DD): The inverse of the scale density $1/s'$ is either integrable or barely nonintegrable;

(SI): square of identity function, ι^2, is either m-integrable[7] or m-barely nonintegrable.



(a) Recurrence Properties (b) Mean Reversion Properties

Figure 3.3: Subfigures (a) and (b) characterize the recurrence properties and the mean reversion properties of $\{y_t\}$ under different combinations of $(\widetilde{\mu}, \widetilde{\sigma}^2) \in R \times R_+$. PR=positive recurrent, NR=null recurrent, TR=transient; SMR=strong mean reversion, WMR=weak mean reversion, NMR=no mean reversion.

Kim and Park (2016) showed that when both ST and DD hold, the process has strong mean reversion (SMR) and if only one of ST and DD holds the process has

[6]A function m is defined to be *barely nonintegrable* if there exists some slowly varying function ℓ such that $m\ell$ is integrable.

[7]The square of the identity function ι^2 is defined by $\iota^2(x) = x^2$; and a function f is defined to be m-*integrable* if mf is integrable.

weak mean reversion (WMR). By checking these conditions, we find that model (3.2.11) satisfies: DD if and only if $\widetilde{\mu}/\widetilde{\sigma}^2 \leq -1/2$, i.e., $\kappa \leq 0$; ST if and only if $\widetilde{\mu}/\widetilde{\sigma}^2 \leq 1/2$, i.e., $\phi \leq 0$; and SI if $\widetilde{\mu}/\widetilde{\sigma}^2 \leq -1/2$, i.e., $\kappa \leq 0$. So in our model the condition that ensures ST is the same as that which ensures SI, and is stronger than that which ensures DD. Thus, if $\kappa \leq 0$, our model has strong mean reversion; if $\phi \leq 0$ but $\kappa > 0$, our model has weak mean reversion; and if $\phi > 0$, our model does not imply mean reversion. Hence, our definition of explosiveness is the same as no mean reversion in Kim and Park (2016). Figure 3.3 summarizes the mean reversion properties of the process, viz., strong mean reversion (SMR), weak mean reversion (WMR), and no mean reversion (NMR) of the diffusion process (3.2.11) in different regions of the respective parameter spaces.

3.3 Model Estimation using Realized Volatility

To estimate the continuous-time model (3.2.2) based on discretely sampled data, we employ the two-stage estimation procedure proposed by Phillips and Yu (2009). In the first stage we make use of the feasible central limit theory for realized volatility to set up a regression model for estimating $\widetilde{\sigma}^2$ and σ^2. In the second stage the in-fill log-likelihood function is maximized to estimate $\widetilde{\mu}$. Consistency and asymptotic distribution theory are established for all estimates.

To explain the estimation method and to establish the large sample theory of the estimators, we assume the time interval $[0, T]$ with span length T can be split into N equispaced blocks. The time span of each block is $h := T/N$ and we assume there are M observations of y_t within each block. So in total $M \times N$ observations on y_t are available over $[0, T]$ and $M \times N = T/\Delta$. Further assume that as $\Delta \to 0$, $M \to \infty$ and $M \times N \to \infty$. Figure 3.4 illustrates this notation and the sampling scheme. For example, if ten years of daily observations are available and we split the data into ten blocks, then $T = 10$, $\Delta = 1/252$, $M = 252$, $h = 1$, and $N = 10$. The total number of observations is 2520 and the number of observations contained

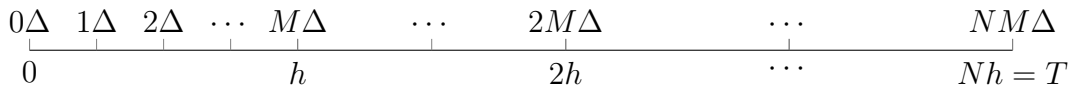


Figure 3.4: Notational schematic for individual observations, block divisions, and full sample span

in each block is 252.

The quadratic variation process $[y]_t$ of $y(t)$ in (3.2.2) satisfies $d[y]_t = (\widetilde{\sigma}^2 y_t^2 + \sigma^2)\, dt$, giving

$$[y]_t = \int_0^t \left(\widetilde{\sigma}^2 y_s^2 + \sigma^2\right) ds. \tag{3.3.1}$$

Barndorff-Nielsen and Shephard (2002) showed that quadratic variation may be consistently estimated using realized variance (RV) when $\Delta \to 0$. Realized variance and realized quarticity (RQ) are computed using increments $y_{(n-1)h+i\Delta} - y_{(n-1)h+(i-1)\Delta}$ in the observed process by means of the following formulae calculated over the nth block

$$RV_n = \sum_{i=1}^{M} \left[y_{(n-1)h+i\Delta} - y_{(n-1)h+(i-1)\Delta}\right]^2, \quad n = 1, 2, \cdots, N,$$
$$RQ_n = \frac{1}{3\Delta} \sum_{i=1}^{M} \left[y_{(n-1)h+i\Delta} - y_{(n-1)h+(i-1)\Delta}\right]^4, \quad n = 1, 2, \cdots, N.$$

From Barndorff-Nielsen (2002) realized variance has the following asymptotic distribution for large M within each block

$$\sqrt{M}\left([y]_{(n-1)h}^{nh} - RV_n\right) \xrightarrow{L} \mathcal{MN}\left(0, 2h \int_{(n-1)h}^{nh} \left(\widetilde{\sigma}^2 y_s^2 + \sigma^2\right)^2 ds\right), \tag{3.3.2}$$

where $\mathcal{MN}$ signifies mixed normal and $[y]_{(n-1)h}^{nh} = \int_{(n-1)h}^{nh} (\widetilde{\sigma}^2 y_s^2 + \sigma^2)\, ds$.

Following the algorithm of Phillips and Yu (2009), the first-stage estimation step aims to estimate $\boldsymbol{\theta} := (\widetilde{\sigma}^2, \sigma^2)'$ by least squares using the criterion

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q_\Delta(\boldsymbol{\theta}), \tag{3.3.3}$$

where

$$Q_\Delta(\boldsymbol{\theta}) = \Delta \sum_{n=1}^{N} \frac{\left(\log RV_n - \log[y]^{nh}_{(n-1)h} + \frac{1}{2}s_n^2\right)^2}{s_n^2},$$

with

$$s_n = \max\left\{\sqrt{2\Delta\frac{RQ_n}{RV_n^2}}, \sqrt{\frac{2}{M}}\right\},$$

and where Θ is a compact subset of R_+^2 containing the true value $\theta_0 = (\widetilde{\sigma}_0^2, \sigma_0^2)'$ as an interior point. The term $s_n^2/2$ in the numerator of $Q_\Delta(\boldsymbol{\theta})$ is a finite sample correction on the asymptotic theory. In practice, the quadratic variation element $[y]^{nh}_{(n-1)h}$ in $Q_\Delta(\boldsymbol{\theta})$ can be approximated by Riemann sums as follows

$$[y]^{nh}_{(n-1)h} = \int_{(n-1)h}^{nh} (\widetilde{\sigma}^2 y_s^2 + \sigma^2) ds = \Delta \sum_{t=1}^{M} \left\{\widetilde{\sigma}^2 y^2_{(n-1)h+t\Delta} + \sigma^2\right\} + O(\sqrt{\Delta}).$$

In the second stage, $\widetilde{\mu}$ is estimated by maximizing the approximate log-likelihood function, viz.,

$$\widehat{\widetilde{\mu}} = \arg\max_{\widetilde{\mu}} \frac{1}{MN} \log \ell_{ALF}(\widetilde{\mu}), \tag{3.3.4}$$

$$\ell_{ALF}(\widetilde{\mu}) = \sum_{t=1}^{M\times N} \frac{\widetilde{\mu} y_{(t-1)\Delta}}{\widehat{\widetilde{\sigma}^2} y^2_{(t-1)\Delta} + \widehat{\sigma^2}} \left(y_{t\Delta} - y_{(t-1)\Delta}\right) - \frac{\Delta}{2} \sum_{t=1}^{M\times N} \frac{\widetilde{\mu}^2 y^2_{(t-1)\Delta}}{\widehat{\widetilde{\sigma}^2} y^2_{(t-1)\Delta} + \widehat{\sigma^2}}., \tag{3.3.5}$$

giving

$$\widehat{\widetilde{\mu}} = \Delta^{-1} \frac{\sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}\left(y_{t\Delta} - y_{(t-1)\Delta}\right)}{\widehat{\widetilde{\sigma}^2} y^2_{(t-1)\Delta} + \widehat{\sigma^2}}}{\sum_{t=1}^{M\times N} \frac{y^2_{(t-1)\Delta}}{\widehat{\widetilde{\sigma}^2} y^2_{(t-1)\Delta} + \widehat{\sigma^2}}} =: \Delta^{-1} \frac{\widehat{A}_N}{\widehat{B}_N}. \tag{3.3.6}$$

This estimator of $\widetilde{\mu}$ has the same form as the weighted least squares estimator used by Hwang and Basawa (2005) in the context of a discrete time RCAR.

3.4 Asymptotic Theory

This section develops asymptotic theory for the diffusion parameter and drift parameter estimates. Limit theory for estimates of the diffusion parameters is first obtained by using infill asymptotic methods where $\Delta \to 0$. We then investigate the asymptotic theory of the drift parameter by resorting to double asymptotic methods in which both $\Delta \to 0$ and $T \to \infty$.

3.4.1 Diffusion Parameters

This section derives asymptotic theory for the estimates $\widehat{\widetilde{\sigma}}^2$ and $\widehat{\sigma}^2$ by assuming $\Delta \to 0$ (with a fixed T) in an infill asymptotic scheme. Since $M = h/\Delta$ it follows that $\Delta \to 0$ implies $M \to \infty$. Realized variance and realized quarticity are therefore consistent estimates of integrated variance and integrated quarticity. Since T is fixed, $N = T/h$ is also fixed. Let $\{y_t\}_{t=\Delta}^{MN\Delta}$ be a discrete sample generated from (3.2.2) where the true parameter values for $\widetilde{\mu}, \widetilde{\sigma}^2, \sigma^2$ are denoted $\widetilde{\mu}_0, \widetilde{\sigma}_0^2, \sigma_0^2$. Assume that $\boldsymbol{\theta_0} = (\widetilde{\sigma}_0^2, \sigma_0^2)' \in \text{Int}(\boldsymbol{\Theta})$ where $\boldsymbol{\Theta}$ is a compact set in R_+^2. Let $\rho_0 = \exp(\widetilde{\mu}_0 \Delta) = E(\rho_{t\Delta})$, and $\widehat{\rho} = \exp\left(\widehat{\widetilde{\mu}}\Delta\right)$. The following result provides within block infill asymptotics as $\Delta \to 0$.

Theorem 3.4.1 *If $\boldsymbol{\theta_0} \in$ Int$(\boldsymbol{\Theta})$ and $\Delta \to 0$,*

$$\frac{1}{\sqrt{\Delta}}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}\right) \xrightarrow{L}$$
$$\left(\sum_{n=1}^{N} \frac{\int_{(n-1)h}^{nh} \frac{\partial \breve{\sigma}^2(y_s;\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} \cdot \frac{\partial \breve{\sigma}^2(y_s;\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}'} ds}{\int_{(n-1)h}^{nh} \breve{\sigma}^4(y_s;\boldsymbol{\theta_0}) ds}\right)^{-1} \left(\sum_{n=1}^{N} \frac{\sqrt{2}\int_{(n-1)h}^{nh} \frac{\partial \breve{\sigma}^2(y_s;\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} \breve{\sigma}^2(y_s;\boldsymbol{\theta_0}) dB_s}{\int_{(n-1)h}^{nh} \breve{\sigma}^4(y_s;\boldsymbol{\theta_0}) ds}\right),$$

where $\breve{\sigma}^2(y_t;\boldsymbol{\theta_0}) = \widetilde{\sigma}_0^2 y_t^2 + \sigma_0^2$ is the spot variance of $y(t)$.

Remark 3.4.1 *In discrete time modeling, it is common for the parameters $\widetilde{\sigma}^2$ and σ^2 to be estimated by MLE or QMLE by imposing ARCH-type innovations, see for example Jensen and Rahbek (2004); Ling and Li (2008); Francq and Zakoïan*

(2012); Chen et al. (2014). This approach provides consistent estimates and associated asymptotics for $\widetilde{\sigma}^2$ rather than σ^2 when y_t is nonstationary. The explanation is that as $T \to \infty$, the log-likelihood function becomes flat because of the dominating scale effects of y_T that occur in the direction where $\widetilde{\sigma}^2$ is fixed and σ^2 varies. Unlike previous work, our approach applies an infill asymptotic scheme which fixes the time span (T) and shrinks the sampling interval (Δ) to 0. These asymptotics ensure that y_T is measurable and finite, so that σ^2 continues to play a role in the limit as $\Delta \to 0$. With this approach it is possible to consistently estimate both variance parameters and establish their asymptotic properties as in Theorem 3.4.1.

Remark 3.4.2 *As found in Phillips and Yu (2009), there is a trade-off between large N and small N. When N is too small, the variance of the estimator is large. However, when N is sufficiently large, further increases in N fail to improve the variance. The choice of the optimal N is therefore of interest but is beyond the scope of the present paper.*

It is interesting in practical applications to test the null hypothesis $\widetilde{\sigma}^2 = 0$, which corresponds to the special case of no randomness in the persistence properties of $y(t)$. To test this boundary condition hypothesis we apply a modified version of the locally best invariant test (*LBI*-test) by Lee (1998) for $\widetilde{\sigma}^2 = 0$, viz.,

$$\widetilde{Z}_N := \frac{\sum_{t=1}^{M\times N}\left(\widetilde{\varepsilon}_{t\Delta}^2 - \left(\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{\varepsilon}_{t\Delta}^2\right)\right)\widetilde{y}_{(t-1)\Delta}^2}{\sqrt{\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{\varepsilon}_{t\Delta}^4 - \left(\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{\varepsilon}_{t\Delta}^2\right)^2}\sqrt{\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{y}_{(t-1)\Delta}^4 - \left(\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{y}_{(t-1)\Delta}^2\right)^2}}$$

where $\widetilde{y}_{t\Delta} = \dfrac{y_{t\Delta}}{\sqrt{1+y_{t\Delta}^2}}$, $\widetilde{\varepsilon}_{t\Delta} = y_{t\Delta} - \widetilde{\rho}y_{(t-1)\Delta}$ and $\widetilde{\rho} = \left(\sum_{t=1}^{M\times N}\widetilde{y}_{(t-1)\Delta}y_{(t-1)\Delta}\right)^{-1}\sum_{i=1}^{M\times N}\widetilde{y}_{(t-1)\Delta}y_{t\Delta}$. Then, as $N \to \infty$,

$$(MN)^{-1/2}\widetilde{Z}_N \xrightarrow{L} \mathcal{N}(0,1), \text{ under } H_0 : \widetilde{\sigma}^2 = 0,$$

and

$$|(MN)^{-1/2}\widetilde{Z}_N| \xrightarrow{p} \infty, \quad \text{under } H_1 : \widetilde{\sigma}^2 > 0.$$

Remark 3.4.3 *Note first that we use the self-normalized variable $\widetilde{y}_{t\Delta}$ for constructing the test statistic. This is because the normalization ensures that $\widetilde{y}_{t\Delta}$ is stationary when $y_{t\Delta}$ is nonstationary, which is crucial for $\widetilde{Z}_N$ to converge under the null hypothesis (Lee, 1998; Nagakura, 2009). In fact, the weighting function $1 + y_{t\Delta}^2$ can be replaced by any function $g(x)$ where $g : [0, \infty) \to (0, \infty)$ is a Borel function satisfying $x^2/g(x) \to 1$ as $|x| \to \infty$. In practice, we follow the usual convention by setting the weighting function to be $1 + y_{t\Delta}^2$ as in Hill and Peng (2014) and Horváth and Trapani (2016).*

Remark 3.4.4 *A second point worthy of note concerns the use of the IV estimate $\widetilde{\rho}$ here. Following Chan et al. (2012), the IV estimate*

$$\widetilde{\rho} = \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta} y_{t\Delta}}{\sqrt{\delta + y_{(t-1)\Delta}^2}} \Bigg/ \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}^2}{\sqrt{\delta + y_{(t-1)\Delta}^2}}$$

is uniformly asymptotically normally distributed for both stationary and nonstationary $y_{t\Delta}$. Further, the IV estimate $\widetilde{\rho}$ includes the Cauchy estimator (So and Shin, 1999) as a special case ($\delta = 0$), which is known to be asymptotically median-unbiased. This helps improve the finite sample performance of the test statistic which depends explicitly on the residuals.

Remark 3.4.5 *The above test for coefficient constancy remains valid in the presence of correlation between the random coefficient and innovations. When the random coefficients are endogenous the quadratic covariation $\langle B_u, B_\varepsilon \rangle_t = \int_0^t \gamma_s ds$ and the conditional variance of $\widetilde{\varepsilon}_{t\Delta}$ under the null is $Var(\widetilde{\varepsilon}_{t\Delta} | y_{(t-1)\Delta}) = \sigma^2$, whereas under the alternative $Var(\widetilde{\varepsilon}_{t\Delta} | y_{(t-1)\Delta}) = \widetilde{\sigma}^2 y_{(t-1)\Delta}^2 + 2\gamma_t \widetilde{\sigma} \sigma y_{(t-1)\Delta} + \sigma^2$. The test may therefore be interpreted as examining evidence for the presence of a relationship between $\widetilde{\varepsilon}_{t\Delta}^2$ and $y_{(t-1)\Delta}^2$ and $y_{(t-1)\Delta}$ – in other words, a test for conditional heteroskedasticity.*

3.4.2 Drift Parameter

This section first derives limit theory for the estimated drift parameter under the double asymptotics scheme with $\Delta \to 0$ and $T \to \infty$. In this framework both $M \to \infty$ and $N \to \infty$. Then we derive limit theory for the expectation of the random autoregressive coefficient $\rho_{t\Delta}$ by requiring only that $\Delta \to 0$. Based on the limit theory for these estimated parameters we construct test statistics for assessing the nature of extreme sample path behavior in data.

Theorem 3.4.2 *In model (3.2.2), assume $\widetilde{\sigma}_0^2 > 0$. When $T \to \infty$ and $\Delta \to 0$, $\widetilde{\mu} \xrightarrow{p} \widetilde{\mu}_0$. Additionally, if $T\Delta \to 0$, the asymptotic distribution of $\widehat{\widetilde{\mu}}$ is given by*

$$\sqrt{T}\left(\widehat{\widetilde{\mu}} - \widetilde{\mu}_0\right) \xrightarrow{L} \mathcal{N}\left(0, V_1^{-1}\right),$$

where

$$V_1 = \begin{cases} E\left(\dfrac{y_t^2}{\widetilde{\sigma}_0^2 y_t^2 + \sigma_0^2}\right), & \text{if } \kappa = \widetilde{\mu}_0 + \dfrac{1}{2}\widetilde{\sigma}_0^2 < 0; \\ \widetilde{\sigma}_0^{-2}, & \text{if } \kappa = \widetilde{\mu}_0 + \dfrac{1}{2}\widetilde{\sigma}_0^2 \geq 0. \end{cases} . \tag{3.4.1}$$

Remark 3.4.6 *The asymptotics (3.4.2) hold regardless of the value of $\widetilde{\mu}_0 + \widetilde{\sigma}_0^2/2$, which may be less than zero, equal zero, or greater than zero. By contrast, it is well-known that in the case of the pure AR(1) model, the asymptotic theory for the least squares estimator of the autoregressive coefficient depends critically on the true value of the coefficient. However, in the RCAR model asymptotic normality may hold in both the stationary and explosive cases under certain conditions, as discussed in Hwang and Basawa (2005). The above result reinforces this finding and extends applicability to the continuous-time random coefficient model examined here.*

The asymptotic theory given in (3.4.2) suggests that consistent estimation of $\widetilde{\mu}$ requires $T \to \infty$. In practical work, however, the time span is often short making large span asymptotics less relevant. The following theorem provides infill asymp-

totics for estimating $\rho = \exp\{\widetilde{\mu}\Delta\}$, which is useful for testing nonstationarity in a finite time span setting.

Theorem 3.4.3 *In model (3.2.2), assume $\widetilde{\sigma}_0^2 > 0$. When T is fixed and $\Delta \to 0$, $\widehat{\rho} \xrightarrow{p} \rho_0 = 1$ and the asymptotic distribution of $\widehat{\rho}$ is given by*

$$\frac{1}{\Delta}\left(\widehat{\rho} - 1\right) \xrightarrow{L} \mathcal{N}\left(0, (TV_1)^{-1}\right). \tag{3.4.2}$$

where

$$\widehat{\rho} = 1 + \frac{\widehat{A}_N}{\widehat{B}_N}, \tag{3.4.3}$$

and V_1 is given in (3.4.1).

Remark 3.4.7 *Although the above result does not deliver a consistent estimate of $\widetilde{\mu}$ with a finite T, the asymptotic theory in (3.4.2) shows that consistent estimation of ρ is possible when $\Delta \to 0$. This result motivates estimation of $\beta_\kappa := \exp\left\{\left(\widetilde{\mu} + \widetilde{\sigma}^2/2\right)\Delta\right\}$ and $\beta_\phi := \exp\left\{\left(\widetilde{\mu} - \widetilde{\sigma}^2/2\right)\Delta\right\}$ instead of the continuous time parameters κ and ϕ when the time span of the data is short.*

Proposition 3.4.1 *For model (3.2.2) with T fixed and $\widetilde{\sigma}_0^2 > 0$, as $\Delta \to 0$,*

$$\frac{1}{\Delta}\left(\widehat{\beta}_\kappa - \beta_\kappa^0\right) \xrightarrow{L} \mathcal{N}\left(0, (TV_1)^{-1}\right), \quad \frac{1}{\Delta}\left(\widehat{\beta}_\phi - \beta_\phi^0\right) \xrightarrow{L} \mathcal{N}\left(0, (TV_1)^{-1}\right).$$

where

$$\widehat{\beta}_\kappa = \exp\left\{\left(\widehat{\widetilde{\mu}} + \frac{1}{2}\widehat{\widetilde{\sigma}^2}\right)\Delta\right\} \text{ and } \widehat{\beta}_\phi = \exp\left\{\left(\widehat{\widetilde{\mu}} - \frac{1}{2}\widehat{\widetilde{\sigma}^2}\right)\Delta\right\}, \tag{3.4.4}$$

$$\beta_\kappa^0 = \exp\left\{\left(\widetilde{\mu}_0 + \widetilde{\sigma}_0^2/2\right)\Delta\right\} \text{ and } \beta_\phi^0 = \exp\left\{\left(\widetilde{\mu}_0 - \widetilde{\sigma}_0^2/2\right)\Delta\right\}.$$

Remark 3.4.8 *To test different forms of unstable/explosive behavior, we need to test whether $\kappa = \widetilde{\mu} + \widetilde{\sigma}^2/2 = 0$, or $\widetilde{\mu} = 0$, or $\phi = \widetilde{\mu} - \widetilde{\sigma}^2/2 = 0$. Testing these restrictions corresponds to testing the hypotheses $\beta_\kappa = 1$, or $\rho = 1$, or $\beta_\phi = 1$. In the spirit of Theorem 3.4.3 and Proposition 3.4.1, as $\Delta \to 0$, we can construct the*

following test statistics and derive their asymptotic distributions as detailed below:

$$t_{\kappa} = \left(\frac{1}{\Delta} \sum_{t=1}^{M \times N} \frac{y_{(t-1)\Delta}^2}{\widehat{\widetilde{\sigma}}^2 y_{(t-1)\Delta}^2 + \widehat{\sigma^2}} \right)^{1/2} \left(\widehat{\beta}_{\kappa} - \beta_{\kappa}^0 \right) \xrightarrow{L} \mathcal{N}(0,1), \tag{3.4.5}$$

$$t_{\widetilde{\mu}} = \left(\frac{1}{\Delta} \sum_{t=1}^{M \times N} \frac{y_{(t-1)\Delta}^2}{\widehat{\widetilde{\sigma}}^2 y_{(t-1)\Delta}^2 + \widehat{\sigma^2}} \right)^{1/2} (\widehat{\rho} - \rho_0) \xrightarrow{L} \mathcal{N}(0,1), \tag{3.4.6}$$

$$t_{\phi} = \left(\frac{1}{\Delta} \sum_{t=1}^{M \times N} \frac{y_{(t-1)\Delta}^2}{\widehat{\widetilde{\sigma}}^2 y_{(t-1)\Delta}^2 + \widehat{\sigma^2}} \right)^{1/2} \left(\widehat{\beta}_{\phi} - \beta_{\phi}^0 \right) \xrightarrow{L} \mathcal{N}(0,1). \tag{3.4.7}$$

These three t-test statistics can be calculated sequentially and compared with the right-tailed critical value of the asymptotic distributions, giving a real-time testing strategy of empirical evidence of instability/explosiveness in the data. Accordingly, the origination and termination dates of different types of extreme behavior may be estimated in the same fashion as Phillips et al. (2015a). More specifically, date estimates can be determined from first crossing times as follows

$$\hat{r}_{un}^{ie} = \inf_{s \in [\hat{r}_{un}^{(i-1)f}, 1]} \{ s : t_{\kappa}(s) > Z_{0.95} \} \quad \text{and} \quad \hat{r}_{un}^{if} = \inf_{s \in [\hat{r}_{un}^{ie}, 1]} \{ s : t_{\kappa}(s) < Z_{0.95} \},$$

$$\hat{r}_{le}^{ie} = \inf_{s \in [\hat{r}_{le}^{(i-1)f}, 1]} \{ s : t_{\widetilde{\mu}}(s) > Z_{0.95} \} \quad \text{and} \quad \hat{r}_{le}^{if} = \inf_{s \in [\hat{r}_{le}^{ie}, 1]} \{ s : t_{\widetilde{\mu}}(s) < Z_{0.95} \},$$

$$\hat{r}_{e}^{ie} = \inf_{s \in [\hat{r}_{e}^{(i-1)f}, 1]} \{ s : t_{\phi}(s) > Z_{0.95} \} \quad \text{and} \quad \hat{r}_{e}^{if} = \inf_{s \in [\hat{r}_{e}^{ie}, 1]} \{ s : t_{\phi}(s) < Z_{0.95} \},$$

where: $Z_{0.95} = 1.645$ is the 95% critical value of the standard normal distribution; $\hat{r}_{un}^{ie}/\hat{r}_{le}^{ie}/\hat{r}_{e}^{ie}$ represent estimates of the origination date of the ith explosive period; and $\hat{r}_{un}^{if}/\hat{r}_{le}^{if}/\hat{r}_{e}^{if}$ represent estimates of the termination date of the ith explosive period. To identify the first unstable/explosive period in the sample, a minimum window is needed to start the recursion. The time-stamping strategy used here is based on the standard normal distribution whereas the PWY and PSY algorithms rely on non-standard unit root and sup unit root distributions.

3.5 The Model with Endogeneity

This section extends the base model (3.2.2) by allowing for endogeneity, quantified by the correlation between the random coefficient and the equation innovation. In the discrete time literature Hwang and Basawa (1997, 1998) described this framework as a generalized random coefficient autoregressive model. With stationarity imposed they studied the local asymptotic normality of the maximum likelihood estimator and the weighted least squares estimator of the autoregressive coefficient. Zhao and Wang (2012) considered empirical likelihood estimation of the stationary model and proposed a likelihood ratio test for testing stationary/ergodicity. Lieberman and Phillips (2017c) studied the effects of endogeneity in a multivariate context and derived the asymptotic distribution for the non-linear least squares (NLLS) estimator for the autoregressive coefficient, showing that NLLS is inconsistent for the autoregressive coefficient under endogeneity. To address inconsistency of NLLS, Lieberman and Phillips (2017b) proposed a non-linear instrumental variable technique and a GMM approach, establishing consistency and deriving the asymptotic distribution for the IV estimator of the autoregressive coefficient.

To incorporate endogeneity in a continuous time random coefficient setting, we rewrite the model (3.2.2) as the following continuous time system

$$\begin{aligned} dy(t) &= y(t)d\widetilde{Z}(t) + dZ(t), \quad y(0) = y_0, \\ d\widetilde{Z}(t) &= \widetilde{\mu}dt + \widetilde{\sigma}dB_u(t), \\ dZ(t) &= \sigma dB_\varepsilon(t), \end{aligned} \tag{3.5.1}$$

where (B_u, B_ε) is a two-dimensional Brownian motion with covariance parameter γ so that the quadratic covariation process satisfies $d\langle B_u, B_\varepsilon\rangle_t = \gamma dt$. Then, $d\langle \widetilde{Z}, Z\rangle_t = \gamma\widetilde{\sigma}\sigma dt := \omega dt$, where $\omega = \gamma\widetilde{\sigma}\sigma$ is the covariance parameter of $(\widetilde{Z}, Z)$.

According to Föllmer et al. (1994), the strong solution to this continuous system is

$$y(t) = \exp\left(\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)t + \widetilde{\sigma}B_u(t)\right)y(0) + J(t), \tag{3.5.2}$$

where

$$\begin{aligned} J(t) &= \sigma \int_0^t \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(t-s) + \widetilde{\sigma}\left(B_u(t) - B_u(s)\right)\right\} dB_\varepsilon(s) \\ &\quad - \omega \int_0^t \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(t-s) + \widetilde{\sigma}\left(B_u(t) - B_u(s)\right)\right\} ds \\ &= K(t) - L(t). \end{aligned}$$

Compared to the model without endogeneity in (3.2.4), the dynamics of the process are now driven by $J(t)$ instead of $K(t)$. $J(t)$ has two components, one being $K(t)$ and the other depending on the covariance of the random coefficient and the innovation, ω. The model specified in the system (3.5.1) is the continuous time limit of the endogenous stochastic unit root (STUR) model of Lieberman and Phillips (2017c) and ω corresponds to the one-sided long-run covariance in the STUR model.

The following proposition shows that the given characterization of instability/explosiveness in the model without endogeneity remains valid for the model with endogeneity.

Proposition 3.5.1 *The sample path characteristics of the process (3.5.2) may be classified into the following three types,*

1. *unstable:* $\kappa = \widetilde{\mu} + \frac{1}{2}\widetilde{\sigma}^2 > 0$*;*
2. *locally explosive:* $\widetilde{\mu} > 0$*;*
3. *explosive:* $\phi = \widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2 > 0$*.*

The fact that sample path characteristics of (3.5.2) are unaffected by endogeneity may be explained intuitively by noting that the model (3.5.1) is observationally

equivalent to the following continuous system

$$dy_t = \widetilde{\mu} y_t dt + \sqrt{\widetilde{\sigma}^2 y_t^2 + 2\omega y_t + \sigma^2} dB_v(t), \tag{3.5.3}$$

where $B_v(t)$ is another standard Brownian motion in an expanded probability space. Note that when the variance of y_t goes to infinity as t increases, the dominant term in the diffusion function $\widetilde{\sigma}^2 y_t^2 + 2\omega y_t + \sigma^2$ is $\widetilde{\sigma}^2 y_t^2$, which explains why $\widetilde{\sigma}^2$ is the key parameter in determining long-run volatility.

Remark 3.5.1 *From the perspective of diffusion process asymptotics, the recurrence and mean reversion characterizations given in Figure 3.3 also remain valid. This robustness is evident by checking the limit of the scale index function:*

$$p = \lim_{y \to \infty} v(y) = \lim_{y \to \infty} \frac{-2\widetilde{\mu} y^2}{\sigma^2 + 2\omega y + \widetilde{\sigma}^2 y^2} = -\frac{2\widetilde{\mu}}{\widetilde{\sigma}^2},$$

which is apparently unaffected by endogeneity in the limit.

We can rewrite the discrete time model in AR(1) format as

$$y_{t\Delta} = \exp\left(\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)\Delta + \widetilde{\sigma}\sqrt{\Delta} u_t\right) y_{(t-1)\Delta} + J_\Delta(t) = \rho_{t\Delta} y_{(t-1)\Delta} + J_{t\Delta}, \tag{3.5.4}$$

where $u_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, and

$$\begin{aligned} J_{t\Delta} = \sigma \int_{(t-1)\Delta}^{t\Delta} \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(t\Delta - s) + \widetilde{\sigma}\left(B_{u,t\Delta} - B_{u,s}\right)\right\} dB_{\varepsilon,s} \\ - \omega \int_{(t-1)\Delta}^{t\Delta} \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(t\Delta - s) + \widetilde{\sigma}\left(B_{u,t\Delta} - B_{u,s}\right)\right\} ds. \end{aligned}$$

From earlier derivations we know that

$$E\left(J_{t\Delta}\right) = \frac{\omega}{\widetilde{\mu}}(1 - \exp(\widetilde{\mu}\Delta)) = -\omega\Delta + O(\Delta^2), \tag{3.5.5}$$

$$Var(J_{t\Delta}) = O(\Delta). \tag{3.5.6}$$

Therefore, when standardizing the model by the factor $1/\sqrt{\Delta}$, the expectation of the

correspondingly standardized error process $J_{t\Delta}/\sqrt{\Delta}$ in (3.5.4) has order $O(\sqrt{\Delta})$ as $\Delta \to 0$. This means that under infill asymptotics we can consistently estimate the expectation of the random coefficient, $\rho_0 = E\rho_{t\Delta} = \exp(\widetilde{\mu}\Delta)$. This result is naturally achieved in the continuous time setup with infill asymptotics and contrasts with inconsistency of least squares estimation in discrete time models with endogeneity (Lieberman and Phillips, 2017c).

As before, we continue to apply the two stage estimation procedure of Phillips and Yu (2009) to estimate the model under endogeneity. Note that the quadratic variation of y_t now satisfies

$$d[y]_t = (\widetilde{\sigma}^2 y_t^2 + 2\omega y_t + \sigma^2)dt. \tag{3.5.7}$$

In light of the argument of Remark 3.4.1 we cannot consistently estimate ω and σ^2 in explosive cases under long-span sampling because the signal of y_t^2 is so strong that it drowns information in the linear and constant terms (i.e., $2\omega y_t$ and σ^2). However, infill asymptotics for $\widehat{\boldsymbol{\theta}^*} := \left(\widehat{\widetilde{\sigma}^2}, \widehat{\gamma}, \widehat{\sigma^2}\right)'$ can be developed in the same way as before and the results are summarized in the following theorem.

Theorem 3.5.1 *Assume $\boldsymbol{\theta}_{\mathbf{0}}^* \in Int(\boldsymbol{\Theta}^*)$ where $\boldsymbol{\Theta}^*$ is a compact set in $R_+ \times [-1,1] \times R_+$. As T is fixed and $\Delta \to 0$, we have*

$$\frac{1}{\sqrt{\Delta}}\left(\widehat{\boldsymbol{\theta}^*} - \boldsymbol{\theta}_{\mathbf{0}}^*\right) \xrightarrow{L}$$
$$\left(\sum_{n=1}^{N} \frac{\int_{(n-1)h}^{nh} \frac{\partial \breve{\sigma}^2(y_s;\boldsymbol{\theta}_{\mathbf{0}}^*)}{\partial \boldsymbol{\theta}^*} \cdot \frac{\partial \breve{\sigma}^2(y_s;\boldsymbol{\theta}_{\mathbf{0}}^*)}{\partial \boldsymbol{\theta}^{*\prime}} ds}{\int_{(n-1)h}^{nh} \breve{\sigma}^4(y_s;\boldsymbol{\theta}_{\mathbf{0}}^*) ds}\right)^{-1} \left(\sum_{n=1}^{N} \frac{\sqrt{2}\int_{(n-1)h}^{nh} \frac{\partial \breve{\sigma}^2(y_s;\boldsymbol{\theta}_{\mathbf{0}}^*)}{\partial \boldsymbol{\theta}^*} \breve{\sigma}^2(y_s;\boldsymbol{\theta}_{\mathbf{0}}^*) dB_s}{\int_{(n-1)h}^{nh} \breve{\sigma}^4(y_s;\boldsymbol{\theta}_{\mathbf{0}}^*) ds}\right),$$

where $\breve{\sigma}^2(y_t;\boldsymbol{\theta}_{\mathbf{0}}^*) = \widetilde{\sigma}_0^2 y_t^2 + 2\omega_0 y_t + \sigma_0^2$ is the spot variance of $y(t)$.

Remark 3.5.2 *In principle at least, this limit theory enables us to construct a test for endogeneity based on the asymptotic distribution of $\widehat{\gamma}$. However, the limit theory above is hard to implement as this distribution is non-standard and non-pivotal and*

$\widehat{\gamma}$ is biased when the frequency is low. Instead, to test the most relevant hypothesis of interest $H_0 : \gamma_0 = 0$ we propose the likelihood ratio test based on the objective function $Q_\Delta(\boldsymbol{\theta}^*)$:

$$LR = \Delta^{-1}\left(Q_\Delta^r - Q_\Delta^{ur}\right) \sim \chi^2(1), \text{ under } H_0 : \gamma_0 = 0. \tag{3.5.8}$$

For consistent estimation of $\widetilde{\mu}$, as in the base model, we maximize the following approximated likelihood

$$\ell_{ALF}(\widetilde{\mu}) = \sum_{t=1}^{M\times N} \frac{\widetilde{\mu} y_{(t-1)\Delta}\left(y_{t\Delta} - y_{(t-1)\Delta}\right)}{\widehat{\widetilde{\sigma}^2} y_{(t-1)\Delta}^2 + 2\widehat{\omega} y_{(t-1)\Delta} + \widehat{\sigma^2}} - \frac{\Delta}{2}\sum_{t=1}^{M\times N} \frac{\widetilde{\mu}^2 y_{(t-1)\Delta}^2}{\widehat{\widetilde{\sigma}^2} y_{(t-1)\Delta}^2 + 2\widehat{\omega} y_{(t-1)\Delta} + \widehat{\sigma^2}}, \tag{3.5.9}$$

where $\widehat{\omega} = \widehat{\gamma}\sqrt{\widehat{\widetilde{\sigma}^2}\widehat{\sigma^2}}$, which gives

$$\widehat{\widetilde{\mu}} = \Delta^{-1}\frac{\displaystyle\sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}\left(y_{t\Delta} - y_{(t-1)\Delta}\right)}{\widehat{\widetilde{\sigma}^2} y_{(t-1)\Delta}^2 + 2\widehat{\omega} y_{(t-1)\Delta} + \widehat{\sigma^2}}}{\displaystyle\sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}^2}{\widehat{\widetilde{\sigma}^2} y_{(t-1)\Delta}^2 + 2\widehat{\omega} y_{(t-1)\Delta} + \widehat{\sigma^2}}} =: \Delta^{-1}\frac{\widehat{A}_N^*}{\widehat{B}_N^*}. \tag{3.5.10}$$

The following theorem provides asymptotic theory for $\widehat{\widetilde{\mu}}$ and $\widehat{\rho} := \exp\left(\widehat{\widetilde{\mu}}\Delta\right)$.

Theorem 3.5.2 *In model (3.5.1) assume $\widetilde{\sigma}_0^2 > 0$. When $T \to \infty$ and $\Delta \to 0$, we have $\widetilde{\mu} \xrightarrow{p} \widetilde{\mu}_0$. Additionally, if $T\Delta \to 0$, we have,*

$$\sqrt{T}\left(\widehat{\widetilde{\mu}} - \widetilde{\mu}_0\right) \xrightarrow{L} \mathcal{N}\left(0, V_2^{-1}\right), \tag{3.5.11}$$

where

$$V_2 = \begin{cases} E\left(\dfrac{y_t^2}{\widetilde{\sigma}_0^2 y_t^2 + 2\omega_0 y_t + \sigma_0^2}\right) & \text{if } \kappa = \widetilde{\mu} + \dfrac{1}{2}\widetilde{\sigma}^2 < 0 \\ \widetilde{\sigma}_0^{-2} & \text{if } \kappa = \widetilde{\mu} + \dfrac{1}{2}\widetilde{\sigma}^2 \geq 0 \end{cases}. \tag{3.5.12}$$

Theorem 3.5.3 *In model (3.5.1), assume $\widetilde{\sigma}_0^2 > 0$. When T is fixed and $\Delta \to 0$, we*

have $\widehat{\rho} \xrightarrow{p} \rho_0$ *and its asymptotic distribution is*

$$\frac{1}{\Delta}\left(\widehat{\rho}-\rho_0\right) \xrightarrow{L} \mathcal{N}\left(0,(TV_2)^{-1}\right), \tag{3.5.13}$$

where

$$\widehat{\rho}=\frac{\sum_{t=1}^{M\times N}\dfrac{y_{(t-1)\Delta}y_{t\Delta}}{\widehat{\widetilde{\sigma}}^2 y_{(t-1)\Delta}^2+2\widehat{\omega}y_{(t-1)\Delta}+\widehat{\sigma^2}}}{\sum_{t=1}^{M\times N}\dfrac{y_{(t-1)\Delta}^2}{\widehat{\widetilde{\sigma}}^2 y_{(t-1)\Delta}^2+2\widehat{\omega}y_{(t-1)\Delta}+\widehat{\sigma^2}}}=:1+\frac{\widehat{A}_N^*}{\widehat{B}_N^*}.$$

According to Theorems 3.5.2 and 3.5.3 the estimates $\widehat{\widetilde{\mu}}$ and $\widehat{\rho}$ continue to have asymptotic normal distributions under infill asymptotics. This convenient feature allows us to apply the testing procedures proposed in the previous section after making a minor change in the variance of the limit distribution to accommodate endogeneity.

3.6 Simulations

This section reports the results of Monte Carlo simulations designed to evaluate the performance of the two-stage estimator. We also examine the finite sample adequacy of the asymptotic theory for the test statistics developed in Sections 3.4 and 3.5. To save space, we only report the results for model (3.2.2) under explosiveness. More tables reporting bias and standard errors under stationary, unstable, and locally explosive cases are given in the online supplement to this paper (Tao et al., 2018).

The simulations involved 10,000 replications of sample paths generated from model (3.2.2) under explosiveness with parameter values $\widetilde{\mu}=1$, $\widetilde{\sigma}=1$, $\sigma=1$, and with initial condition $y_0=10$.[8] Since $\phi>0$ this generating process leads to explosiveness. In the first experiment, we set the time span $T=5$, but varied Δ from $1/252$ to $1/19656$ and varied M from 21, 63 to 252. $\Delta=1/252$ corresponds to daily observations whereas $\Delta=1/19656$ corresponds to 5-minute (high

[8]We also report bias and standard errors under stationary, unstable, and locally explosive cases in the online appendix.

frequency) observations. When $\Delta = 1/252$, $M = 21$, 63 and 252 implies a corresponding block size that is monthly, quarterly, and annual, respectively. When $\Delta = 1/19656$, we report the estimation bias and standard errors by holding the number of observations for calculating the realized volatilities (M) constant as in a daily frequency. In panel A of Table 3.2, we report the bias and the standard errors of the two-stage estimates when there is no endogeneity in the model, i.e. when $\gamma = 0$, and in panel B, we report the corresponding results for the model with endogeneity, specifically with $\gamma = 0.8$. The bias and the standard errors are computed using 5,000 replications.

Table 3.2: Bias and standard errors of the two-stage estimates for different Δ and M and a fixed $T(= 5)$. The parameter values are $\widetilde{\mu} = 1$, $\widetilde{\sigma}^2 = 1$, $\sigma^2 = 1$. $y_0 = 10$.

Params	Case I: $\Delta = 1/252$					
	$M = 21$		$M = 63$		$M = 252$	
	Bias	S.E.	Bias	S.E.	Bias	S.E.
Panel A: $\gamma = 0$						
$\widetilde{\mu}$	-0.0328	0.4942	-0.0347	0.4946	-0.0455	0.4938
$\widetilde{\sigma}^2$	-0.0093	0.0471	-0.0135	0.0493	-0.0190	0.0611
σ^2	4.6415	13.0303	6.1285	18.5101	28.7040	138.7416
κ	-0.0375	0.4952	-0.0414	0.4958	-0.0549	0.4962
ρ	-1.3e-04	0.0020	-1.4e-04	0.0020	-1.8e-04	0.0020
ϕ	-0.0282	0.4943	-0.0279	0.4945	-0.0360	0.4933
Panel B: $\gamma = 0.8$						
$\widetilde{\mu}$	-0.0487	0.5213	-0.0511	0.5214	0.0208	1.5405
$\widetilde{\sigma}^2$	0.0326	0.0974	0.0350	0.1119	0.0776	0.2239
σ^2	19.2645	45.3508	28.3990	76.9436	193.6521	1.1e+03
γ	-0.4179	0.6293	-0.4519	0.6521	-0.6296	0.7499
κ	-0.0324	0.5208	-0.0336	0.5199	0.0596	1.5431
ρ	-1.9e-04	0.0021	-2.0e-04	0.0021	9.9e-05	0.0064
ϕ	-0.0650	0.5262	-0.0686	0.5289	-0.0180	1.5461
Params	Case II: $\Delta = 1/19656$					
	$M = 21$		$M = 63$		$M = 252$	
	Bias	S.E.	Bias	S.E.	Bias	S.E.
Panel A: $\gamma = 0$						
$\widetilde{\mu}$	-0.0279	0.5173	-0.0279	0.5173	-0.0279	0.5173
$\widetilde{\sigma}^2$	0.0014	0.0056	4.7e-04	0.0055	-8.3e-05	0.0055
σ^2	0.2518	1.3496	0.2385	1.3241	0.2414	1.3443
κ	-0.0272	0.5172	-0.0276	0.5172	-0.0279	0.5172
ρ	-1.4e-06	2.6e-05	-1.4e-06	2.6e-05	-1.4e-06	2.6e-05
ϕ	-0.0285	0.5174	-0.0281	0.5174	-0.0278	0.5174
Panel B: $\gamma = 0.8$						
$\widetilde{\mu}$	-0.0368	0.5194	-0.0368	0.5194	-0.0368	0.5194
$\widetilde{\sigma}^2$	0.0037	0.0105	0.0028	0.0119	0.0028	0.0104
σ^2	1.4455	3.7584	1.4215	3.6797	1.4215	3.7170
γ	-0.0884	0.1778	-0.0874	0.1750	-0.0874	0.1762
κ	-0.0349	0.5193	-0.0354	0.5193	-0.0354	0.5193
ρ	-4.9e-05	1.1e-05	-4.1e-05	2.4e-05	-4.1e-05	2.4e-05
ϕ	-0.0386	0.5196	-0.0382	0.5196	-0.0382	0.5196

First, from Table 3.2 it is apparent that when the sampling frequency increases the parameters $\widetilde{\sigma}^2$, γ and σ^2 are all better estimated in terms of bias and standard error. On the other hand, there is little improvement in the estimation of $\widetilde{\mu}$ because the

time span does not change. This finding corroborates the asymptotic theory for $\widetilde{\mu}$ given in Theorem 3.4.2 and also supports results found in Yu (2012). Furthermore, due to the difference in the convergence rates shown in Theorems 3.4.1 and 3.4.2, the bias and the standard errors of $\widehat{\kappa}$ and $\widehat{\phi}$ are mainly determined by those of $\widehat{\widetilde{\mu}}$, which explains why estimation performance of $\widehat{\kappa}$ and $\widehat{\phi}$ does not improve as sampling frequency increases. Finally, bias and standard errors both appear reasonably robust across different values of M.

In the second experiment, we fix $\Delta = 1/252$, but vary T from 30 to 60 and M from 21, 63 to 252. In Panel A of Table 3.3, we report the bias and the standard errors of the two-stage estimators across 5,000 simulated samples for the model without endogeneity. The same experiment is repeated for the model with endogeneity and the results are reported in Panel B. Several findings are evident from Table 3.3. First, as the time span enlarges, sharp reductions occur in the bias and standard error of $\widehat{\widetilde{\mu}}$. Combined with the results of Table 3.2, this finding suggests that time span, not sampling frequency, is the primary influence on performance of $\widehat{\widetilde{\mu}}$. Second, the bias and standard errors of $\widehat{\widetilde{\sigma}^2}$, $\widehat{\gamma}$ and $\widehat{\sigma^2}$ do not change significantly as T increases. Finally, both bias and standard errors are again robust with respect to M.

From Table 3.2 and Table 3.3, it is evident that the proposed two-stage method is effective in estimating $\widetilde{\mu}$, γ, $\widetilde{\sigma}^2$, κ, ρ, ϕ even in the presence of endogeneity. While the estimate of σ^2 is less satisfactory,[9] this outcome is unsurprising because when $\kappa > 0$, y_t^2 grows exponentially with t. Hence, estimates of γ and σ^2 are dominated by the component $\widetilde{\sigma}^2 y_t^2$ in $\widetilde{\sigma}^2 y_t^2 + 2\omega y_t + \sigma^2$ when t is large. More importantly, the three forms of explosive behavior do not depend on γ and σ^2 in that case. Hence, it is expected that the performance of $\widehat{\gamma}$ and $\widehat{\sigma^2}$ will have little impact on the performance of the proposed t-tests and the time-stamping strategy.

The third experiment is designed to evaluate the performance of the test statistics

[9] In both Table 2 and Table 3, the bias and S.E. of $\widetilde{\sigma}^2$ and σ^2 evidently increase with M. The explanation is that, given Δ and T, as M increases the effective sample size in the first stage estimation (N) decreases, which harms performance of $\widehat{\widetilde{\sigma}^2}$ and $\widehat{\sigma^2}$.

Table 3.3: Bias and standard error of the two-stage estimates for different T and M and fixed $\Delta(= 1/252)$. The parameter values are $\widetilde{\mu} = 1$, $\widetilde{\sigma}^2 = 1$, $\sigma^2 = 1$. $y_0 = 10$.

Params	Case I: $T = 30$					
	$M = 21$		$M = 63$		$M = 252$	
	Bias	S.E.	Bias	S.E.	Bias	S.E.
Panel A: $\gamma = 0$						
$\widetilde{\mu}$	-0.0058	0.1861	-0.0060	0.1861	-0.0075	0.1860
$\widetilde{\sigma}^2$	-0.0031	0.0175	-0.0066	0.0179	-0.0077	0.0208
σ^2	3.8582	11.6119	5.1172	16.6821	24.0530	136.4710
κ	-0.0073	0.1864	-0.0093	0.1862	-0.0113	0.1860
ρ	-2.3e-05	7.4e-04	-2.4e-05	7.4e-04	-2.9e-05	7.4e-04
ϕ	-0.0042	0.1862	-0.0027	0.1864	-0.0036	0.1866
Panel B: $\gamma = 0.8$						
$\widetilde{\mu}$	-9.7e-04	0.1796	-0.0013	0.1796	0.0045	0.6583
$\widetilde{\sigma}^2$	-1.6e-04	0.0188	-0.0035	0.0193	-0.0034	0.0225
σ^2	12.4166	32.3889	17.3784	53.4253	81.2221	396.9127
γ	-0.2879	0.5323	-0.3136	0.5496	-0.3924	0.6223
κ	-0.0010	0.1799	-0.0030	0.1798	0.0028	0.6584
ρ	-3.6e-06	7.2e-04	5.0e-06	7.2e-04	2.2e-05	0.0027
ϕ	-8.9e-04	0.1797	4.2e-04	0.1800	0.0063	0.6584
Params	Case II: $T = 60$					
	$M = 21$		$M = 63$		$M = 252$	
	Bias	S.E.	Bias	S.E.	Bias	S.E.
Panel A: $\gamma = 0$						
$\widetilde{\mu}$	-0.0014	0.1301	-0.0015	0.1301	-0.0023	0.1302
$\widetilde{\sigma}^2$	-0.0023	0.0123	-0.0058	0.0126	-0.0068	0.0147
σ^2	3.8333	11.6501	5.0750	16.5899	24.4001	137.0690
κ	-0.0026	0.1304	-0.0044	0.1303	-0.0057	0.1303
ρ	-5.6e-06	5.2e-04	-6.0e-06	5.2e-04	-9.0e-06	5.2e-04
ϕ	-2.7e-04	0.1302	0.0013	0.1303	0.0011	0.1304
Panel B: $\gamma = 0.8$						
$\widetilde{\mu}$	0.0016	0.1277	0.0014	0.1277	0.0127	0.5476
$\widetilde{\sigma}^2$	-0.0010	0.0126	-0.0043	0.0130	-0.0048	0.0153
σ^2	12.1161	31.6064	16.9538	52.3937	80.3542	393.2385
γ	-0.2768	-0.5228	-0.3011	0.5417	-0.3848	0.6138
κ	0.0011	0.1279	-7.8e-04	0.1278	0.0103	0.5475
ρ	6.4e-06	5.1e-04	5.7e-06	5.1e-04	5.3e-05	0.0022
ϕ	0.0021	0.1278	0.0036	0.1279	0.0151	0.5477

introduced in Remark 3.4.8. To do so, we simulate 5,000 sample paths from model (3.2.2) and from model (3.5.1) with $\gamma = 0.8$, and calculate the power and size of the three tests. In the first case, we set the nominal size to 5%, $M = 21$ and $\Delta = 1/252$, but vary the time span T. In the second case, we set the nominal size to 5%, $M = 21$ and $T = 10$, but vary the sampling interval Δ. The values of Δ correspond to 1-day (1/252), 1-hour (1/1638), and 30-minute (1/3276), respectively. Results for the power and size are reported in Table 3.4 and 3.5.

Table 3.4: Power and size of the t tests under different forms of unstable/explosive behavior with fixed Δ ($= 1/252$) and growing time span T.

T		t_κ				$t_{\widetilde{\mu}}$				t_ϕ		
$\gamma = 0$	Size	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 2$	Size	$\widetilde{\mu} = 0.5$	$\widetilde{\mu} = 1$	$\widetilde{\mu} = 2$	Size	$\phi = 0.5$	$\phi = 1$	$\phi = 2$
5	0.039	0.256	0.658	0.995	0.044	0.281	0.706	0.997	0.047	0.292	0.717	0.998
10	0.034	0.330	0.807	1.000	0.039	0.441	0.912	1.000	0.047	0.468	0.933	1.000
15	0.037	0.409	0.895	1.000	0.038	0.547	0.966	1.000	0.047	0.600	0.984	1.000
30	0.038	0.580	0.984	1.000	0.038	0.747	0.999	1.000	0.048	0.853	1.000	1.000
T		t_κ				$t_{\widetilde{\mu}}$				t_ϕ		
$\gamma = 0.8$	Size	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 2$	Size	$\widetilde{\mu} = 0.5$	$\widetilde{\mu} = 1$	$\widetilde{\mu} = 2$	Size	$\phi = 0.5$	$\phi = 1$	$\phi = 2$
5	0.045	0.261	0.634	0.986	0.045	0.273	0.669	0.994	0.044	0.270	0.682	0.996
10	0.041	0.438	0.890	1.000	0.046	0.482	0.915	1.000	0.044	0.453	0.918	1.000
15	0.041	0.583	0.975	1.000	0.054	0.665	0.981	1.000	0.046	0.593	0.981	1.000
30	0.040	0.838	0.999	1.000	0.053	0.926	1.000	1.000	0.047	0.861	1.000	1.000

Table 3.5: Power and size of the t tests under different forms of unstable/explosive behavior with fixed T ($= 10$) and shrinking sampling interval Δ.

Δ		t_κ				$t_{\widetilde{\mu}}$				t_ϕ		
$\gamma = 0$	Size	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 2$	Size	$\widetilde{\mu} = 0.5$	$\widetilde{\mu} = 1$	$\widetilde{\mu} = 2$	Size	$\phi = 0.5$	$\phi = 1$	$\phi = 2$
1/252	0.034	0.330	0.807	1.000	0.039	0.441	0.912	1.000	0.047	0.468	0.933	1.000
1/1638	0.035	0.326	0.813	0.999	0.044	0.441	0.916	1.000	0.048	0.473	0.930	1.000
1/3276	0.036	0.334	0.816	1.000	0.048	0.446	0.910	1.000	0.050	0.474	0.927	1.000
Δ		t_κ				$t_{\widetilde{\mu}}$				t_ϕ		
$\gamma = 0.8$	Size	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 2$	Size	$\widetilde{\mu} = 0.5$	$\widetilde{\mu} = 1$	$\widetilde{\mu} = 2$	Size	$\phi = 0.5$	$\phi = 1$	$\phi = 2$
1/252	0.041	0.438	0.890	1.000	0.046	0.482	0.915	1.000	0.044	0.453	0.918	1.000
1/1638	0.041	0.427	0.887	1.000	0.052	0.473	0.915	1.000	0.047	0.459	0.921	1.000
1/3276	0.043	0.442	0.885	0.999	0.054	0.483	0.911	1.000	0.051	0.460	0.916	1.000

The simulation results show that size distortion of the proposed tests for different types of explosive behavior becomes smaller when data is sampled more frequently. In addition, the power of the tests rises rapidly as the sample size increases (by either increasing the time span T or shrinking the sampling interval Δ) and for greater departures of the true parameters from the null. The reason for smaller size distortions as Δ shrinks is that the approximation errors of the distributions of $\Delta^{-1}(\widehat{\beta}_\kappa - \beta_\kappa^0)$ and $\Delta^{-1}(\widehat{\beta}_\phi - \beta_\phi^0)$ to the limiting distributions are of order $O(\sqrt{\Delta})$; see (A.2.9).

Next, we check size and power of the tests under endogeneity. Simulations are generated by setting $\widetilde{\sigma}^2 = 1$ and $\sigma^2 = 100$ with $y_0 = 10$ under the sampling scheme $\Delta = 1/252$, $M = 21$ and $T = \{5, 10, 15, 30\}$. Sample paths are generated in 1000 replications for parameter values $\widetilde{\mu} = \{-1, -0.5, 0.5, 1\}$ and for correlation coefficients $\gamma = \{0, 0.04, 0.08, 0.4, 0.8\}$. The results in Table 3.6 show that size distortion is very small under all parameter scenarios and that test power grows more slowly as the process becomes more unstable. This phenomenon is due to the structure of the quadratic variation $\widetilde{\sigma}^2 y_t^2 + 2\omega y_t + \sigma^2$ under endogeneity. When the process y_t is more unstable, the signal from y_t^2 is stronger and a much larger value of ω is needed for the component $2\omega y_t$ in the quadratic variation to enhance the probability of rejecting the null. Also, as expected, the power of the test increases with the increase in sample size.

Table 3.6: Power and size of the LR test for endogeneity.

T	Stationary ($\widetilde{\mu} = -1$)					Unstable ($\widetilde{\mu} = -0.5$)				
	$\gamma = 0$	0.04	0.08	0.4	0.8	$\gamma = 0$	0.04	0.08	0.4	0.8
5	0.0470	0.0780	0.1950	0.9510	0.9940	0.0500	0.0810	0.1920	0.9210	0.9850
10	0.0440	0.1290	0.4110	0.9990	1.0000	0.0490	0.1200	0.4030	0.9970	1.0000
15	0.0540	0.1960	0.6000	1.0000	1.0000	0.0560	0.2160	0.5970	1.0000	1.0000
30	0.0490	0.3820	0.8790	1.0000	1.0000	0.0540	0.3990	0.8920	1.0000	1.0000
T	Locally Explosive ($\widetilde{\mu} = 0.5$)					Explosive ($\widetilde{\mu} = 1$)				
	$\gamma = 0$	0.04	0.08	0.4	0.8	$\gamma = 0$	0.04	0.08	0.4	0.8
5	0.0600	0.0880	0.1580	0.7260	0.8710	0.0460	0.0580	0.1070	0.5940	0.7850
10	0.0460	0.1120	0.2770	0.8650	0.9490	0.0530	0.0690	0.1640	0.7120	0.8800
15	0.0540	0.1210	0.3680	0.9090	0.9690	0.0540	0.0870	0.1780	0.7310	0.9020
30	0.0520	0.1910	0.5270	0.9480	0.9860	0.0540	0.0810	0.1900	0.7530	0.9130

The final experiment checks performance of the proposed tests of coefficient constancy, i.e. $\mathcal{H}_0 : \widetilde{\sigma}^2 = 0$. To do so, we simulate 10,000 sample paths from model (3.5.1) with different parameter values to cover the various explosive scenarios. Both size and power are calculated. More specifically, we vary $\widetilde{\mu}$ from -0.1 to 0.1, $\widetilde{\sigma}$ from 0 to 0.2 and $\gamma \in \{0, 0.8\}$ holding $\sigma = 1$, which covers all explosive scenarios. In these experiments, we set nominal size to 5%, $M = 21$ and $\Delta = 1/252$, but vary the time span T to control for sample sizes. Test size and power are reported in Table 3.7.

Table 3.7: Power and size of the modified *LBI*-test for different null models.

T	$\widetilde{\mu}=-0.1$				$\widetilde{\mu}=0$				$\widetilde{\mu}=0.1$			
	$\widetilde{\sigma}=0$	0.04	0.10	0.20	$\widetilde{\sigma}=0$	0.04	0.10	0.20	$\widetilde{\sigma}=0$	0.04	0.10	0.20
Panel A: $\gamma=0$												
5	0.049	0.173	0.868	0.998	0.047	0.155	0.787	0.998	0.050	0.494	0.925	0.998
10	0.047	0.285	0.993	1.000	0.050	0.338	0.984	1.000	0.050	0.966	0.999	1.000
15	0.049	0.345	0.998	1.000	0.046	0.504	0.998	1.000	0.046	0.998	1.000	1.000
30	0.047	0.391	1.000	1.000	0.047	0.794	1.000	1.000	0.053	1.000	1.000	1.000
Panel B: $\gamma=0.8$												
5	0.049	0.576	0.944	0.999	0.047	0.434	0.913	0.999	0.050	0.719	0.950	0.999
10	0.047	0.926	0.999	1.000	0.050	0.800	0.998	1.000	0.050	0.982	1.000	1.000
15	0.049	0.986	1.000	1.000	0.046	0.944	1.000	1.000	0.046	0.999	1.000	1.000
30	0.047	0.990	0.998	0.999	0.047	0.997	0.997	0.998	0.053	1.000	1.000	1.000

3.7 Empirical Studies

3.7.1 Daily data

As a practical illustration of our methods with real data, we analyze daily S&P 500 real prices over the period from December 31, 1927 to June 29, 2018 with $T = 91.5$. For the first stage estimation we use daily prices to calculate monthly realized variance and monthly realized quarticity within each month. Thus 1098 blocks are chosen with each block containing daily observations within each calendar month, i.e., $N = 1098$, $h = 1/12$. While on average there are 21 trading days within each month, the actual number of days within a month is year-dependent as well as month-dependent. Similarly, while on average there are 252 trading days within each year, the actual number of days within a year is year-dependent. If there are M days within a month, we set $\Delta = h/M$ for that month.

We first apply our estimation, testing, and time-stamping strategies to these daily price data based on the model with no endogeneity, that is, model (3.2.2).[10] Following PWY (2011), the initial window is taken as the first 5-year of the full sample. For comparison purposes, we also implement the BADF test of PWY and the BSADF test of PSY. The empirical results are shown in Figure 3.5, where we plot the test statistic sequences under the three forms of explosiveness and the test statistic se-

[10]We used model (3.2.2) largely because the bias in estimation of ω is relatively large in long-span, low-frequency samples and the bias becomes severe when the process is explosive (c.f., Lieberman and Phillips (2017a)). The methods of the present paper are more relevant in models without endogeneity when high-frequency data are unavailable.

quences under the assumption of time-invariant coefficients. We also plot the 95% critical values and the data in each panel.

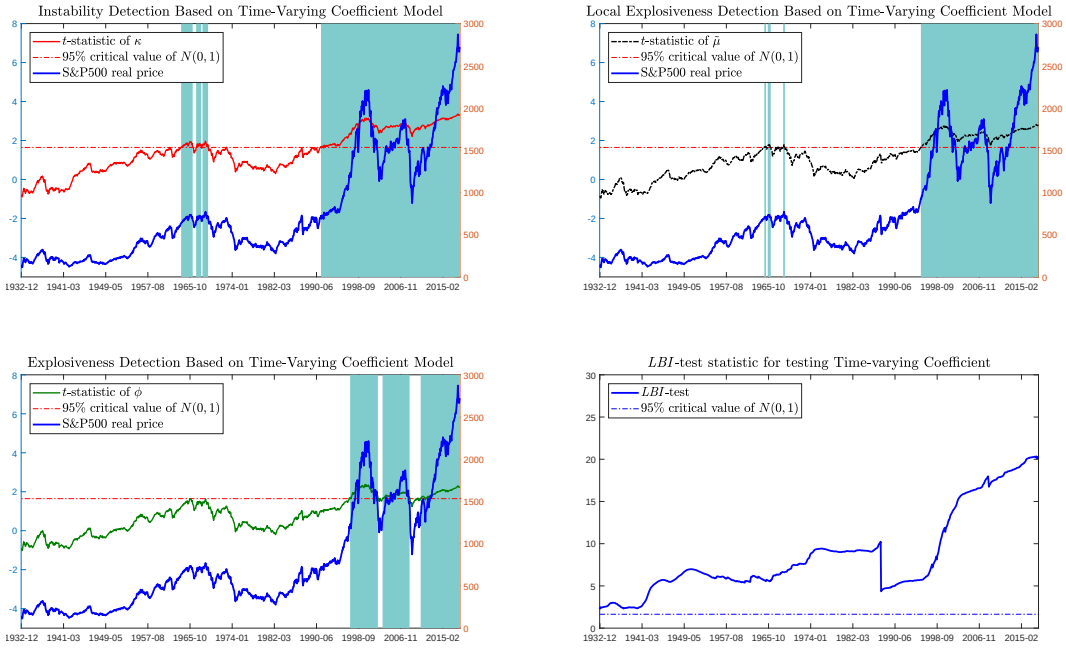


Figure 3.5: Date-stamping Explosive Periods in the S&P500 Real Price Index with Time-Varying Coefficient Model without Endogeneity.

The last panel in Figure 3.5 plots the recursive test statistic sequence for testing a time-invariant autoregressive coefficient. The test results suggest that throughout the whole sample period the data are well described by a model with time varying coefficients as the test statistic for a time varying coefficient always exceeds the 95% critical value. In fact, the test statistic generally increases as the time period expands. Although the value of the test statistic drops in October 1987, indicative of a possible change in market behavior at that point, it is still larger than the critical value. Evidence for time variation becomes much stronger from July 1995 onwards. This dating coincides well with the estimated origination dates of the three forms of extreme behavior indicated by the other three panels in Figure 3.5. For example, the first panel in Figure 3.5 indicates that real stock prices have experienced two major unstable periods. The unstable behavior is first detected at February 1964 and is interrupted twice over the succeeding period to May 1969. Afterwards real stock prices are not unstable or explosive until May 1991, from which point the instability continues until the end of the sample.

The second panel in Figure 3.5 indicates that real prices are not locally explosive between January 1928 to January 1965, at which point locally explosive (submartingale) behavior is detected. This behavior is interrupted twice over the succeeding period to January 1969. The locally explosive behavior is detected again in July 1995 which continues to the end of the sample. Since local explosiveness is more extreme behavior than instability, its duration tends to be much shorter than that of instability.

The third panel in Figure 3.5 indicates that the real price index does not experience explosive behavior between January 1928 and July 1997. An episode of explosive behavior is detected in July 1997, lasting for a few years and ending in June 2002. The end date corresponds to the termination of the well-known internet bubble period. Then, the test detects a second explosive period starting from May 2003 to August 2008, which corresponds to the pre global financial crisis period. The panel also suggests a further explosive episode starting in October 2010 and continuing to the end of the sample period. These three periods of major price escalation in the sample (namely, the second half of 1990s, the pre global financial crisis period, and the recovery from the global financial crisis) are all deemed to have explosiveness which seems to be a reasonable empirical finding. Time horizons for the three forms of extreme behavior detected by our method are summarized in a table in the online supplement (Tao, Phillips, and Yu (2018)).

For comparison purposes, Figure 3.6 plots the recursive BADF statistic (used in PWY), the recursive BSADF statistic (used in PSY), and corresponding 95% critical values together with the sample data in each panel. It is clear that both PWY and PSY tests identify explosive behavior in the second half of the 1990s earlier than the method proposed in the present paper. This early origination date identification is achieved by using a more restrictive reduced form autoregressive model. Interestingly, the PSY test recursion indicates two similar pronounced periods of explosive behavior, one in the second half of the 1990s and the other at the end of the sample, both matching those identified by methods of the present paper using a more

complex modeling framework. Time horizons for the explosive behavior detected by PWY and PSY are summarized in a table in Tao et al. (2018).

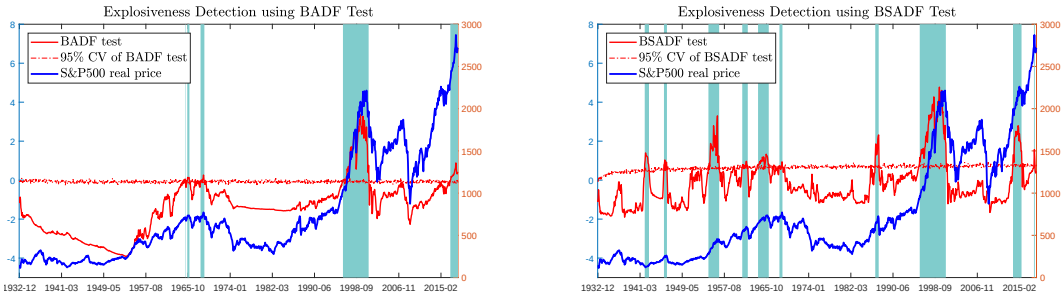


Figure 3.6: Date-stamping Explosive Periods in the S&P500 Real Price Index with Fixed Coefficient Model.

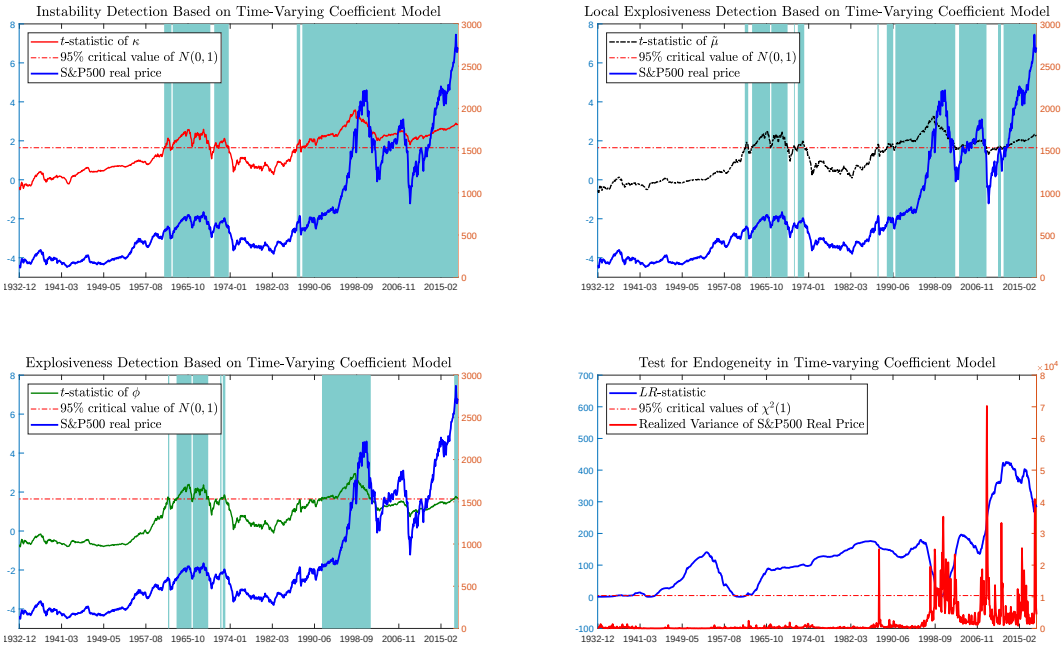


Figure 3.7: Date-stamping Explosive Periods in the S&P500 Real Price Index with Time-varying Coefficient Model with Endogeneity.

To address the possible presence of endogeneity we estimate the more general model in which endogeneity effects are permitted, i.e., model (3.5.1). The results are summarized by the recursions plotted in Figure 3.7. First, from the test for endogeneity it is apparent that the null of exogeneity is rejected almost everywhere throughout the entire sample, confirming that endogeneity is important in the generating mechanism for these data. From the plotted realized variance graphic in Figure 3.7 it is further evident that the rejection of exogeneity is closely associated

with the behavior of the quadratic variation of the process. This is explained by the fact that the likelihood ratio statistic is based on an objective function that is constructed using a central limit theorem (CLT) for the realized variance time series. The test statistic for endogeneity therefore captures differences in the realized variance estimates using different models, as is shown in Figure 3.8. Further, more nonstationary horizons are detected in the data when using the model with endogeneity. In particular, extreme behavior is detected in the form of both unstable and locally explosive behavior during price spike in 1987. Note that this explosive behavior is also reported in the PSY test. However, previous empirical evidence for explosiveness before global financial crisis (namely, from May 2003 to August 2008) disappears in the fitted model where endogeneity effects are incorporated in the autoregressive response mechanism. These findings indicate that endogeneity feedbacks in the random coefficient autoregressive model framework can play an important role in assessing evidence for various types of instability and explosiveness in the data. These time horizons for different form of extreme behavior are summarized in a table in Tao, Phillips, and Yu (2018).

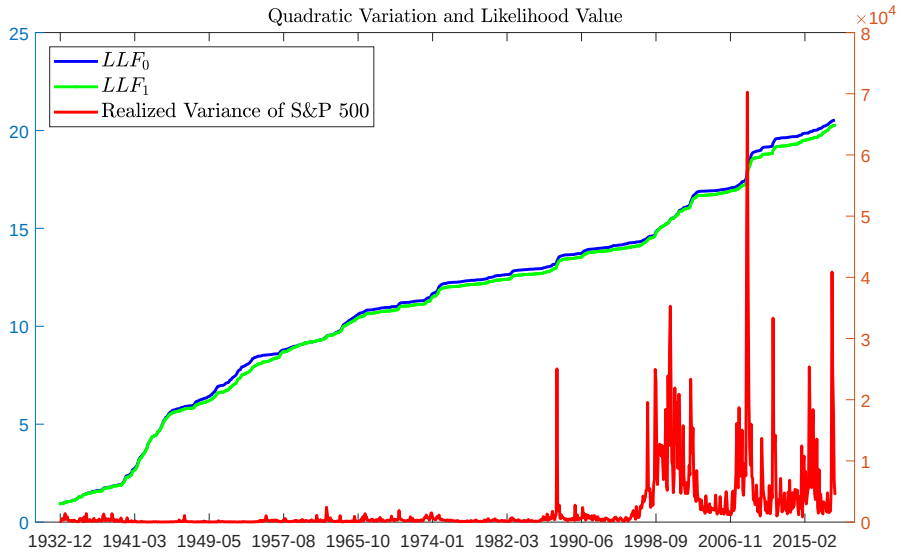


Figure 3.8: Realized Variance and Likelihood Values for Random Coefficient Autoregression.

3.7.2 Intra-day data

To further assess evidence for endogeneity and to reduce bias in the estimation of γ, we estimate the same model (3.5.1) using 5-minute high-frequency data for the real prices of the E-mini S&P 500 futures contract over the period from the closing on October 31, 1997 to the closing on October 31, 2013. Use of this high frequency intra-day data leads to a substantial increase in sample size, accords more closely with infill asymptotic theory, but has the limitation that the model itself abstracts from possible intra-day effects that are known to be present in ultra high frequency data. On the other hand, use of 5-minute data (rather than even higher frequency observations) helps to mitigate some of these intra-day effects and gives the benefit of bias reduction in estimation of the correlation between the equation errors and the random autoregressive coefficient, thereby improving estimation of the degree of endogeneity in the random coefficient driver variables.

We collect 16 years of 5-minute high-frequency S&P 500 real prices with $T = 16$. That is, in each trading day we have 79 real prices, sampled at 9:30am, 9:35am, ..., 3:55pm, 4:00pm. The overnight price movement is captured by the difference between the price observed at 4:00pm on day t and that at 9:30am on day $t + 1$. By doing so, we treat overnight price movements in the same way as intra-day price movements over 5-minute intervals.

For the first stage estimation we use price changes over 5-minute intervals and overnight to calculate monthly realized variance and monthly realized quarticity within each month. That is, 192 blocks are chosen with each block containing high-frequency data within each calendar month, i.e., $N = 192$, $h = 1/12$. If there are M days within a month we set $\Delta = h/(78 \times M)$ for that month. Model (3.5.1) is fitted recursively with an initial window size of 5 years. The empirical results are summarized in Figure 3.9 on monthly basis.

The recursive test statistic graphics in Figure 3.9 indicate that over this sample period and allowing for high frequency fluctuations the data are unstable but not

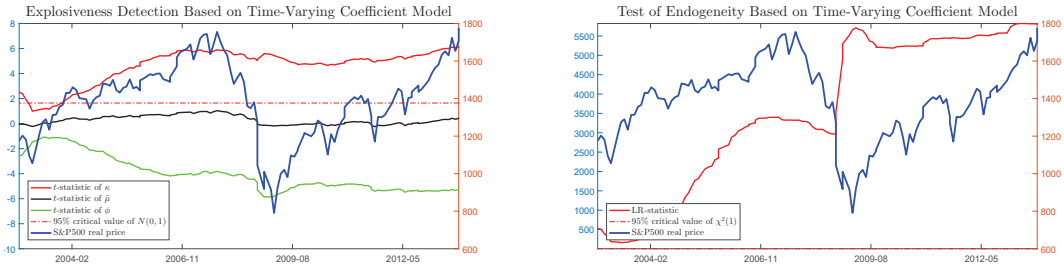


Figure 3.9: Testing Explosiveness and Endogeneity in the 5-minute S&P 500 Real Price Index with Time-varying Coefficient Model.

locally explosive or explosive. Based on the simulation findings in the previous section, estimates of the endogeneity parameter γ can be expected to have reasonably small bias at this frequency and the t-tests to have good size and power. From the second panel in Figure 3.9, the LR-test for endogeneity always exceeds the 5% critical value of the χ^2_1 distribution, which reinforces from the 5-minute high-frequency data the evidence in support of endogenous effects on the autoregressive coefficient found in the daily-frequency sample.

3.8 Summary and Conclusions

This paper introduces a continuous time model for financial data where the persistence parameter is allowed to be random and time varying. The model has an analytical solution and an exact discrete time representation which make analysis convenient for studying the properties of the system that are associated with extreme sample path behavior. The discrete time model relates to some models already in the literature, including the stochastic unit root model (Granger and Swanson (1997); Lieberman and Phillips (2014); Lieberman and Phillips (2017c)) and the near-explosive random coefficient model of Banerjee et al. (2017). The statistical properties of our model reveal three different forms of potential extreme behavior in generated sample paths: instability, local explosiveness, and explosiveness. These forms of extreme behavior depend directly on the values of model parameters, including the possible presence of endogeneity in the random autoregressive coefficient.

A novel two-stage estimation method that relies on empirical quadratic variation is developed to estimate the model parameters. Limit theory is developed using an infill asymptotic scheme that provides a convenient basis for testing parameter constancy and the various forms of extreme sample path behavior. The test statistics all have asymptotically pivotal standard normal distributions which makes implementation of the tests straightforward in practical work. Similar to other recent work in the literature on bubbles, a time-stamping strategy is proposed to detect origination and termination dates of extreme behavior.

In an empirical application to daily S&P 500 real prices between December 31, 1927 and June 29, 2018. Strong evidence against parameter constancy is found in the whole sample period and this evidence strengthens after July 1997, leading to a finding of long durations of parameter instability in the model. Three periods of explosive instability in the data match well with observed periods of major price escalation in the data and these largely overlap with the periods of price exuberance identified in earlier work. Tests for endogeneity in these data provide strong evidence in support of endogenous feedbacks in the random coefficient model framework that materially influence quadratic variation and hence recursive estimates of realized variation in the data. The empirical findings of extreme sample path behavior in real S&P 500 stock prices are broadly in line with the conclusions of other recent work on stock price exuberance but now provide new evidence against parameter constancy and in support of the role of endogenous feedbacks that influence autoregressive behavior and the time forms of extreme sample paths.

Chapter 4 A Dynamic Network Perspective on Cryptocurrencies

4.1 Introduction

The invention of Bitcoin (Nakamoto, 2008) spurred the creation of many cryptocurrencies (cryptos hereafter) commonly known as *Altcoins*. As of December 31, 2018, more than 1500 cryptos are actively traded worldwide, with a market capitalization of more than 200 billion USD. The growing number of Altcoins led investors to investigate interrelationships between Altcoins to make a profit. Unlike stocks that we can group into different industries by GIC or SIC, there are no stringent criteria to classify cryptos. By virtue of network analysis, we develop a covariate-assisted spectral clustering (CASC) method that accommodates important network features such as connection sparsity, degree heterogeneity, and relation asymmetry, to study the interrelationships between cryptos systematically. We thereby provide a novel angle to study the market segmentation problem of cryptos and other financial instruments.

The crypto market is distinct from the equity market in various aspects, which hinders the application of traditional classification methodology. Given that both cryptos and stocks are traded at high frequency, return information is particularly important as it serves as timely information to understand the dynamics of the market structure. According to market efficiency, the covariance between the prices of speculative assets cannot exceed the covariance between their fundamental information. Consequently, in the equity market, return co-movement is frequently

adopted to project the fundamental similarity between stocks. However, excess return co-movement has been widely documented in the literature (see, e.g., Kumar and Lee, 2006; Boyer, 2011) and it is more significant in the crypto market given the strong behavioral bias of market participants and high information uncertainty of its future cash flows. Inspired by Hoberg and Phillips (2016), who ameliorate industry classification by studying a set of dynamic industry structures generated from product differentiation and competition, we use crypto's contract information to help identify the fundamental similarities between cryptos. In particular, we extract the most fundamental characteristics of each mining contract, that is, the cryptographic algorithm and proof types, as additional input for clustering analysis. As we expected, our method shows superior classification accuracy over state-of-the-art methods available in the literature. In particular, cryptos in the same group show stronger return co-movement than the cross-group return co-movement across all empirical settings. Moreover, within-group cryptos show stronger connections in algorithms and proof types than cross-group cryptos do.

To understand the economic meaning of the latent group structure, we conduct several tests to verify the asset pricing implication of the grouping results. Acemoglu et al. (2012) proposed a theoretical framework to model spillover effects through sector-level shocks. The model suggests that if the linkages in the intersectoral network are sufficiently asymmetric, then sectoral shocks might not cancel out through diversification, but aggregate into macroeconomic fluctuations. Ahern (2013) further pointed out that the stocks with higher incoming linkages tend to receive more shocks from linked stocks and thus require a higher risk premium. Motivated by these results, we construct a cross-sectional portfolio by sorting on group centrality and show that high-centrality cryptos require a higher risk premium than the low-centrality ones. Next, we investigate whether other factors such as liquidity (Amihud and Mendelson, 1986), investor attention (Liu and Tsyvinski, 2018), and macro uncertainty (Baker et al., 2016) could possibly explain this augmented risk premium. Our results suggest that the return predictability of centrality survives

after controlling for all of these factors. Hence, it provides an important empirical implication for both academic studies and participants in the crypto market.

This paper develops statistical theory for dynamic networks and thereby makes several important contributions to classical finance as well as FinTech. First, we offer a network angle to study the crypto market by connecting cryptos according to their inter-predictive relationship estimated by adaptive LASSO. Second, we provide a new set of quantitative tools to study crypto market segmentation that can be applied to a wide variety of assets. Specifically, we extend the static spectral clustering methods (Binkiewicz et al., 2017; Zhang et al., 2018, among others,) to identify communities in dynamic networks with both time-evolving membership and node covariates. To make full use of the relevant information, we address the challenges of the features of the real data, namely, time dependency, degree heterogeneity, sparsity, and node covariates. Our proposed community detection method can resolve the aforementioned data issues. The methodology we present can also be extended to cover more asset-specific characteristics to achieve higher classification accuracy.

In addition, we deepen the understanding of the crypto market in terms of both market segmentation and portfolio management. Intensive research in this area considers asset pricing inferences from different angles, but there is limited work that shows the economic link between crypto fundamentals and its performance. Härdle et al. (2019) suggest crypto dynamics as an extraordinary research opportunity for academia and provide some insights into the mechanics of this market. Härdle and Trimborn (2015) construct the CRIX (thecrix.de), a market index consisting of a selection of cryptos representative of the whole crypto market. Given the low liquidity in the current Altcoin market compared to traditional assets, Trimborn et al. (2019) propose a Liquidity Bounded Risk-return Optimization (LIBRO) approach that accounts for liquidity issues by studying the Markowitz framework under liquidity constraints. Chen et al. (2018) propose an option pricing technique for cryptos based on a stochastic volatility model with correlated jumps. Lee et al. (2018) compare cryptos with traditional asset classes and find that cryptos provide additional

diversification to mainstream assets, hence improving the portfolio performance. Petukhina et al. (2018) characterize the effects of adding cryptos to the set of traditional eligible assets in portfolio management and find that cryptos can significantly improve the risk-return profile of mainstream asset portfolios. Our results provide new insights into the fundamentals of the crypto market structure by dividing them into different groups. We find that cryptos' fundamentals have very different features from those of traditional assets, and these features indeed affect a crypto's price evolution.

The remainder of the paper is organized as follows. In section 4.2, we introduce the model and method to estimate the dynamic group structure and demonstrate the effectiveness of our method via simulation. In section 4.3, we employ our method to identify the latent group structure of cryptos and provide its economic interpretation. Then, in section 4.4, we check the time series and cross-sectional return predictability and demonstrate its portfolio implications. We conclude in section 4.5. All proofs and technical details are provided in the supplement. R codes to implement the algorithms are available at QuantNet (quantlet.de) by searching the keyword "CASC."

4.2 Models and Methodology

In the equity market, network structures are powerful in revealing risk percolations in assets such as firms, industries, and financial instruments (Cohen and Frazzini, 2008; Aobdia et al., 2014; Acemoglu et al., 2015; Chen et al., 2019, see, e.g.,). The latest study, Herskovic (2018), constructs a sector level network based on the Bureau of Economic Analysis (BEA) Input-Output Accounts. Here, we borrow the network idea to model the interdependencies in between cryptos, such as technological similarities and return co-movements. However, just applying a network view on cryptos will not give us any insights into the dominant elements of the market. We therefore represent the adjacency matrices stochastically via a block structure to identify

the latent communities. To build such a stochastic blockmodel with time-varying communities, we need to establish a more advanced methodology to identify group memberships. Based on adaptive LASSO in a 60-day rolling window, we generate a time series of adjacency matrices. By imposing an assumption on the switch in group memberships, we can uniformly identify communities consistently. We base our numerical implementation of this procedure on spectral clustering. Binkiewicz et al. (2017) show that the classification accuracy of the spectral clustering method can be improved by introducing covariate assistance. Here, we present an extension of the static covariate-assisted spectral clustering (CASC) algorithm to deal with the dynamic stochastic blockmodel and co-blockmodel. The theoretical justification and simulations also demonstrate the consistency of this method.

4.2.1 Dynamic network model with covariates

Undirected network

Consider a dynamic network defined as a sequence of random undirected graphs with N nodes, $G_{N,t}$, $t = 1, \cdots, T$, on the vertex set $V_N = \{v_1, v_2, \cdots, v_N\}$, which does not change over horizons. For each period, we model the unipartite network structure with the *spectral-contextualized degree-corrected stochastic blockmodel* (SC-DCBM) introduced by Zhang et al. (2018). Specifically, we generate the adjacency matrices A_t by

$$A_t(i,j) = \begin{cases} \text{Bernoulli}\{P_t(i,j)\}, & \text{if } i < j \\ 0, & \text{if } i = j \\ A_t(j,i), & \text{if } i > j \end{cases} \tag{4.2.1}$$

where $P_t(i,j) = \Pr\{A_t(i,j) = 1\}$. To reflect the group structure, the probabilities of a connection $P_t(i,j)$ at period t are blocked. In particular, denote $z_{i,t}$ as the group label of node i at time t; then, if $z_{i,t} = k$ and $z_{j,t} = k'$, then $P_t(i,j) = B_t(z_{i,t}, z_{j,t}) = B_t(k,k')$. Hence, for any $t = 1, \cdots, T$, we can obtain the population adjacency

matrix

$$\mathcal{A}_t \stackrel{\text{def}}{=} E(A_t) = Z_t B_t Z_t^\top, \tag{4.2.2}$$

where $Z_t \in \{0,1\}^{N\times K}$ is the *clustering matrix* such that there is only one 1 in each row and at least one 1 in each column.

Since the conventional stochastic blockmodel presumes that each node in the same group should have the same expected degrees, following Karrer and Newman (2011), we introduce the *degree parameters* $\psi = (\psi_1, \cdots, \psi_N)$ to capture the degree heterogeneity of the groups. In particular, the edge probability between node i and j at time t is

$$P_t(i,j) = \psi_i \psi_j B_t(z_{i,t}, z_{j,t}), \tag{4.2.3}$$

with the identifiability restriction

$$\sum_{i\in\mathcal{G}_k} \psi_i = 1, \quad \forall k \in \{1,2,\cdots,K\}. \tag{4.2.4}$$

where $\mathcal{G}_k$ is the set of nodes that belongs to the kth group. Denote Diag(ψ) by Ψ. The population adjacency matrices for the dynamic SC-DCBM is then:

$$\mathcal{A}_t = \Psi Z_t B_t Z_t^\top \Psi, \tag{4.2.5}$$

Define the regularized graph Laplacian as

$$L_{\tau,t} = D_{\tau,t}^{-1/2} A_t D_{\tau,t}^{-1/2}, \tag{4.2.6}$$

where $D_{\tau,t} = D_t + \tau_t I$ and D_t is a diagonal matrix with $D_t(i,i) = \sum_{j=1}^N A_t(i,j)$. As Chaudhuri et al. (2012) shows, regularization improves the spectral clustering performance, especially for sparse networks. We fix τ_t as the value of average node degree, that is, $\tau_t = N^{-1}\sum_{i=1}^N D_t(i,i)$.

Recent developments suggest that using node features or covariates can greatly improve classification accuracy. For example, Binkiewicz et al. (2017) add the co-

variance $XX^\top$, with $X \in [-J, J]^R$ being the node covariate matrix, to the regularized graph Laplacian and perform the spectral clustering on the static *similarity matrix*. We extend the static similarity matrix to cover the dynamic case below:

$$S_t = L_{\tau,t} + \alpha_t C. \tag{4.2.7}$$

where $C = XX^\top$ and $\alpha_t \in [0, \infty)$ is a tuning parameter that controls the informational balance between $L_{\tau,t}$ and X in the leading eigenspace of S_t. As a generalization of the model, Zhang et al. (2018) refines this by replacing C with $C_w = XWX^\top$, where W is some weight matrix. Finally, we substitute C with the new covariate-assisted component $C_t^w = XW_tX^\top$, and the population similarity matrix now becomes

$$\mathcal{S}_t = \mathcal{L}_{\tau,t} + \alpha_t \mathcal{C}_t^w, \tag{4.2.8}$$

where $\mathcal{L}_{\tau,t} = \mathcal{D}_{\tau,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{\tau,t}^{-1/2}$ and $\mathcal{C}_t^w = \mathcal{X}\mathcal{W}_t\mathcal{X}$.

The setup in (4.2.8) addresses several extensions of existing methods. First, W_t creates a time-varying interaction between different covariates. For instance, we may think of different refined algorithms that stem from the same origins. Such inheritance relationships will potentially lead to an interaction between the cryptos. In addition, over time, some algorithms may become more popular while the others may near extinction. Thus, this interaction would also change over time. These interactions are not included in C.

Second, we can easily select covariates by setting certain elements of W_t to zero. This is necessary as it helps us to model the evolution of technologies. At some point in time, some cryptographic technology may be eliminated due to upgrades or cracking. Therefore, W_t offers us the flexibility to exclude covariates, which we cannot do easily with C.

Lastly, the role of C is to link similarity in covariates to a high probability of node connection. However, this is questionable in crypto networks. Due to the open source nature of the blockchain, crypto developers can easily copy and paste the

source code and launch a new coin without any costs. Consequently, we observe a high degree of homogeneity in the crypto market. However, this homogeneity does not necessarily result in a co-movement of prices: some cryptos are negatively correlated. In this case, we may set $W_t(i,i)$ to be negative and C_t^w will eventually bring the cryptos with different technologies closer in the similarity matrix.

Directed network

To model the dynamic block structure in a directed network, we employ the dyanmic spectral-contextualized degree-corrected stochastic co-blockmodel (SC-DCcBM). For a directed network, the adjacency matrix A_t is not necessarily symmetric; that is,

$$A_t(i,j) = \begin{cases} \text{Bernoulli}\{P_t(i,j)\}, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases} \tag{4.2.9}$$

Similarly, define the regularized graph Laplacian $L_{\tau,t} \in \mathbb{R}^{N\times N}$ for the directed network as

$$L_{\tau,t} = D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2}, \tag{4.2.10}$$

where $D_{R,t}$ and $D_{C,t}$ are diagonal matrices with $D_{R,t}(i,i) = \sum_{j=1}^{N} A_t(i,j) + \tau_{R,t}$ and $D_{C,t}(i,i) = \sum_{j=1}^{N} A_t(j,i) + \tau_{C,t}$, where $\tau_{R,t}$ and $\tau_{C,t}$ are set to be the average row and column degrees at each period, respectively.

We now include the node covariates by constructing a similarity matrix from regularized graph Laplacian $L_{\tau,t}$ and covariate matrix X in the same way as in an undirected network; that is, for each $t = 1, \cdots, T$,

$$S_t = L_{\tau,t} + \alpha_t X W_t X^\top = D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2} + \alpha_t X W_t X^\top, \tag{4.2.11}$$

where $\alpha_t \in [0,\infty)$ is the tuning parameter. Then, let $Z_{R,t} \in \{0,1\}^{N_R\times K_R}$ and $Z_{C,t} \in \{0,1\}^{N_C\times K_C}$, such that there is only one 1 in each row and at least one 1 in each column. Let the block probability matrix in each period be $B_t \in [0,1]^{K_R\times K_C}$

with rank $K = \min\{K_R, K_C\}$. Then, the population adjacency matrix is

$$\mathcal{A}_t = E(A_t) = \Psi^R Z_{R,t} B_t Z_{C,t}^\top \Psi^C, \tag{4.2.12}$$

and the population regularized graph Laplacian is

$$\mathcal{L}_{\tau,t} = \mathcal{D}_{R,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{C,t}^{-1/2}. \tag{4.2.13}$$

Therefore, the population similarity matrix is

$$\mathcal{S}_t = \mathcal{L}_{\tau,t} + \alpha_t \mathcal{X} \mathcal{W}_t \mathcal{X}^\top. \tag{4.2.14}$$

By construction, we know $D_{R,t}(i,i) = \sum_{j=1}^N \mathbb{1}_{\{i\to j\}}^{(t)} + \tau_{R,t}$, which controls for the number of the parents of node j, and $D_{C,t}(i,i) = \sum_{i=1}^N \mathbb{1}_{\{j\to i\}}^{(t)} + \tau_{C,t}$, which controls the number of the offspring of node j. To analyze the asymmetric adjacency matrix A_t caused by directional information, Rohe et al. (2016) propose using the *singular value decomposition* instead of eigen-decomposition for the regularized graph Laplacian. The intuition behind this methodology is to use both the eigenvectors of $L_{\tau,t}^\top L_{\tau,t}$ and $L_{\tau,t} L_{\tau,t}^\top$, which contains information about "the number of common parents" and "the number of common offspring"; that is, for each $t = 1, \cdots, T$,

$$(L_{\tau,t}^\top L_{\tau,t})_{ab} = \frac{1}{\sqrt{D_{C,t}(a,a) D_{C,t}(b,b)}} \sum_{i=1}^N \frac{\mathbb{1}_{\{i\to a \text{ and } i\to b\}}^{(t)}}{D_{R,t}(i,i)},$$
$$(L_{\tau,t} L_{\tau,t}^\top)_{ab} = \frac{1}{\sqrt{D_{R,t}(a,a) D_{R,t}(b,b)}} \sum_{i=1}^N \frac{\mathbb{1}_{\{a\to i \text{ and } b\to i\}}^{(t)}}{D_{C,t}(i,i)}.$$

4.2.2 Dynamic CASC

To set up a dynamic CASC, we face two major difficulties: (i) defining W_t and (ii) estimating the similarity matrix with dynamic network information. For the first issue, we follow Zhang et al. (2018) by setting $W_t = X^\top L_{\tau,t} X$, which measures

the correlation between covariates along the graph. For the second issue, we follow Pensky and Zhang (2017) by constructing the estimator of $\mathcal{S}_t$ with a discrete kernel to bring in historical network information. Klochkov et al. (2019) present a similar idea. Specifically, we first pick an integer $r \geq 0$, obtain two sets of integers

$$\mathcal{F}_r = \{-r, \cdots, 0\}, \quad \mathcal{D}_r = \{T - r + 1, \cdots, T\},$$

and assume that $|W_{r,l}(i)| \leq W_{\max}$, where $W_{\max}$ is independent of r and i, and satisfies

$$\frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} i^k W_{r,l}(i) = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{if } k = 1, 2, \cdots, l. \end{cases} \tag{4.2.15}$$

Obviously, the $W_{r,l}$ is a discretized version of the continuous boundary kernel that weighs only the historical observations. This kernel assigns more recent similarity matrices higher scores. To choose an optimal bandwidth r, Pensky and Zhang (2017) propose an adaptive estimation procedure using Lepski et al. (1997)'s method. Here, we also employ their method and construct the estimator for edge connection matrices:

$$\widehat{\mathcal{S}}_{t,r} = \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,l}(i) S_{t+i}. \tag{4.2.16}$$

Once we obtain $\widehat{\mathcal{S}}_{t,r}$, we create an eigen-decomposition of $\widehat{S}_{t,r} = \widehat{U}_t \widehat{\Lambda}_t \widehat{U}_t^\top$ for each $t = 1, 2, \cdots, T$. As Lei and Rinaldo (2015) discuss, the matrix $\widehat{U}_t$ may now have more than K distinct rows due to the degree correction, whereas the rows of $\widehat{U}_t$ still only point to at most K directions. Therefore, we apply the spherical clustering algorithm to find a cluster structure among the rows of the normalized matrix $\widehat{U}_t^+$ with $\widehat{U}_t^+(i, *) = \widehat{U}_t(i, *)/\|\widehat{U}_t(i, *)\|$. More specifically, we consider the following spherical k-means spectral clustering:

$$\left\| \widehat{Z}_t^+ \widehat{Y}_t - \widehat{U}_t^+ \right\|_F^2 \leq (1 + \varepsilon) \min_{\substack{Z_t^+ \in \mathcal{M}_{N_+,K} \\ Y_t \in \mathbb{R}^{K \times K}}} \left\| Z_t^+ Y_t - \widehat{U}_t^+ \right\|_F^2 \tag{4.2.17}$$

where Y_t is some rotation matrix. In the last step, we extend $\widehat{Z}_t^+$ to obtain $\widehat{Z}_t$ by adding $N - N_+$ canonical unit row vectors at the end. $\widehat{Z}_t$ is the estimate of Z_t from this method. We summarize the algorithm in detail below.

Algorithm 1: CASC in the Dynamic SC-DCBM

Input : Adjacency matrices A_t for $t = 1, \cdots, T$;
Covariates matrix X;
Number of communities K;
Approximation parameter ε.

Output: Membership matrices Z_t for any $t = 1, \cdots, T$.

1 Calculate regularized graph Laplacian $L_{\tau,t}$ and weight matrix W_t.

2 Estimate $\mathcal{S}_t$ by $\widehat{\mathcal{S}}_{t,r}$ as in (4.2.16).

3 Let $\widehat{U}_t \in \mathbb{R}^{N \times K}$ be a matrix representing the first K eigenvectors of $\widehat{\mathcal{S}}_{t,r}$.

4 Let N_+ be the number of nonzero rows of $\widehat{U}_t$. Then, obtain $\widehat{U}^+ \in \mathbb{R}^{N_+ \times K}$ consisting of normalized nonzero rows of $\widehat{U}_t$; that is, $\widehat{U}_t^+(i,*) = \widehat{U}_t(i,*) / \left\|\widehat{U}_t(i,*)\right\|$ for i such that $\left\|\widehat{U}_t(i,*)\right\| > 0$.

5 Apply the $(1+\varepsilon)$-approximate k-means algorithm to the row vectors of $\widehat{U}_t^+$ to obtain $\widehat{Z}_t^+ \in \mathcal{M}_{N_+,K}$.

6 Extend $\widehat{Z}_t^+$ to obtain $\widehat{Z}_t$ by arbitrarily adding $N - N_+$ canonical unit row vectors at the end, such as $\widehat{Z}_t(i) = (1, 0, \cdots, 0)$ for i such that $\left\|\widehat{U}_t(i,*)\right\| = 0$.

7 Output $\widehat{Z}_t$.

Similar to the dynamic SC-DCBM case, we estimate the block structure of the dynamic SC-DCcBM by analyzing the normalized singular vectors on both sides. Then, using the spherical k-means analysis, we can also obtain the clustering matrices. The spectral clustering algorithm for the dynamic SC-DCcBM is below.

4.2.3 Uniform consistency

Undirected case

In the subsequent analysis, we illustrate that the dynamic CASC is uniformly consistent over time for both undirected and directed networks. We first make some assumptions on the graph that generates the dynamic network. The major assumption we need here is *assortativity*, which ensures that the nodes within the same cluster are more likely to share an edge than nodes in two different clusters.

Algorithm 2: CASC in the Dynamic SC-DCcBM

Input : Adjacency matrices A_t for $t = 1, \cdots, T$;
Covariates matrix X;
Number of row clusters K_R and number of column clusters K_C;
Approximation parameter ε.

Output: Membership matrices of rows and columns $Z_{R,t}$ and $Z_{C,t}$ for $t = 1, \cdots, T$.

1 Calculate regularized graph Laplacian $L_{\tau,t}$.

2 Estimate $\mathcal{S}_t$ by $\widehat{\mathcal{S}}_{t,r}$ as in (4.2.16).

3 Compute the singular value decomposition of $\widehat{\mathcal{S}}_{t,r} = U_t \Sigma_t V_t^\top$ for $t = 1, \cdots, T$.

4 Extract the first K columns of U_t and V_t that correspond to the K largest singular values in Σ_t, where $K = \min\{K_R, K_C\}$. Denote the resulting matrices $U_t^K \in \mathbb{R}^{N \times K}$ and $V_t^K \in \mathbb{R}^{N \times K}$.

5 Let N_+^R be the number of nonzero rows of U_t^K; then, obtain $U_{t+}^K \in \mathbb{R}^{N_+^R \times K}$ consisting of normalized nonzero rows of U_{t+}^K; that is, $U_{t+}^K(i,*) = U_t^K(i,*) / \left\| U_t^K(i,*) \right\|$ for i such that $\left\| U_t^K(i,*) \right\| > 0$.

6 Similarly, let N_+^C be the number of nonzero rows of V_t^K; then, obtain $V_{t+}^K \in \mathbb{R}^{N_+^C \times K}$ consisting of normalized nonzero rows of V_{t+}^K; that is, $V_{t+}^K(i,*) = V_t^K(i,*) / \left\| V_t^K(i,*) \right\|$ for i such that $\left\| V_t^K(i,*) \right\| > 0$.

7 Apply the $(1+\varepsilon)$-approximate k-means algorithm to cluster the rows (columns) of $\widehat{\mathcal{S}}_t$ into K_R (K_C) clusters by treating each row of U_{t+}^K (V_{t+}^K) as a point in $\mathbb{R}^K$ to obtain $\widehat{Z}_{R,t}^+$ ($\widehat{Z}_{C,t}^+$).

8 Extend $\widehat{Z}_{R,t}^+$ ($\widehat{Z}_{C,t}^+$) to obtain $\widehat{Z}_{R,t}$ ($\widehat{Z}_{C,t}$) by arbitrarily adding $N - N_+^R$ ($N - N_+^C$) canonical unit row vectors at the end, such as $\widehat{Z}_{R,t}(i) = (1, 0, \cdots, 0)$ ($\widehat{Z}_{C,t}(i) = (1, 0, \cdots, 0)$) for i such that $\|U_t(i,*)\| = 0$ ($\|V_t(i,*)\| = 0$).

9 Output $\widehat{Z}_{R,t}$ and $\widehat{Z}_{C,t}$.

Assumption 4.2.1 *1 The dynamic network is composed of a series of assortative graphs that are generated under the stochastic blockmodel with covariates whose block probability matrix B_t is positive definite for all $t = 1, \cdots, T$.*

Intuitively, the more frequent the group membership changes, the less stable the network will be. Consequently, it becomes harder to make use of the information from the historical and future network structures to detect the communities in the present network structure. In Assumption 4.2.2, we restrict the maximum number of nodes that switch memberships (s) to some finite number. Based on this assumption, the proportion of nodes that switch their memberships shrinks to 0 as the size of the network grows to infinity. Additionally, we can easily bound the dynamic behavior of clustering matrices ($Z_{t+r} - Z_t$) by noting that there are at most rs nonzero rows in the differenced matrix.

Assumption 4.2.2 *2 At most, $s < \infty$ number of nodes can switch their memberships between any consecutive time instances.*

Assumption 4.2.3 *3 For $1 \leq k \leq k' \leq K$, there exists a function $f(\cdot; k, k')$ such that $B_t(k, k') = f(\varsigma_t; k, k')$ and $f(\cdot; k, k') \in \Sigma(\beta, L)$, where $\Sigma(\beta, L)$ is a Hölder class of functions $f(\cdot)$ on $[0, 1]$ such that $f(\cdot)$ are ℓ times differentiable and*

$$|f^{(\ell)}(x) - f^{(\ell)}(x')| \leq L|x - x'|^{\beta-\ell}, \text{ for any } x, x' \in [0, 1], \tag{4.2.18}$$

with ℓ being the largest integer smaller than β.

Assumption 4.2.3 states that neither the connection probabilities nor the cluster memberships change drastically over the horizons. Lastly, to guarantee the performance of our clustering method, we impose some conditions to regularize the behavior of the covariate matrix and the eigenvalues of the similarity matrices.

Assumption 4.2.4 *4 Let $\lambda_{1,t} \geq \lambda_{2,t} \geq \cdots \geq \lambda_{K,t} > 0$ be the K largest eigenvalues of $\mathcal{S}_t$ for each $t = 1, \cdots, T$. In addition, assume that*

$$\underline{\delta} = \inf_t \{\min_i \mathcal{D}_{\tau,t}(i, i)\} > 3\log(8NT/\epsilon) \quad \textit{and} \quad \alpha_{\max} = \sup_t \alpha_t \leq \frac{a}{NRJ^2\xi},$$

with

$$a = \sqrt{\frac{3\log(8NT/\epsilon)}{\underline{\delta}}} \quad \textit{and} \quad \xi = \max(\sigma^2\|L_\tau\|_F\sqrt{\log(TR)}, \sigma^2\|L_\tau\|\log(TR), NRJ^2/\underline{\delta}),$$

where $\sigma = \max_{i,j}\|X_{ij} - \mathcal{X}_{ij}\|_{\phi_2}$, $L_\tau = \sup_t L_{\tau,t}$.

To establish the consistency of the CASC for the dynamic SC-DCBM, we need to determine the upper bounds for the misclustering rates. Following Binkiewicz et al. (2017), we denote $C_{i,t}$ and $\mathcal{C}_{i,t}$ as the cluster centroids of the ith node at time t generated using k-means clustering on the sample eigenvector U_t and the population $\mathcal{U}_t$, respectively. Then, we define the set of mis-clustered nodes at each period as

$$\mathbb{M}_t = \left\{i\colon \left\|C_{i,t}\mathcal{O}_t^\top - \mathcal{C}_{i,t}\right\| > \left\|C_{i,t}\mathcal{O}_t^\top - \mathcal{C}_{j,t}\right\|, \text{for any } j \neq i\right\}, \tag{4.2.19}$$

where $\mathcal{O}_t$ is a rotation matrix that minimizes $\|U_t\mathcal{O}_t^\top - \mathcal{U}_t\|_F$ for each $t = 1, \cdots, T$.

The misclustering error in $\mathbb{M}_t$ has two sources: the estimation error of $\mathcal{S}_t$ using the discrete kernel estimator and from spectral clustering. In Theorem 4.2.1, we provide the uniform upper bound of the misclustering rate for the undirected and directed networks separately.

Theorem 4.2.1 *Let clustering proceed according to Algorithm 1 based on the estimator* $\widehat{\mathcal{S}}_{t,r}$ *of* $\mathcal{S}_t$. *Let* $Z_t \in \mathcal{M}_{N,K}$ *and* $P_{\max} = \max_{i,t}(Z_t^\top Z_t)_{ii}$ *denote the size of the largest block over the horizons. Then, under Assumptions 4.2.1-4.2.4, the misclustering rate satisfies*

$$\sup_t \frac{|\mathbb{M}_t|}{N} \leq \frac{c_1(\varepsilon)KW_{\max}^2}{m_z^2 N\lambda_{K,\max}^2}\left\{(6+c_w)\frac{b}{\underline{\delta}^{1/2}} + \frac{2K}{\underline{\delta}}(\sqrt{2P_{\max}rs} + 2P_{\max}) + \frac{NL}{\underline{\delta}\cdot l!}\left(\frac{r}{T}\right)^\beta\right\}^2.$$

with a probability of at least $1-\epsilon$, *where* $c_1(\varepsilon) = 2^9(2+\varepsilon)^2$, $b = \sqrt{3\log(8NT/\epsilon)}$, *and* $\lambda_{K,\max} = \max_t\{\lambda_{K,t}\}$ *with* $\lambda_{K,t}$ *being the* K*th largest absolute eigenvalue of* $\mathcal{S}_t$.

Directed case

Analogous to the undirected case, we modify Assumption 4.2.4 to accommodate the stochastic co-blockmodel setup.

Assumption 4.2.5 *4' Let* $\lambda_{1,t} \geq \lambda_{2,t} \geq \cdots \geq \lambda_{K,t} > 0$ *be the* $K = \min K_R, K_C$ *largest singular values of* $\mathcal{S}_t$ *for each* $t = 1, \cdots, T$. *In addition, assume that*

$$\underline{\delta}' = \inf_t \{\min\{\min_i \mathcal{D}_{R,t}(i,i), \min_i \mathcal{D}_{C,t}(i,i)\}\} > 3\log(16NT/\epsilon)$$

and

$$\alpha_{\max} = \sup_t \alpha_t \leq \frac{a}{NRJ^2\xi},$$

with $a = \sqrt{\frac{3\log(16NT/\epsilon)}{\underline{\delta}'}}$ *and* $\xi = \max(\sigma^2\|L_\tau\|_F\sqrt{\log(TR)}, \sigma^2\|L_\tau\|\log(TR), NRJ^2/\underline{\delta}')$, *where* $\sigma = \max_{i,j}\|X_{ij} - \mathcal{X}_{ij}\|_{\phi_2}$, $L_\tau = \sup_t L_{\tau,t}$.

Following Rohe et al. (2016), we define the "R-mis-clustered" and "C-mis-clustered" vertices as

$$\mathbb{M}_t^p = \left\{i\text{: } \left\|C_{i,t}^p - \mathcal{C}_{i,t}^p\mathcal{O}_t^p\right\| > \left\|C_{i,t}^p - \mathcal{C}_{j,t}^p\mathcal{O}_t^p\right\|, \text{for any } j \neq i\right\}, \quad p \in \{R, C\}, \tag{4.2.20}$$

where $C_{i,t}^p$ and $\mathcal{C}_{i,t}^p$ for $p \in \{R, C\}$ are the cluster centroids of the ith node at time t generated using the k-means clustering on the left/right singular vectors and the population left/right singular vectors, respectively.

Theorem 4.2.2 *Assuming* $K_R \leq K_C$, *let* $Z_{R,t} \in \mathcal{M}_{N,K_R}$, $Z_{C,t} \in \mathcal{M}_{N,K_C}$, *and* $P_{\max} = \max\{\max_{i,t}(Z_{R,t}^\top Z_{R,t})_{ii}, \max_{i,t}(Z_{C,t}^\top Z_{C,t})_{ii}\}$ *denote the size of the largest block over the horizons. Then, under Assumptions 4.2.1-4.2.3 and 4.2.5, the mis-clustering rate satisfies*

$$\sup_t \frac{\left|\mathbb{M}_t^R\right|}{N} \leq \frac{c_2(\varepsilon)KW_{\max}^2}{m_r^2 N\lambda_{K,\max}^2}\left\{(6+c_w')\frac{b'}{\underline{\delta}'^{1/2}} + \frac{2K_C}{\underline{\delta}'}(\sqrt{2P_{\max}rs} + 2P_{\max}) + \frac{NL}{\underline{\delta}'\cdot\ell!}\left(\frac{r}{T}\right)^\beta\right\}^2,$$
$$\sup_t \frac{\left|\mathbb{M}_t^C\right|}{N} \leq \frac{c_3(\varepsilon)KW_{\max}^2}{m_c^2 N\gamma_c^2\lambda_{K,\max}^2}\left\{(6+c_w')\frac{b'}{\underline{\delta}'^{1/2}} + \frac{2K_C}{\underline{\delta}'}(\sqrt{2P_{\max}rs} + 2P_{\max}) + \frac{NL}{\underline{\delta}'\cdot\ell!}\left(\frac{r}{T}\right)^\beta\right\}^2,$$

with a probability of at least $1-\epsilon$*, where* $c_2(\varepsilon) = 2^6(2+\varepsilon)^2$*,* $c_3(\varepsilon) = 2^7(2+\varepsilon)^2$*,* $b' = \sqrt{3\log(16NT/\epsilon)}$*,* γ_c *are defined in supplement equation (44), and* $\lambda_{K,\max} = \max_t\{\lambda_{K,t}\}$ *with* $\lambda_{K,t}$ *being the* K*th largest absolute singular value of* $\mathcal{S}_t$*.*

4.2.4 Choice of tuning parameters

Obviously, we must choose the tuning parameters r, α, and K carefully. For the choice of r, we first need to determine the upper bound of the variance proportion of the estimation error $\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t\|$, which is $\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\|$. In the following lemma, we derive a sharp probabilistic upper bound on $\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\|$ using the device provided in Lei and Rinaldo (2015).

Lemma 4.2.1 *Let* $d = rN\|\mathcal{S}_t\|_\infty$ *and* $\eta \in (0,1)$*. Then,*

$$\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\| \leq (1-\eta)^{-2}\frac{W_{\max}\sqrt{d}}{r \vee 1},$$

with probability $1-\epsilon$*, where* $\epsilon = N^{\left(\frac{3}{16\|\mathcal{S}_t\|_\infty} - 2\log\left(\frac{7}{\eta}\right)\right)}$.

From Lemma 4.2.1 and the proofs of the previous theorems, we can see that $\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\|$ is decreasing, while $\|\mathcal{S}_{t,r} - \mathcal{S}_t\|$ is increasing in r. Therefore, there exists an optimal r^* that achieves the best bias-variance balance; that is,

$$r^* = \arg\min_{0\leq r\leq T/2}\left((1-\eta)^{-2}\frac{W_{\max}\sqrt{d}}{r\vee 1} + \|\mathcal{S}_{t,r} - \mathcal{S}_t\|\right). \tag{4.2.21}$$

Then, we can apply Lepski's method (Lepski et al., 1997) to construct the adaptive estimator for r^*. Without loss of generality, we choose $\eta = 1/2$. The, we define the adaptive estimator as

$$\widehat{r} = \max\left\{0 \leq r \leq T/2 : \left\|\widehat{\mathcal{S}}_{t,r} - \widehat{\mathcal{S}}_{t,\rho}\right\| \leq 4W_{\max}\sqrt{\frac{N\|\mathcal{S}_t\|_\infty}{\rho\vee 1}},\ \text{for any } \rho < r\right\}. \tag{4.2.22}$$

Next, for the choice of α_t, we select α_t to achieve a balance between $L_{\tau,t}$ and

C_t^w:

$$\alpha_t = \frac{\lambda_K(L_{\tau,t}) - \lambda_{K+1}(L_{\tau,t})}{\lambda_1(C_t^w)}. \tag{4.2.23}$$

Lastly, to determine K, we have several choices. Wang and Bickel (2017) implement a pseudo likelihood approach to choose the number of clusters in a stochastic blockmodel without covariates. Chen and Lei (2017) propose a network cross-validation procedure to estimate the number of clusters by utilizing adjacency information. Li et al. (2016) refine the network cross-validation approach by proposing an edge sampling algorithm. In our case, we apply the network cross-validation approach directly by inputting the similarity matrix instead of the adjacency matrix because the covariate matrix C_t^w behaves just like an adjacency matrix when we use dummy variables to indicate different technology attributes. Therefore, the network cross-validation applies to the similarity matrix in our study.

4.2.5 Monte Carlo simulations

In this section, we carry out several simulation studies using our algorithm and existing clustering methods under different model setups. Our benchmark algorithms for undirected networks are the dynamic degree-corrected spectral clustering for the sum of the squared adjacency matrix (DSC-DC) by Bhattacharyya and Chatterjee (2018) and the dynamic spectral clustering method (DSC-PZ) by Pensky and Zhang (2017). For the directed networks, as we do not have a fair competitor for a dynamic model, we choose several algorithms designed for a static model. In particular, we compete with the degree-corrected DI-SIM (DI-SIM-DC) by Rohe et al. (2016) and the covariate-assisted DI-SIM (CA-DI-SIM-St) method by Zhang et al. (2018) for the adjacency matrix in each period.

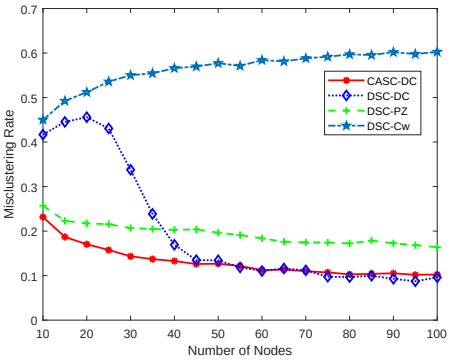
First, we set the block probability matrix B_t as

$$B_t = \frac{t}{T}\begin{bmatrix} 0.9 & 0.6 & 0.3 \\ 0.6 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.8 \end{bmatrix}, \text{ with } 1 \leq t \leq T.$$

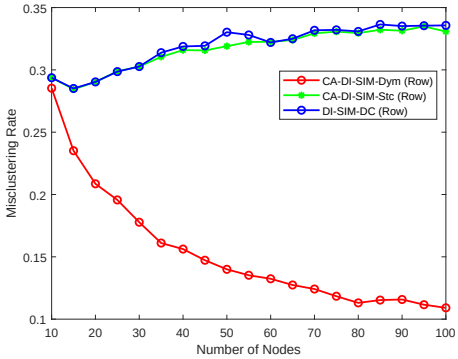
and set the order of the polynomials for kernel construction at $L = 4$ for all simulations. In the next step, for the undirected network, we simulate the first period's clustering matrix Z_1 by randomly choosing one entry in each row and assign it to 1 to generate clustering matrices (Z_t). Then, for $t = 2, \cdots, T$, we fix the last $N - s$ rows of Z_{t-1} and re-assign 1s in the first s rows of Z_1 to mimic the group membership change behaviors. Similarly, for the directed network, we generate each period's row/column clustering matrix ($Z_{R,t}$ or $Z_{C,t}$) in the same way, separately. Lastly, we assume that the number of communities $K = 3$ (or K_R and K_C for directed network) is known throughout the simulations. The time-invariant node covariates are $R = \lfloor \log(N) \rfloor$ dimensional with values $X \sim U(0, 10)$. We replicate all experiments 100 times and the misclustering rate we report is the temporal average of the misclustering rates; that is, $T^{-1} \sum_{t=1}^{T} |\mathbb{M}_t|/N$ (or $T^{-1} \sum_{t=1}^{T} |\mathbb{M}_t^R|/N$ and $T^{-1} \sum_{t=1}^{T} |\mathbb{M}_t^C|/N$ for the directed network).

We first examine the clustering performance with a growing network size. The number of vertices in the network varies from 10 to 100 with step size 5. The time span is $T = 10$. We summarize the results in Figure 4.1. Evidently, as the size of the undirected network becomes larger (panel (a)), the misclustering rates of the CASC-DC decrease sharply and dominate DSC-PZ in all cases. DSC-DC only performs as well as CASC-DC when the network is large, while CASC-DC retains an acceptable misclustering rate in small networks. It also shows that although using the covariate per se for clustering (DSC-Cw) is unsatisfactory, we can still add covariates to the adjacency matrix for better grouping.

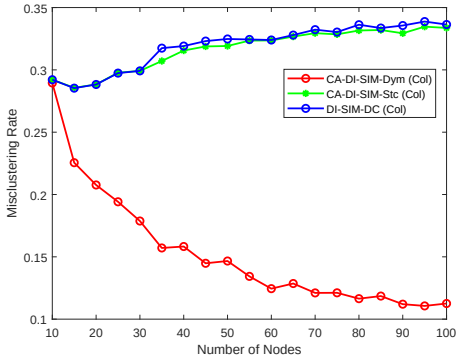
Next, we check the relative performance for a growing maximal number of group membership changes. Here, we fix the total number of vertices at 100 and we vary the group membership changes for each period, s, in $\{0, N/50, N/25, N/20, N/10, N/5, N/4, N/$ The total number of horizons is $T = 10$. We summarize the results in Figure 4.2. Obviously, our methods are sensitive to the total number of group membership changes. In other words, the more unstable the group membership is, the higher the misclustering rate will be. Despite the result, our method still achieves the lowest



(a) Undirected Network



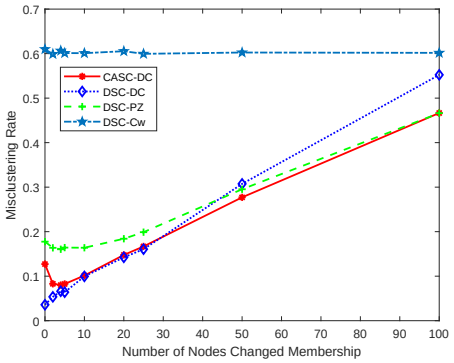
(b) Directed Network (Row Cluster)



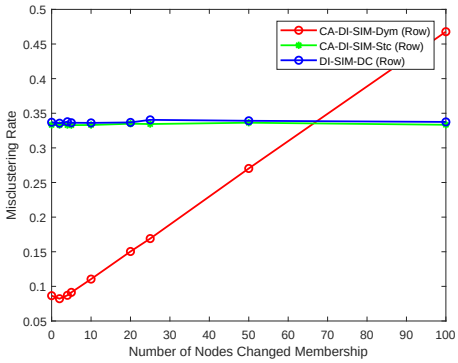
(c) Directed Network (Column Cluster)

Figure 4.1: This figure reports the misclustering rate of different spectral clustering algorithms for networks with a growing number of vertices. Panel (a) reports the results for undirected networks, while Panels (b) and (c) report the results for directed networks. CASC-DC represents Algorithm 1. DSC-DC denotes the dynamic spectral clustering in Bhattacharyya and Chatterjee (2018). DSC-PZ denotes the dynamic spectral clustering methods in Pensky and Zhang (2017). DSC-Cw is the spectral clustering based on only covariates. CA-DI-SIM-Dym represents Algorithm 2. DI-SIM-DC is the degree-corrected DI-SIM in Rohe et al. (2016) and CA-DI-SIM-Stc is the static covariate-assisted DI-SIM method in Zhang et al. (2018). In all cases, the number of nodes varies from 10 to 100, and the number of membership changes is fixed at $s = N^{1/2}$. The horizon $T = 10$ and all simulations are repeated 100 times.

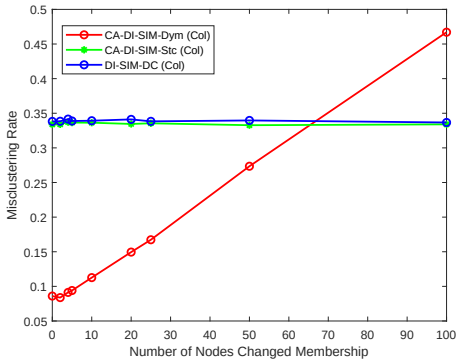
misclustering rate amongst all methods when the group memberships are relatively stable ($s \leq N/2$).



(a) Undirected Network



(b) Directed Network (Row Cluster)



(c) Directed Network (Column Cluster)

Figure 4.2: This figure reports the misclustering rate of different spectral clustering algorithms for networks with a growing number of membership changes. Panel (a) reports the results for undirected networks, while Panels (b) and (c) report the results directed networks. CASC-DC represents Algorithm 1. DSC-DC denotes the dynamic spectral clustering in Bhattacharyya and Chatterjee (2018). DSC-PZ denotes the dynamic spectral clustering methods in Pensky and Zhang (2017). DSC-Cw is the spectral clustering based on only covariates. CA-DI-SIM-Dym represents Algorithm 2. DI-SIM-DC is the degree-corrected DI-SIM in Rohe et al. (2016) and CA-DI-SIM-Stc is the static covariate-assisted DI-SIM method in Zhang et al. (2018). In all cases, the network size is fixed at 100, and the number of membership changes varies in $\{0, N/50, N/25, N/20, N/10, N/5, N/4, N/2, N\}$. The horizon is $T = 10$ and all simulations are repeated 100 times.

4.3 Crypto Networks and Clusters

In this section, we illustrate how we construct a dynamic network structure using crypto returns and its contract information. Specifically, we first form a return-based network using the inter-predictive relations between cryptos. In addition, we

add linkages between the cryptos that adopt similar cryptography techniques. We then perform clustering with our new algorithm.

4.3.1 Data and variables

We collected data on the historical daily prices, trading volumes, and contract attributes of the top 200 cryptos by market capitalization from an interactive platform (*Cryptocompare.com*) with free API access. After excluding cryptos with incomplete contract information, we obtain a sample of 199 cryptos. The sample covers August 31, 2015 to March 31, 2018, and we used an in-sample period for community detection from August 31, 2015 to December 31, 2017and an out-of-sample period of three months (2018-01-01 to March 31, 2018) for return predictability tests and portfolio construction. In term of the time-invariant attributes, we mainly collected algorithm and proof types from each crypto's contract:

Algorithm, which is short for the *hashing algorithm*, plays a central role in determining the security of the crypto. For each crypto, there is a hash function in mining; for example, Bitcoin (BTC) uses double SHA-256 and Litecoin (LTC) uses Scrypt. As security is one of the most important features of cryptos, the hashing algorithm naturally–in terms of trust–determines the intrinsic value of a crypto. In the example above, the Scrypt system was used with cryptos to improve upon the SHA256 protocol. The SHA256 preceded the Scrypt system and was the basis for BTC. Specifically, Scrypt was employed as a solution to prevent specialized hardware from brute-force efforts to out-mine others. Thus, Scrypt-based Altcoins require more computing effort per unit, on average, than the equivalent coin using SHA256. The relative difficulty of the algorithm confers a relative value.

Proof Types, or proof system/protocol, is an economic measure to deter denial of service attacks and other service abuses such as spam on a network by requiring some work from the service requester, usually the equivalent to processing time by a computer. For each crypto, at least one of the protocols will be chosen as a transaction verification method; for example, BTC and Ethereum (ETH) currently use the

Proof-of-Work (PoW), and Diamond (DMD) and Blackcoin use the Proof-of-Stake (PoS). PoW-based cryptos such as BTC use mining–the solving of computationally intensive puzzles–to validate transactions and create new blocks. In PoS-based cryptos, the creator of the next block is chosen through various combinations of random selection and wealth (in terms of crypto) or age (i.e., the stake). In summary, the proof protocol determines the reliability, security, and effectiveness of the transactions.

4.3.2 Crypto network construction

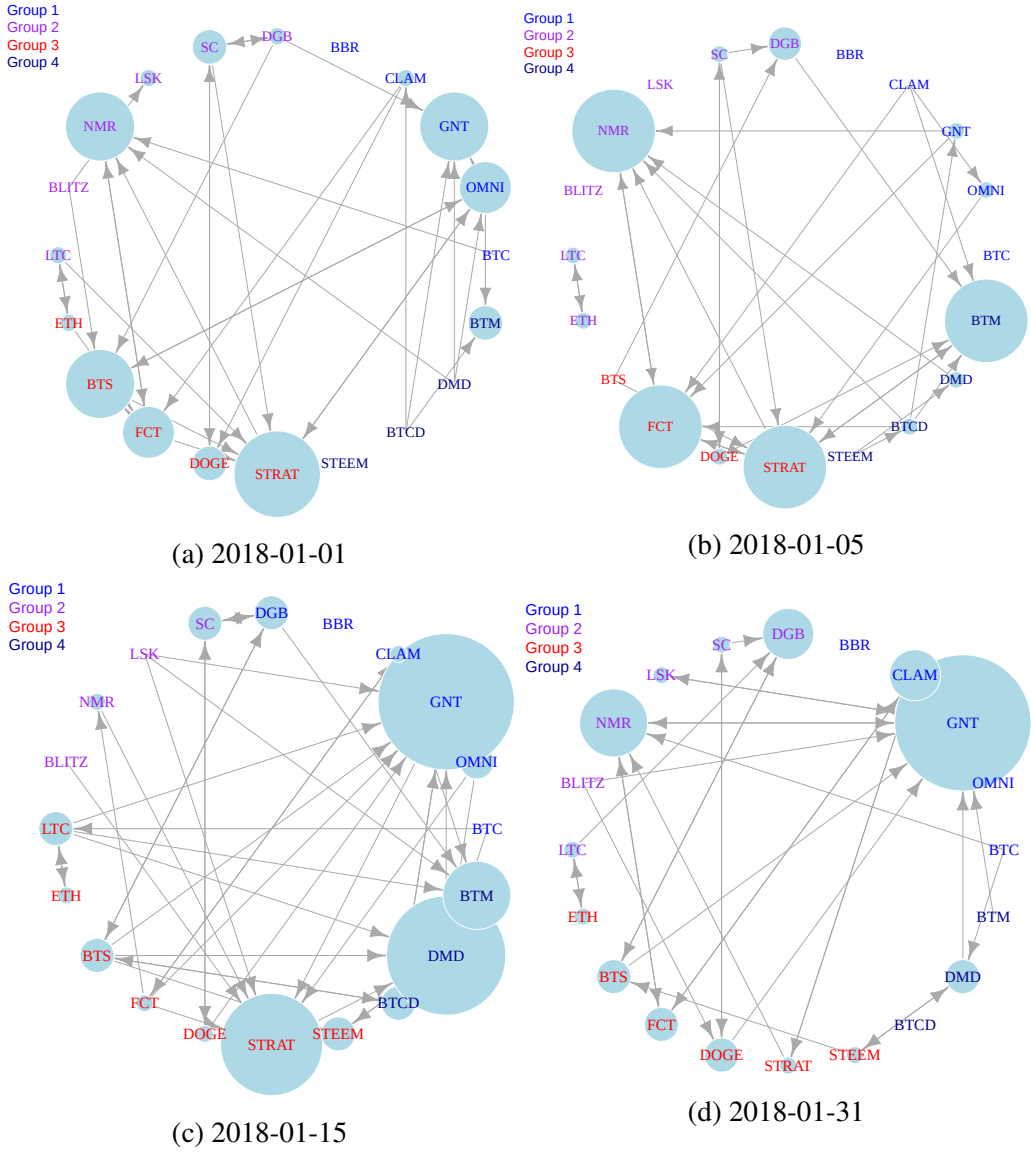
To study how risk or information propagates through the network, we construct it from the interrelations between the crypto returns. More precisely, we focus on one crypto and regresses its returns on the other cryptos' lagged returns in a 60-day estimation window. We employ adaptive LASSO (Zou, 2006) to estimate the regression coefficient; that is,

$$\hat{b}_i^* = \arg\min \left\{ \left\| r_{i,t+1}^s - \alpha_i - \sum_{j \neq i} b_{i,j} r_{j,t}^s \right\|^2 + \lambda_i \sum_{j \neq i} \hat{w_{i,j}} |b_{i,j}| \right\}, \tag{4.3.1}$$

where $r_{j,t}^s$ is the standardized return for crypto j, $\hat{b}_i^* = (\hat{b}_{i,1}^*, \cdots, \hat{b}_{i,N}^*)^\top$ is the adaptive LASSO estimate, λ_i are non-negative regularization parameters, and $\hat{w}_{i,j}$ are the weights corresponding to $|b_{i,j}|$ for $j = 1, \cdots, N$ in the penalty term. Conventionally, one defines $\hat{w}_{i,j} = 1/|\hat{b}_{i,j}^{ols}|^\gamma$ with some $\gamma > 0$. The LASSO technique yields an active set that has "parental" influence on the focal crypto. Thus, we obtain an adjacency matrix for each period, $\mathcal{A}_t$, $t = 1, \cdots, T$.

In Figure 4.3, we visualize a subgroup of 20 cryptos on selected dates to illustrate the structural features revealed by (4.3.1). The node color indicates the estimated group membership and the node size denotes its degree centrality from the receiver's perspective. Evidently, the predictive relations between cryptos are highly asymmetrical (rare double-sided arrows). Acemoglu et al. (2012) also observe this feature, which will later help us argue that sectoral shocks might not cancel out

through diversification, but aggregate into a systematic fluctuation. Therefore, determining the centered cryptos and the group structure is crucial for understanding how information or shocks propagate in the crypto market.



(a) 2018-01-01

(b) 2018-01-05

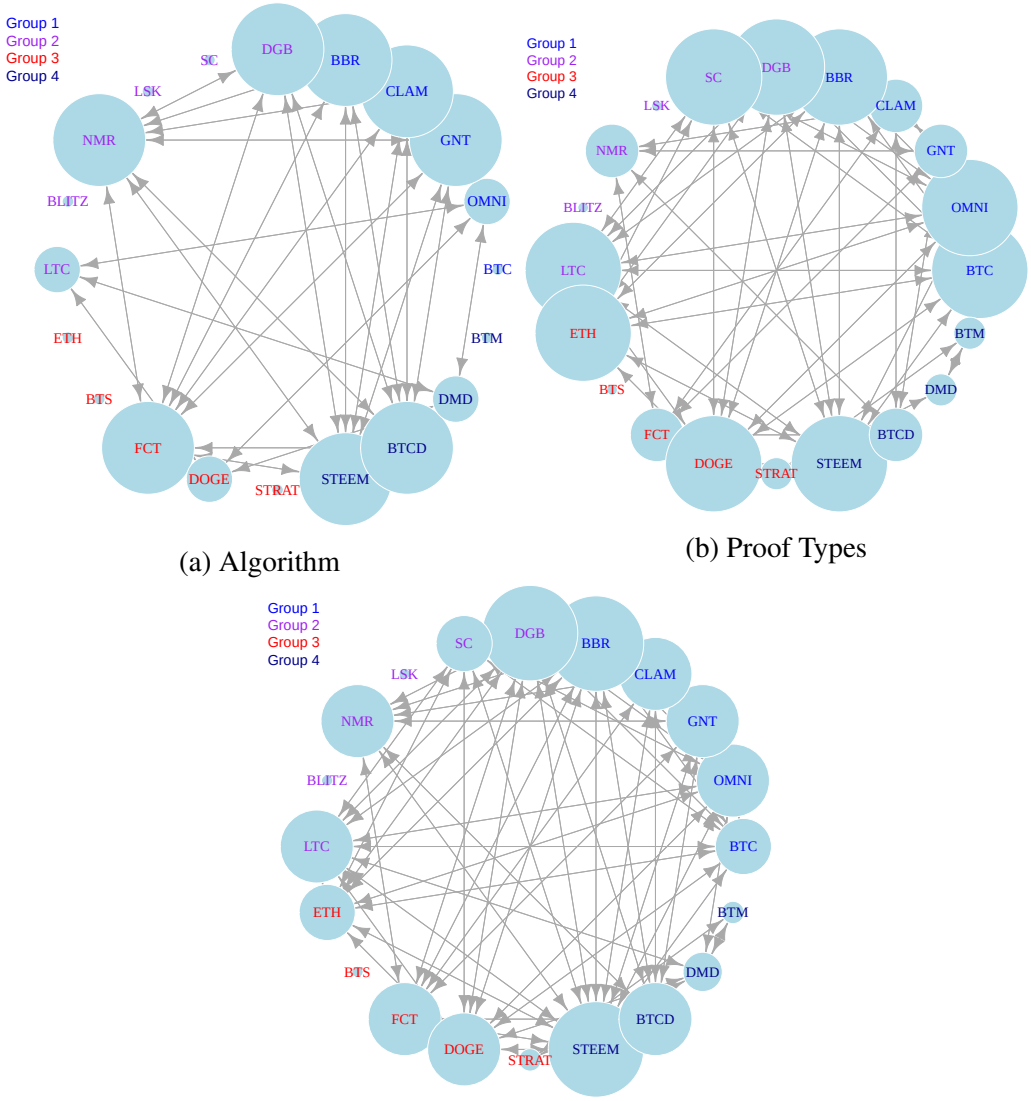
(c) 2018-01-15

(d) 2018-01-31

Figure 4.3: This figure presents the return-based network structure on selected dates in January 2018. We selected 20 cryptos, including BTC, ETH, LTC and other top cryptos by market capitalization as of December 31, 2017 within each group estimated by dynamic CASC. We obtained the connections from the predictive regression $r^s_{i,t+1} = \alpha_i + \sum_{j=1,j\neq i}^{N-1} b_{i,j} r^s_{j,t} + \epsilon_{i,t}$, where $r^s_{i,t}$ is the standardized daily return on crypto i and N is the total number of cryptos. Adaptive LASSO is employed to estimate the regression above and only the cryptos selected by adaptive LASSO will be linked to crypto i.

As Figure 4.3 shows, the return-inferred network is time-varying and sparse in general. Taking subfigures (a) and (d) as an example, the interrelation between

BTC and DMD vanishes on January 1, 2018, and the connections on 2018-01-01 are sparser than those on January 15, 2018 are. This observation requires a more refined clustering and the use of node attributions. To demonstrate how node attribution assists classification, we replot the network with the same cryptos in Figure 4.3 and link the cryptos that share at least one fundamental characteristic to obtain Figure 4.4. Both LTC and DOGE adopt the Scrypt algorithm; hence, these two cryptos are fundamentally connected.



(a) Algorithm

(b) Proof Types

(c) Combined Fundamentals

Figure 4.4: This figure depicts the contract-based network structure. We link two cryptos if they share the same fundamental technology, that is, algorithm and proof types. Node size denotes the degree centrality of the crypto.

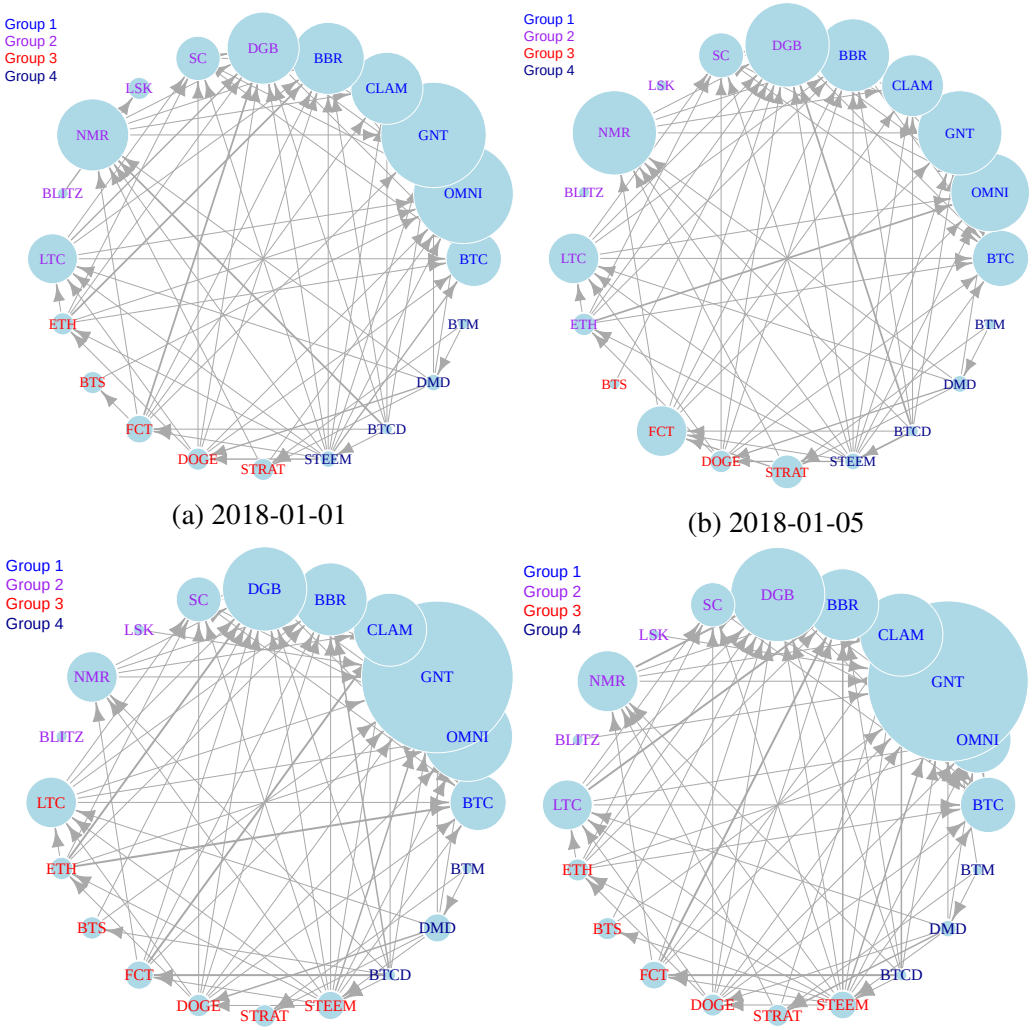
Clearly, due to the limited choices of algorithms and other attributes, the cryp-

tos are more likely to connect with each other when using attribute commonality to form linkages. However, using contract information alone is enough to identify the group structure, as crypto returns carry information on investors' beliefs, which is particularly important for the crypto market. In addition, the relationship between a crypto's fundamental characteristics to its value is more complicated than is a firm's fundamental to its equity. It is possible that a new algorithm does not add any valuable features to the existing algorithms. In fact, many developers simply copy and paste the blockchain source code with minor modifications on the parameters to launch a new coin for speculative purposes through an initial coin offering (ICO). Although these Altcoins may show little differences between their fundamental characteristics, their abilities to generate future cash flows vary considerably. A good example is IXCoin, the first BTC *clonecoin*. While IXCoin copied every detail from Bitcoin, IXCoin was unable to replicate the success of BTC. The developers stopped working on IXCoin for months after its ICO. This example shows that a clonecoin could be more risky than its protocoin for speculation reasons. www.deadcoins.com provides other similar cases.

To address the issues raised above fully, we combine the return-based network and the contract-based network using a similarity matrix. Figure 4.5 illustrates the combined network for selected dates. Compared to the network based on a single information set, the combined network is denser and the degrees of the cryptos are distributed more evenly. Consequently, the similarity matrix will most likely improve classification accuracy.

4.3.3 Clusters in crypto networks

The combined network structure and application of the CASC created four groups. Table 4.2 summarizes the grouping results for one example. The table indicates that as of December 31, 2017, the largest cryptos (BTC, ETH, and LTC) in terms of market capitalization are not necessarily categorized into the same group. Take LTC and BTC as an example. Although their return patterns are closely related, the



(a) 2018-01-01

(b) 2018-01-05

(c) 2018-01-15

(d) 2018-01-31

Figure 4.5: This figure depicts the dynamic combined network structure based on a similarity matrix, which combines return information and contract information simultaneously. The color of the node labels indicates the group estimated by dynamic CASC and the node size denotes the degree centrality of the crypto.

fundamental attributes between them are rather different: BTC employs SHA256 while LTC uses Scrypt. As a comparison, we also show the grouping results for the same 20 cryptos under DISIM from Rohe et al. (2016) in Table 4.2.

Table 4.1: Representative Cryptos of Groups Estimated by the Dynamic CASC.

Group ID	Group 1	Group 2	Group 3	Group 4
Cryptocurrencies	BBR	BLITZ	BTS	BTCD
	BTC	DGB	DOGE	BTM
	CLAM	LSK	*ETH*	DMD
	GNT	NMR	FCT	STEEM
	OMNI	SC	*LTC*	STRAT

Table 4.2: Representative Cryptos of Groups Estimated by DISIM from Rohe et al. (2016).

Group ID	Group 1	Group 2	Group 3	Group 4
Cryptocurrencies	BBR	BTC	BLITZ	BTCD
	LSK	DGB	STEEM	CLAM
	DOGE	LTC	SC	GNT
	ETH	NMR	BTS	
	OMNI	DMD		
	BTM	STRAT		

To illustrate the performance of our method, we check the differences between the within- and cross-group connections of each group, defined as

$$\textit{Within-Group Connection}_i = \frac{\text{\# of Degrees of Coins within Group } i}{4N_i},$$
$$\textit{Cross-Group Connection}_i = \frac{\text{\# of Degrees of Coins between Group } i \text{ and other Groups}}{4\bar{N}_i},$$

where N_i is the number of cryptos in group i and $\bar{N}_i$ is the number of cryptos not in group i. Intuitively, if the clustering method correctly classifies all cryptos, then the within-group connections should be stronger than the cross-group connections; that is, the difference between them should be positive. Table 4.3 summarizes the within- and cross-group connections of different information sets based on DISIM from Rohe et al. (2016) and dynamic CASC, respectively. Panel A reports the average return-based connection over the sample period. Panels B and C report the

algorithm-inferred connections and proof-types-inferred connections, respectively. The differences between the within- and cross-group connection (W-C difference) are reported in the last column of each panel.

Table 4.3: Within- and Cross-group Connections using DISIM and Dynamic CASC

Panel A reports the average return-based connection across the sample period. Panels B and C report the algorithm-inferred connections and proof-type-inferred connections, respectively. Statistical significance indicated by 1% 5% 10% for the positive signs and 1% 5% 10% for the negative signs.

	Panel A: Return			Panel B: Algorithm			Panel C: Proof Types		
	Within	Cross	Diff.	Within	Cross	Diff.	Within	Cross	Diff.
DISIM by Rohe et al. (2016)									
Group 1	0.033	0.051	−0.018	0.252	0.204	0.048	0.251	0.229	0.021
Group 2	0.084	0.074	0.010	0.216	0.198	0.018	0.277	0.238	0.039
Group 3	0.086	0.075	0.011	0.215	0.196	0.019	0.279	0.238	0.041
Group 4	0.084	0.073	0.011	0.216	0.197	0.019	0.278	0.238	0.040
Overall	0.072	0.068	0.004	0.225	0.199	0.026	0.271	0.236	0.035
Dynamic CASC									
Group 1	0.029	0.026	0.003	0.232	0.202	0.030	0.266	0.232	0.033
Group 2	0.029	0.025	0.004	0.243	0.203	0.041	0.272	0.232	0.040
Group 3	0.031	0.025	0.005	0.240	0.202	0.038	0.274	0.233	0.041
Group 4	0.031	0.026	0.006	0.240	0.202	0.038	0.273	0.233	0.040
Overall	0.030	0.025	0.005	0.239	0.202	0.037	0.271	0.233	0.039

Evidently, the dynamic CASC method has superior classification efficiency than DISIM does given that it delivers higher overall differences in both return-inferred connections and contract-inferred connections. For example, the overall W-C difference of DISIM is 0.004, 0.026, and 0.035, while that of dynamic CASC is 0.005, 0.037, and 0.039, respectively. Indeed, dynamic CASC utilizes fundamental information better in the sense that the contract-inferred network structure (Panels B and C) generates a higher W-C difference without discounting the grouping information from the return-inferred network. These facts indicate that fundamental information introduces an extra dimension of commonality for classifying cryptos, and improves the information extraction from return dynamics by emphasizing the return co-movement induced by fundamental commonality.

4.4 Asset Pricing Inference

In this section, we apply the classifications we obtained to asset pricing. We first study whether the group structure achieves good risk diversification. Then, we sort the cryptos into 4 quartiles according to eigenvector centrality and construct a portfolio that goes long on the high-centrality cryptos and short on the low-centrality cryptos. Lastly, we conduct several robustness tests to exclude alternative explanations of the centrality measure.

4.4.1 Risk diversification

Risk diversification is one of the most important issues in portfolio management. Portfolio managers seek to achieve a target return with the smallest variance possible. Therefore, it is crucial to invest in different assets or equity sectors that are not highly correlated with each other. We calculate the correlation coefficients of cryptos within the same group and those of the cryptos across groups. Table 4.4 summarizes the results.

Table 4.4: Within- and Cross-group Cryptos' Average Return Correlations by Dynamic CASC.

This table reports the within- and cross-group average return correlation based on dynamic CASC. Each trading day, we balance the portfolio according to the clustering results and calculate the within- and cross-group correlations. The number in brackets below are the t-statistics, which are adjusted by the Newey-West lags(4) method. Statistical significance is indicated by 1% 5% 10% for the positive signs and 1% 5% 10% for the negative signs. The sample period spans from August 31, 2015 to March 31, 2018.

	Within Group	Cross Group	Diff.
Group 1	0.169	0.154	0.014
	(7.626)	(7.423)	(6.856)
Group 2	0.179	0.154	0.021
	(8.077)	(7.423)	(6.077)
Group 3	0.181	0.157	0.021
	(8.191)	(7.506)	(10.374)
Group 4	0.188	0.157	0.027
	(8.114)	(7.416)	(5.607)
Overall	0.188	0.157	0.021
	(7.697)	(7.381)	(6.331)

In Table 4.4, we compare the average pair-wise correlations between two groups. For the within-group portfolio, we randomly pick 10 cryptos from the same group, and for the cross-group portfolio, we randomly pick 5 cryptos in one group and pick the remaining 5 cryptos from other groups. Then, for each trading day, we balance the portfolio according to the clustering results and calculate the within- and cross-group correlations. Table 4.4 demonstrates that the correlations between cryptos within the same group are on average significantly higher than those across groups are. Indeed, the average correlation coefficient within a group is 0.18, while it is 0.15 across groups. In economic terms, this result indicates a 17% reduction in return co-movement when investing in cross-group cryptos. The difference is statistically significant at the 1% level with a Newey-West adjusted t-statistic of 6.33. The result suggests that investment practitioners can find attractive upside and diversification possibility through allocating portfolio weights on cryptos from different groups. As buying all cryptos is costly, the findings provide portfolio managers the opportunity to select group representatives with a significant diversification effect.

4.4.2 Centrality and crypto return

One major advantage of jointly modelling cryptos with a dynamic network is its convenience for studying how risk and trading information propagates from one crypto to another. Acemoglu et al. (2012) propose a theoretical model to explain the spillover effects through sector-level shocks. The model suggests that if the linkages in the inter-sectoral network are sufficiently asymmetric, then sectoral shocks might not cancel out through diversification, but aggregate into macroeconomic fluctuations. Ahern (2013) also finds that idiosyncratic shocks could travel between linked stocks following the direction of the linkages. Therefore, stocks with more "receive linkages" tend to bear more risks in the network and thus require a higher risk premium. Similarly, we would expect that cryptos in a more central position in the network require a higher risk premium.

Centrality, as the key measure describing the importance of the nodes in the

network, best proxies the concentration of risks or trading information. There are several measures of centrality, such as degree, closeness, betweenness, and eigenvector centrality. Among them, eigenvector centrality is the most appropriate measure for an asymmetric network for two reasons. First, shocks that transmit across the crypto market do not have final recipients and are unlikely to follow the shortest path between nodes. Therefore, we cannot use closeness and betweenness centrality to describe market shocks as they implicitly assume that traffic follows geodesic paths (Borgatti, 2005). Second, cross-asset shocks are likely to have feedback effects evidenced by the two-way connections between paired cryptos in Figure 4.3. Thus, using degree centrality tends to overestimate the importance of cryptos with more asymmetric linkages. Eigenvector centrality is calculated via the principal eigenvector of the network's adjacency matrix (Bonacich, 1972). Nodes are more central if they are connected to other nodes that are themselves more central. Figure 4.6 plots the average return of each group portfolio, labelled as high-, median- (2 groups in the middle), and low-centrality groups. Based on the thoughts on portfolio performance above, we find that the group with a higher centrality wins the horse race.

Next, we formally test this discovery by studying cross-sectional portfolio returns. We first sort cryptos into quartile portfolios based on the eigenvector centrality calculated from the similarity matrix on each trading day. We then look at each portfolio's average future returns. Next, we test the statistical significance of the difference in average future return between the high and low portfolios. To show the informativeness of our centrality measure, we construct the portfolio for several formation periods, ranging from day $t+1$ to $t+7$ days. Table 4.5 reports the results.

In line with the observations from Figure 4.6, the cryptos with a higher quartile of centrality receive a higher portfolio return. Particularly, the average portfolio return is 39.78 bps for the highest-centrality group, while it is -0.01 bps for the lowest-centrality group. The difference is statistically significant at the 1% level.

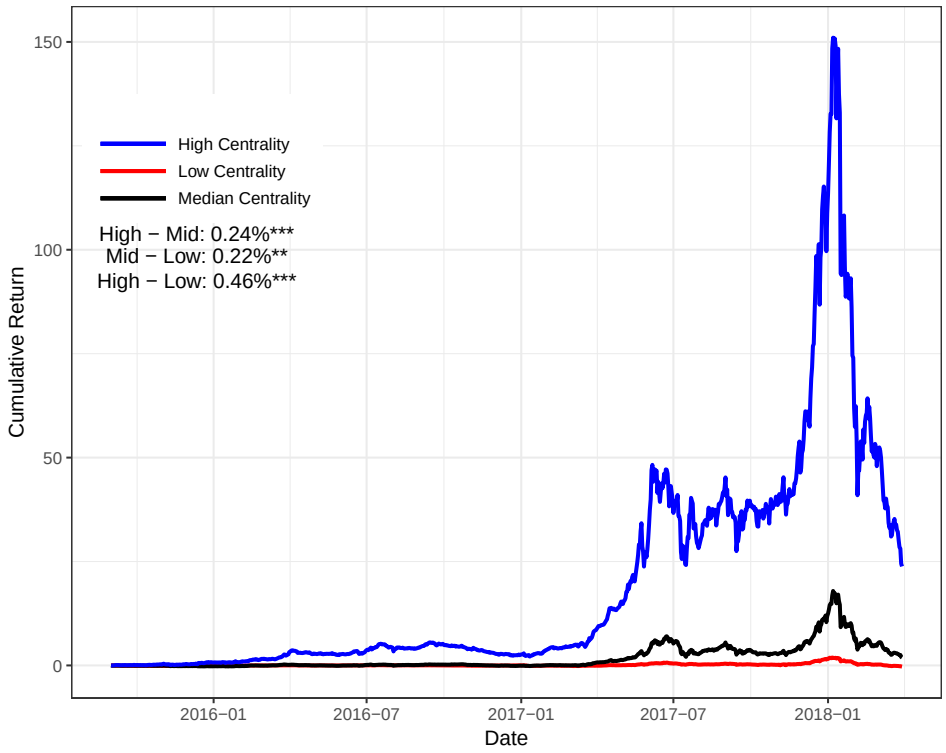


Figure 4.6: This figure depicts the cumulative portfolio return of the high-, median-, and low-centrality groups. Centrality is based on the similarity matrix, which combines return information and contract information simultaneously. The sample period spans from August 31, 2015 to March 31, 2018.

Table 4.5: Average Future Returns of the Cross-sectional Portfolios by Centrality Sorting.

This table reports the average future return for quartile portfolios sorted by the centrality measure. Each trading day, we balance the portfolio according to the centrality score of the previous trading day and calculate the average portfolio returns for both short and long legs. Statistical significance is indicated by ■ 1% ■ 5% ■ 10% for the positive signs and ■ 1% ■ 5% ■ 10% for the negative signs. The t-statistics in parentheses are computed based on standard errors with a Newey-West lags(4) adjustment. The sample period spans from August 31, 2015 to March 31, 2018.

Centrality	Ret_{t+1}	Ret_{t+2}	Ret_{t+3}	Ret_{t+4}	Ret_{t+5}	Ret_{t+6}	Ret_{t+7}
Low	0.00%	0.00%	-0.03%	-0.01%	0.03%	0.02%	0.06%
2	0.15%	0.18%	0.18%	0.19%	0.16%	0.18%	0.16%
3	0.34%	0.34%	0.28%	0.36%	0.38%	0.28%	0.29%
High	0.40%	0.36%	0.48%	0.38%	0.34%	0.42%	0.39%
High - Low	0.40%	0.36%	0.51%	0.39%	0.32%	0.40%	0.33%
t-statistic	(3.53)	(3.10)	(4.24)	(3.33)	(2.74)	(3.44)	(2.85)

We find similar results across different portfolio formation periods. The result provides strong evidence that an informational channel, such as risk and liquidity, should be applied to interpret the eigenvector centrality measure.

4.4.3 Alternative Interpretation

We showed that the centrality measure is economically meaningful as a risk factor. However, it does not rule out other explanations. We therefore conduct several tests to seek other possibilities to link the centrality measure to economic theory. In particular, we test if limit-to-arbitrage, investor attention, and macroeconomic uncertainty can deliver meaningful explanatory power of the anomaly.

The first typical explanation for asset return anomaly is the limit-to-arbitrage. According to Shleifer and Vishny (1997), sophisticated investors would quickly eliminate any return predictability arising from anomalies in a liquid market without impediments to arbitrage. Therefore, when cryptos are illiquid, an arbitrage opportunity is more likely to exist between central and non-central cryptos. As a formal test, we proxy liquidity with trading volume and first sort the cryptos into two groups (high and low) according to their previous day's trading volume. Then, for each group, we sort cryptos by their eigenvector centrality as in the previous sections, and report the corresponding portfolio returns in the first two columns of Table 4.6.

We find that the centrality portfolio return (High–Low) remains significantly positive for both high- and low-volume cryptos. For example, in the low-volume group, the portfolio return is 5 bps for the low-centrality group, while it increases to 28 bps for the high-centrality group. The significantly positive portfolio returns in both groups indicate that the limit-to-arbitrage does not fully explain the centrality measure.

The recent study of Liu and Tsyvinski (2018) provides an alternative explanation. The authors find that investor attention is a powerful predictor of crypto returns. Barber and Odean (2008) point out that excess attention usually drives in-

vestors to overreact to information and thus causes mispricing. Guo et al. (2018) show that investor attention could spill over along the network linkages. Hence, cryptos in a high-investor-attention period are more likely to be mispriced. Following Liu and Tsyvinski (2018), we proxy investor attention by constructing the deviation of Google searches for the word "crypto" on a given day compared to the average of those in the preceding four weeks. We split the sample into two periods (high and low) and test for the existence of the anomaly in each period. We summarize the results in the middle columns of Table 4.6.

In general, the proposed centrality measure—under both high- and low-attention periods—is a better choice. The effect seems to be stronger in high-attention periods. For example, the centrality portfolio achieves a 0.45% daily return during a high-attention period, while it retains a 0.35% return, if not higher, for the low-attention period. However, we can observe that the results are not fully explained by investor attention, as our centrality measure shows significant cross-sectional return predictability.

Last, observing that government policy and crypto price movement has a strong synchronization (Demir et al., 2018), we must check whether the centrality measure relates to underlying economic uncertainty. Naturally, when macroeconomic conditions become uncertain, investing in a certain asset is more risky and investors will require a higher risk premium (Brogaard and Detzel, 2015). We employ Baker et al. (2016) policy uncertainty index, which is constructed from three types of underlying components: media news; the Congressional Budget Office (CBO), which compiles lists of temporary federal tax code provisions; and the Federal Reserve Bank of Philadelphia's Survey of Professional Forecasters. Similarly, we divide the sample into two parts, high- and low-uncertainty periods, and test the existence of abnormal returns in each period. The last two columns of Table 4.6 report the results.

Evidently, the centrality portfolio return remains significantly positive under both high- and low-economic-uncertainty periods. Specifically, in a high-period,

the portfolio return is 1 bps for the low-centrality group and 49 bps for the high-centrality group, which reveals a difference of 48 bps with a Newey-West adjusted t-statistic of 2.71. The results are a bit weaker in the low-uncertainty period, but the overall pattern remains. In this case, the centrality measure cannot be fully explained by economic uncertainty.

In summary, the proposed centrality measure is not driven by the pricing factors listed above. Although we did not exhaust all possibilities, the facts suggest that the centrality measure serves well as an idiosyncratic risk factor to predict future crypto returns.

Table 4.6: Portfolio Returns: Trading Volume, Investor Attention, and Macro Uncertainty

This table reports the quartile portfolio returns sorted by the centrality measure for cryptos with high and low trading volume, in high- and low-investor-attention periods, or under high- and low-macro-uncertainty circumstances. Statistical significance is indicated by 1% 5% 10% for the positive signs and 1% 5% 10% for the negative signs. t-statistics in parentheses are computed based on standard errors with Newey-West lags(4) adjustment. The sample period spans from August 31, 2015 to March 31, 2018.

Centrality	Trading Volume		Investor Attention		Macro Uncertainty	
	Low	High	Low	High	Low	High
Low	0.05%	-0.04%	0.04%	0.01%	0.01%	0.01%
2	0.16%	0.27%	0.11%	0.21%	0.02%	0.26%
3	0.38%	0.12%	0.46%	0.32%	0.22%	0.54%
High	0.56%	0.28%	0.39%	0.46%	0.32%	0.49%
High - Low	0.51%	0.33%	0.35%	0.45%	0.31%	0.48%
t-statistic	(3.62)	(2.73)	(2.27)	(3.06)	(2.23)	(2.71)

4.5 Conclusion

This study examined the market segmentation problem in the crypto market. To solve the problem, we constructed a dynamic network of cryptos using return inter-predictive relationship and proposed a dynamic CASC method to make full use of the dynamic linkage information, as well as the node attributions, to improve classification accuracy. Based on the fitted crypto network and in the spirit of Ahern (2013), we proposed using eigenvector centrality as the idiosyncratic risk factor for

predicting future returns. We find that the cross-sectional portfolio constructed from eigenvector centrality sorting can deliver a persistent 40 bps daily return.

Chapter 5 Summary of Conclusions

In Chapter 2, studies the limit properties of information criteria for distinguishing between unit root model and three types of explosive models. Both the OLS estimator and the indirect inference estimator are employed to estimate the AR coefficient in the candidate model. This paper contributes to the literature in three aspects. First, our results extends results in the literature to the explosive side of the unit root, and we find that information criteria consistently choose the unit root model when the unit root model is the true model. Second, we show that the limiting probabilities for information criteria to select the explosive model depends on both the distance of autoregressive coefficient from unity and the size of penalty term in the information criteria. When the penalty term is not too large and the root is not too close to unit root, all the information criteria consistently select the true model. It is known that the indirect inference method is effective in reducing the bias in OLS estimation in all cases as well as reducing the variance in OLS estimation in the UR model and in the LTU model. However, when information criteria are used in connection with the indirect inference estimation, the limiting probabilities for information criteria to select the correct model can go up or down relative to that with the OLS estimation, depending on the true DGP. When the true DGP is the UR model, the indirect inference estimation increases the probability. When the true DGP is the LTUE model or the ME model or the EX model, the indirect inference estimation decreases the probability. This rather surprising result suggests that the superiority in estimation does not necessarily translate to the superiority in model selection.

In Chapter 3, introduces a continuous time model for financial data where the

persistence parameter is allowed to be random and time varying. The model has an analytical solution and an exact discrete time representation which make analysis convenient for studying the properties of the system that are associated with extreme sample path behavior. The discrete time model relates to some models already in the literature, including the stochastic unit root model (Granger and Swanson (1997); Lieberman and Phillips (2014); Lieberman and Phillips (2017c)) and the near-explosive random coefficient model of Banerjee et al. (2017). The statistical properties of our model reveal three different forms of potential extreme behavior in generated sample paths: instability, local explosiveness, and explosiveness. These forms of extreme behavior depend directly on the values of model parameters, including the possible presence of endogeneity in the random autoregressive coefficient.

A novel two-stage estimation method that relies on empirical quadratic variation is developed to estimate the model parameters. Limit theory is developed using an infill asymptotic scheme that provides a convenient basis for testing parameter constancy and the various forms of extreme sample path behavior. The test statistics all have asymptotically pivotal standard normal distributions which makes implementation of the tests straightforward in practical work. Similar to other recent work in the literature on bubbles, a time-stamping strategy is proposed to detect origination and termination dates of extreme behavior.

In an empirical application to daily S&P 500 real prices between December 31, 1927 and June 29, 2018. Strong evidence against parameter constancy is found in the whole sample period and this evidence strengthens after July 1997, leading to a finding of long durations of parameter instability in the model. Three periods of explosive instability in the data match well with observed periods of major price escalation in the data and these largely overlap with the periods of price exuberance identified in earlier work. Tests for endogeneity in these data provide strong evidence in support of endogenous feedbacks in the random coefficient model framework that materially influence quadratic variation and hence recursive estimates of

realized variation in the data. The empirical findings of extreme sample path behavior in real S&P 500 stock prices are broadly in line with the conclusions of other recent work on stock price exuberance but now provide new evidence against parameter constancy and in support of the role of endogenous feedbacks that influence autoregressive behavior and the time forms of extreme sample paths.

In Chapter 4, we examine the market segmentation problem in the crypto market. To solve the problem, we constructed a dynamic network of cryptos using return inter-predictive relationship and proposed a dynamic CASC method to make full use of the dynamic linkage information, as well as the node attributions, to improve classification accuracy. Based on the fitted crypto network and in the spirit of Ahern (2013), we proposed using eigenvector centrality as the idiosyncratic risk factor for predicting future returns. We find that the cross-sectional portfolio constructed from eigenvector centrality sorting can deliver a persistent 40 bps daily return.

Bibliography

ACEMOGLU, D., V. M. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016.

ACEMOGLU, D., A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2015): "Systemic Risk and Stability in Financial Networks," *American Economic Review*, 105, 564–608.

AHERN, K. R. (2013): "Network Centrality and the Cross Section of Stock Returns," *SSRN Electronic Journal, DOI: 10.2139/ssrn.2197370*.

AKAIKE, H. (1969): "Fitting autoregressive models for prediction," *Annals of the institute of Statistical Mathematics*, 21, 243–247.

——— (1974): "A new look at the statistical model identification," *Automatic Control, IEEE Transactions on*, 19, 716–723.

AMIHUD, Y. AND H. MENDELSON (1986): "Asset Pricing and the Bid-ask Spread," *Journal of Financial Economics*, 17, 223–249.

ANDERSON, T. W. (1959): "On asymptotic distributions of estimates of parameters of stochastic difference equations," *The Annals of Mathematical Statistics*, 676–687.

AOBDIA, D., J. CASKEY, AND N. B. OZEL (2014): "Inter-industry Network Structure and the Cross-predictability of Earnings and Stock Returns," *Review of Accounting Studies*, 19, 1191–1224.

AUE, A. (2008): "Near-integrated random coefficient autoregressive time series," *Econometric Theory*, 24, 1343–1372.

AUE, A. AND L. HORVÁTH (2011): "Quasi-likelihood estimation in stationary and nonstationary autoregressive models with random coefficients," *Statistica Sinica*, 21, 973–.

AUE, A., L. HORVÁTH, AND J. STEINEBACH (2006): "Estimation in random coefficient autoregressive models," *Journal of Time Series Analysis*, 27, 61–76.

BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2016): "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131, 1593–1636.

BANERJEE, A. N., G. CHEVILLON, AND M. KRATZ (2017): "Probabilistic Forecasting of Bubbles and Crashes," *Working Paper*, –.

BARBER, B. M. AND T. ODEAN (2008): "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," *Review of Financial Studies*, 21, 785–818.

BARNDORFF-NIELSEN, O. E. (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 253–280.

BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2002): "Estimating quadratic variation using realized variance," *Journal of Applied econometrics*, 17, 457–477.

——— (2005): *Identification and inference for econometric models. A Festschrift in honour of TJ Rothenberg*, chap. How accurate is the asymptotic approximation to the distribution of realized variance, 306–311.

BERKES, I., L. HORVÁTH, AND S. LING (2009): "Estimation in nonstationary random coefficient autoregressive models," *Journal of Time Series Analysis*, 30, 395–416.

BHATTACHARYYA, S. AND S. CHATTERJEE (2018): "Spectral Clustering for Multiple Sparse Networks: I," *arXiv preprint arXiv:1805.10594*.

BINKIEWICZ, N., J. T. VOGELSTEIN, AND K. ROHE (2017): "Covariate-assisted Spectral Clustering," *Biometrika*, 104, 361–377.

BONACICH, P. (1972): "Technique for Analyzing Overlapping Memberships," *Sociological Methodology*, 4, 176–185.

BORGATTI, S. P. (2005): "Centrality and Network Flow," *Social Networks*, 27, 55–71.

BOYER, B. H. (2011): "Style-Related Comovement: Fundamentals or Labels?" *The Journal of Finance*, 66, 307–332.

BROGAARD, J. AND A. DETZEL (2015): "The Asset-pricing Implications of Government Economic Policy Uncertainty," *Management Science*, 61, 3–18.

BYKHOVSKAYA, A. AND P. C. B. PHILLIPS (2018): "Boundary Limit Theory for Functional Local to Unity Regression," *Journal of Time Series Analysis*, 39, 523–562.

——— (2019): "Point Optimal Testing with Roots That Are Functionally Local to Unity," *Journal of Econometrics, Forthcoming*.

CAVALIERE, G., I. GEORGIEV, AND A. M. R. TAYLOR (2016): "Sieve-based inference for infinite-variance linear processes," *The Annals of Statistics*, 44, 1467–1494.

CHAN, N. H., D. LI, AND L. PENG (2012): "Toward a unified interval estimation of autoregressions," *Econometric Theory*, 28, 705–717.

CHAN, N. H. AND C.-Z. WEI (1987): "Asymptotic inference for nearly nonstationary AR (1) processes," *The Annals of Statistics*, 1050–1063.

CHAUDHURI, K., F. CHUNG, AND A. TSIATAS (2012): "Spectral Clustering of Graphs with General Degrees in the Extended Planted Partition Model," in *Conference on Learning Theory*, 35–1.

CHEN, C. Y.-H., W. K. HÄRDLE, A. J. HOU, AND W. WANG (2018): "Pricing Cryptocurrency Options: The Case of CRIX and Bitcoin," *SSRN Electronic Journal, DOI: 10.2139/ssrn.3159130*.

CHEN, C. Y.-H., W. K. HRDLE, AND Y. OKHRIN (2019): "Tail Event Driven Networks of SIFIs," *Journal of Econometrics*, 208, 282–298.

CHEN, K. AND J. LEI (2017): "Network Cross-validation for Determining the Number of Communities in Network Data," *Journal of the American Statistical Association*, 1–11.

CHEN, M., D. LI, AND S. LING (2014): "Non-Stationarity and Quasi-Maximum Likelihood Estimation on a Double Autoregressive Model," *Journal of Time Series Analysis*, 35, 189–202.

CHEN, Y., P. C. PHILLIPS, AND J. YU (2017): "Inference in Continuous Systems with Mildly Explosive Regressors," *Journal of Econometrics*, 201, 400–416.

CHONG, T. T.-L. (2001): "Structural Change in AR (1) Models," *Econometric Theory*, 87–155.

COHEN, L. AND A. FRAZZINI (2008): "Economic links and predictable returns," *The Journal of Finance*, 63, 1977–2011.

DEMIR, E., G. GOZGOR, C. K. M. LAU, AND S. A. VIGNE (2018): "Does Economic Policy Uncertainty Predict the Bitcoin Returns? an Empirical Investigation," *Finance Research Letters*, 26, 145–149.

DIBA, B. T. AND H. I. GROSSMAN (1988): "Explosive rational bubbles in stock prices?" *The American Economic Review*, 78, 520–530.

FÖLLMER, H., W. CHEUNG, AND M. A. H. DEMPSTER (1994): "Stock price fluctuation as a diffusion in a random environment [and discussion]," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 347, 471–483.

FÖLLMER, H. AND M. SCHWEIZER (1993): "A Microeconomic Approach to Diffusion Models for Stock Prices," *Mathematical Finance*, Vol. 3, 1–23.

FRANCQ, C. AND J.-M. ZAKOÏAN (2012): "Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models," *Econometrica*, 80, 821–861.

GIRAITIS, L., G. KAPETANIOS, AND T. YATES (2014): "Inference on stochastic time-varying coefficient models," *Journal of Econometrics*, 179, 46–65.

GOURIEROUX, C., A. MONFORT, AND E. RENAULT (1993): "Indirect inference," *Journal of Applied Econometrics*, 8, S85–S85.

GRANGER, C. W. J. (1980): "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227–238.

GRANGER, C. W. J. AND N. R. SWANSON (1997): "An introduction to stochastic unit-root processes," *Journal of Econometrics*, 80, 35–62.

GUO, L., L. PENG, Y. TAO, AND J. TU (2018): "News Co-Occurrence, Attention Spillover and Return Predictability," *SSRN Electronic Journal, DOI: 10.2139/ssrn.2927561*.

HANNAN, E. J. AND B. G. QUINN (1979): "The determination of the order of an autoregression," *Journal of the Royal Statistical Society. Series B (Methodological)*, 190–195.

HÄRDLE, W., C. HARVEY, AND R. REULE (2019): "Understanding Cryptocurrencies," *Journal of Financial Econometrics, Forthcoming*.

HÄRDLE, W. K. AND S. TRIMBORN (2015): "CRIX or Evaluating Blockchain Based Currencies," Tech. Rep. Oberwolfach Report No. 42/2015, The Mathematics and Statistics of Quantitative Risk.

HERSKOVIC, B. (2018): "Networks in Production: Asset Pricing Implications," *The Journal of Finance*, 73, 1785–1818.

HILL, J. AND L. PENG (2014): "Unified Interval Estimation for Random Coefficient Autoregressive Models," *Journal of Time Series Analysis*, 35, 282–297.

HOBERG, G. AND G. PHILLIPS (2016): "Text-based Network Industries and Endogenous Product Differentiation," *Journal of Political Economy*, 124, 1423–1465.

HÖPFNER, R. AND Y. KUTOYANTS (2003): "On a problem of statistical inference in null recurrent diffusions," *Statistical Inference for Stochastic Processes*, 6, 25–42.

HORVÁTH, L. AND L. TRAPANI (2016): "Statistical inference in a random coefficient panel model," *Journal of Econometrics*, 193, 54–75.

HWANG, S. Y. AND I. V. BASAWA (1997): "The local asymptotic normality of a class of generalized random coefficient autoregressive processes," *Statistics & probability letters*, 34, 165–170.

——— (1998): "Parameter estimation for generalized random coefficient autoregressive processes," *Journal of Statistical Planning and Inference*, 68, 323–337.

——— (2005): "Explosive Random-Coefficient AR (1) Processes and Related Asymptotics for Least-Squares Estimation," *Journal of Time Series Analysis*, 26, 807–824.

JENSEN, S. T. AND A. RAHBEK (2004): "Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case," *Econometrica*, 72, 641–646.

JIANG, L., X. WANG, AND J. YU (2017): "In-fill Asymptotic Theory for Structural Break Point in Autoregression: A Unified Theory," *Working Paper*.

KARRER, B. AND M. E. J. NEWMAN (2011): "Stochastic Blockmodels and Community Structure in Networks," *Physical Review E*, 83, 016107.

KIM, J. AND J. PARK (2016): "Mean Reversion and Stationarity in Financial Time Series Generated from Diffusion Models," *Working Paper*.

KLOCHKOV, Y., W. HÄRDLE, AND X. XIU (2019): "Localising Multivariate CAViaR," *IRTG 1792 Discussion Paper, under review in Journal of Econometrics*.

KRISTENSEN, D. (2012): "Non-parametric detection and estimation of structural change," *The Econometrics Journal*, 15, 420–461.

KUMAR, A. AND C. M. LEE (2006): "Retail Investor Sentiment and Return Comovements," *The Journal of Finance*, 61, 2451–2486.

LEE, D. K. C., L. GUO, AND Y. WANG (2018): "Cryptocurrency: A New Investment Opportunity?" *The Journal of Alternative Investments*, 20, 16–40.

LEE, S. (1998): "Coefficient constancy test in a random coefficient autoregressive model," *Journal of statistical planning and inference*, 74, 93–101.

LEI, J. AND A. RINALDO (2015): "Consistency of Spectral Clustering in Stochastic Block Models," *The Annals of Statistics*, 43, 215–237.

LEPSKI, O. V., E. MAMMEN, AND V. G. SPOKOINY (1997): "Optimal Spatial Adaptation to Inhomogeneous Smoothness: An Approach Based on Kernel Estimates with Variable Bandwidth Selectors," *The Annals of Statistics*, 929–947.

LI, T., E. LEVINA, AND J. ZHU (2016): "Network Cross-validation by Edge Sampling," *arXiv preprint arXiv:1612.04717*.

LIEBERMAN, O. AND P. C. B. PHILLIPS (2014): "Norming Rates and Limit Theory for Some Time-Varying Coefficient Autoregressions," *Journal of Time Series Analysis*, 35, 592–623.

——— (2017a): "Hybrid Stochastic Local Unit Roots," *Working Paper, Yale University*.

——— (2017b): "IV and GMM Estimation and Testing of Multivariate Stochastic Unit Root Models." *Econometric Theory, Forthcoming*.

——— (2017c): "A Multivariate Stochastic Unit Root Model with an Application to Derivative Pricing," *Journal of Econometrics*, 196, 99–110.

LING, S. AND D. LI (2008): "Asymptotic Inference for a Nonstationary Double AR(1) Model," *Biometrika*, 95, 257–263.

LIU, Y. AND A. TSYVINSKI (2018): "Risks and Returns of Cryptocurrency," *NBER Working Paper*.

MAGDALINOS, T. (2012): "Mildly explosive autoregression under weak and strong dependence," *Journal of Econometrics*, 169, 179–187.

NAGAKURA, D. (2009): "Testing for Coefficient Stability of AR(1) Model When the Null Is an Integrated or a Stationary Process," *Journal of Statistical Planning and Inference*, 139, 2731–2745.

NAKAMOTO, S. (2008): "Bitcoin: A Peer-to-peer Electronic Cash System," *Working Paper*.

NG, S. AND P. PERRON (1995): "Unit Root Tests in ARMA Models with Data-dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268–281.

NICHOLLS, D. AND B. QUINN (1980): "The Estimation of Random Coefficient Autoregressive Models. I," *Journal of Time Series Analysis*, 1, 37–46.

PANG, T., D. ZHANG, AND T. T.-L. CHONG (2014): "Asymptotic Inferences for an AR(1) Model with a Change Point: Stationary and Nearly Non-Stationary Cases," *Journal of Time Series Analysis*, 35, 133–150.

PENSKY, M. AND T. ZHANG (2017): "Spectral Clustering in the Dynamic Stochastic Block Model," *Arxiv Preprint Arxiv:1705.01204*.

PETUKHINA, A., S. TRIMBORN, W. K. HÄRDLE, AND H. ELENDNER (2018): "Investing With Cryptocurrencies – Evaluating the Potential of Portfolio Allocation Strategies," *IRTG 1792 Discussion Paper 2018-058*.

PHILLIPS, P. C. (1987a): "Time Series Regression with a Unit Root," *Econometrica: Journal of the Econometric Society*, 277–301.

——— (2008): "Unit Root Model Selection," *Journal of Japan Statistical Society*, 38, 65–74.

——— (2012): "Folklore Theorems, Implicit Maps, and Indirect Inference," *Econometrica*, 80, 425–.

PHILLIPS, P. C. AND J. H. LEE (2015): "Limit Theory for Vars with Mixed Roots near Unity," *Econometric Reviews*, 34, 1035–1056.

PHILLIPS, P. C. AND T. MAGDALINOS (2007): "Limit Theory for Moderate Deviations from a Unit Root," *Journal of Econometrics*, 136, 115–130.

——— (2009): "Unit Root and Cointegrating Limit Theory When Initialization Is in the Infinite Past," *Econometric Theory*, 25, 1682–1715.

PHILLIPS, P. C., Y. WU, AND J. YU (2011): "Explosive Behavior in the 1990s NASDAQ: When did Exuberance Escalate Asset Values?" *International Economic Review*, 52, 201–226.

PHILLIPS, P. C. AND J. YU (2011): "Dating the timeline of financial bubbles during the subprime crisis," *Quantitative Economics*, 2, 455–491.

PHILLIPS, P. C. B. (1987b): "Time Series Regression with a Unit Root," *Econometrica*, 277–301.

PHILLIPS, P. C. B. AND S.-P. SHI (2017): "Financial Bubble Implosion and Reverse Regression," *Econometric Theory*, 1–49.

PHILLIPS, P. C. B., S.-P. SHI, AND J. YU (2015a): "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&p 500," *International Economic Review*, 56, 1043–1078.

——— (2015b): "Testing for Multiple Bubbles: Testing for Multiple Bubbles: Limit Theory of Real Time Detectors," *International Economic Review*, 56, 1079–1134.

PHILLIPS, P. C. B. AND J. YU (2009): "A two-stage realized volatility approach to estimation of diffusion processes with discrete data," *Journal of Econometrics*, 150, 139–150.

PLOBERGER, W. AND P. C. PHILLIPS (2003): "Empirical limits for time series econometric models," *Econometrica*, 71, 627–673.

QIN, T. AND K. ROHE (2013): "Regularized Spectral Clustering under the Degree-corrected Stochastic Blockmodel," in *Advances in Neural Information Processing Systems*, 3120–3128.

ROHE, K., T. QIN, AND B. YU (2016): "Co-clustering Directed Graphs to Discover Asymmetries and Directional Communities," *Proceedings of the National Academy of Sciences*, 113, 12679–12684.

RUDELSON, M. AND R. VERSHYNIN (2013): "Hanson-wright Inequality and Subgaussian Concentration," *Electronic Communications in Probability*, 18.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

SHLEIFER, A. AND R. W. VISHNY (1997): "The Limits of Arbitrage," *The Journal of Finance*, 52, 35–55.

SMITH, A. A. (1993): "Estimating nonlinear time-series models using simulated vector autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.

SO, B. S. AND D. W. SHIN (1999): "Cauchy Estimators for Autoregressive Processes with Applications to Unit Root Tests and Confidence Intervals," *Econometric Theory*, 15, 165–176.

TAO, Y., P. C. PHILLIPS, AND J. YU (2018): "Online Supplement to: "Random Coefficient Continuous Systems: Testing for Extreme Sample Paths Behaviour"," *Working Paper Supplement*.

TRIMBORN, S., M. LI, AND W. K. HÄRDLE (2019): "Investing with Cryptocurrencies - A Liquidity Constrained Investment Approach," *Journal of Financial Econometrics, Forthcoming*.

WANG, X. AND J. YU (2016): "Double Asymptotics for Explosive Continuous Time Models," *Journal of Econometrics*, 193, 35–53.

WANG, Y. X. R. AND P. J. BICKEL (2017): "Likelihood-based Model Selection for Stochastic Block Models," *The Annals of Statistics*, 45, 500–528.

YU, J. (2012): "Bias in the estimation of the mean reversion parameter in continuous time models," *Journal of Econometrics*, 169, 114–122.

ZHANG, Y., M. POUX-BERTHE, C. WELLS, K. KOC-MICHALSKA, AND K. ROHE (2018): "Discovering Political Topics in Facebook Discussion Threads with Graph Contextualization," *Annals of Applied Statistics*, 12, 1096–1123.

ZHAO, Z. W. AND D. H. WANG (2012): "Statistical inference for generalized random coefficient autoregressive model," *Mathematical and Computer Modelling*, 56, 152–166.

ZOU, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429.

Appendix A: Appendix

A.1 Proofs in Chapter 2

A.1.1 Proof of Theorem 2.3.1

The proof is same as the proof for Theorem 1 in Phillips (2008), and hence omitted.

A.1.2 Proof of Theorem 2.3.2

When the true DGP is the LTUE model, we have $0 < c < \infty$ and

$$\begin{aligned} IC_0 = \log \widehat{\sigma}_0^2 &= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} \left(X_t - X_{t-1} \right)^2 \right\} \\ &= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\left(\rho_n - 1 \right) X_{t-1} + u_t \right]^2 \right\} \\ &= \log \left\{ \frac{1}{n} \left(\rho_n - 1 \right)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} \left(\rho_n - 1 \right) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\}. \end{aligned}$$

By Lemma 1 in Phillips (1987b), when the process is initialized at X_0, we know

$$\frac{1}{n^2} \sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 J_c^2, \tag{A.1.1}$$

and

$$\frac{1}{n} \sum_{t=1}^{n} X_{t-1} u_t \Rightarrow \sigma^2 \int_0^1 J_c dB, \tag{A.1.2}$$

where

$$J_c(r) = \int_0^r e^{c(r-s)} dB(s).$$

Therefore, by Equation (A.1.1) and (A.1.2) we have

$$\begin{aligned} IC_0 &= \log\left\{\frac{\sigma^2 c^2}{n}\int_0^1 J_c^2 + \frac{2c\sigma^2}{n}\int_0^1 J_c dB + \sigma^2 + o_p(n^{-1})\right\} \\ &= \log\sigma^2 + \log\left\{1 + \frac{2c}{n}\int_0^1 J_c(r)dB + \frac{c^2}{n}\int_0^1 J_c^2 + o_p(n^{-1})\right\}. \end{aligned} \tag{A.1.3}$$

We also know from Phillips (1987b) that

$$n\left(\widehat{\rho}_n - \rho_n\right) \Rightarrow \frac{\int_0^1 J_c dB}{\int_0^1 J_c^2}, \tag{A.1.4}$$

Hence,

$$\begin{aligned} IC_1 &= \log\widehat{\sigma}_1^2 + \frac{p_n}{n} \\ &= \log\left\{n^{-1}\sum_{t=1}^{n}\left(X_t - \widehat{\rho}_n X_{t-1}\right)^2\right\} + \frac{p_n}{n} \\ &= \log\left\{\frac{1}{n}\sum_{t=1}^{n}\left[\left(\rho_n - \widehat{\rho}_n\right)X_{t-1} + u_t\right]^2\right\} + \frac{p_n}{n} \\ &= \log\left\{\frac{1}{n}\left(\rho_n - \widehat{\rho}_n\right)^2\sum_{t=1}^{n}X_{t-1}^2 + \frac{2}{n}\left(\rho_n - \widehat{\rho}_n\right)\sum_{t=1}^{n}X_{t-1}u_t + \frac{1}{n}\sum_{t=1}^{n}u_t^2\right\} + \frac{p_n}{n} \\ &\Rightarrow \log\left\{-\frac{\sigma^2}{n}\frac{\left(\int_0^1 J_c dB\right)^2}{\int_0^1 J_c^2} + \sigma^2\right\} + \frac{p_n}{n} \\ &= \log\sigma^2 + \log\left\{1 - \frac{1}{n}\frac{\left(\int_0^1 J_c dB\right)^2}{\int_0^1 J_c^2}\right\} + \frac{p_n}{n}. \end{aligned} \tag{A.1.5}$$

Therefore, by Equation (A.1.3) and (A.1.5), we have

$$IC_1 - IC_0 \Rightarrow \log\left\{1 - \frac{1}{n}\frac{\left(\int_0^1 J_c dB\right)^2}{\int_0^1 J_c^2}\right\} - \log\left\{1 + \frac{2c}{n}\int_0^1 J_c dB + \frac{c^2}{n}\int_0^1 J_c^2\right\} + \frac{p_n}{n}.$$

Hence, if $p_n = 2$ (as in AIC), as $n \to \infty$, we have

$$n\left(IC_1 - IC_0\right) \Rightarrow 2 - \frac{\left(\int_0^1 J_c dB\right)^2}{\int_0^1 J_c^2} - 2c\int_0^1 J_c dB - c^2\int_0^1 J_c^2. \tag{A.1.6}$$

If $p_n \to \infty$ and $\frac{p_n}{n} \to 0$, we have

$$\frac{n}{p_n}\left(IC_1 - IC_0\right) \Rightarrow 1.$$

A.1.3 Proof of Theorem 2.3.3

When the true DGP is the ME model, we have

$$\begin{aligned} IC_0 &= \log \widehat{\sigma}_0^2 = \log\left\{\frac{1}{n}\sum_{t=1}^{n}\left(X_t - X_{t-1}\right)^2\right\} \\ &= \log\left\{\frac{1}{n}\sum_{t=1}^{n}\left[\left(\rho_n - 1\right)X_{t-1} + u_t\right]^2\right\} \\ &= \log\left\{\frac{1}{n}\left(\rho_n - 1\right)^2\sum_{t=1}^{n}X_{t-1}^2 + \frac{2}{n}\left(\rho_n - 1\right)\sum_{t=1}^{n}X_{t-1}u_t + \frac{1}{n}\sum_{t=1}^{n}u_t^2\right\}. \end{aligned} \tag{A.1.7}$$

According to Phillips and Magdalinos (2007), when the process is initialized at $X_0 = o_p(\sqrt{n/c_n})$, we have

$$\frac{c_n^2\rho_n^{-2n}}{n^2}\sum_{t=1}^{n}X_{t-1}^2 \Rightarrow \frac{\sigma^2}{4}Y^2, \tag{A.1.8}$$

$$\frac{c_n\rho_n^{-n}}{n}\sum_{t=1}^{n}X_{t-1}u_t \Rightarrow \frac{\sigma^2}{2}XY, \tag{A.1.9}$$

and

$$\frac{n\rho_n^n}{2c_n}\left(\widehat{\rho}_n - \rho_n\right) \Rightarrow \frac{X}{Y} \sim \mathcal{C}, \tag{A.1.10}$$

where $X, Y \sim \mathcal{N}\left(0,1\right)$ and $\mathcal{C}$ is a standard Cauchy variate.

Therefore, by (A.1.8) and (A.1.9) we have

$$\begin{aligned} IC_0 &\Rightarrow \log\left\{\frac{\sigma^2}{4n\rho_n^{-2n}}Y^2 + \frac{\sigma^2}{n\rho_n^{-n}}XY + \sigma^2\right\} \\ &= \log\sigma^2 + \log\left\{\frac{1}{4n\rho_n^{-2n}}Y^2 + \frac{1}{n\rho_n^{-n}}XY + 1\right\}. \end{aligned} \tag{A.1.11}$$

On the other hand,

$$\begin{aligned} IC_1 &= \log \widehat{\sigma}_1^2 + \frac{p_n}{n} \\ &= \log \left\{ n^{-1} \sum_{t=1}^{n} \left(X_t - \widehat{\rho}_n X_{t-1} \right)^2 \right\} + \frac{p_n}{n} \\ &= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\left(\rho_n - \widehat{\rho}_n \right) X_{t-1} + u_t \right]^2 \right\} + \frac{p_n}{n} \\ &= \log \left\{ \frac{1}{n} \left(\rho_n - \widehat{\rho}_n \right)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} \left(\rho_n - \widehat{\rho}_n \right) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}. \end{aligned}$$

By equation (A.1.8) to (A.1.10), we obtain

$$\begin{aligned} IC_1 &\Rightarrow \log \left\{ \frac{1}{n} \frac{4c_n^2}{n^2 \rho_n^{2n}} \mathcal{C}^2 \frac{n^2 \sigma^2}{4c_n^2 \rho_n^{-2n}} Y^2 - \frac{2}{n} \frac{2c_n}{n \rho_n^n} \mathcal{C} \frac{n\sigma^2}{2c_n \rho_n^{-n}} XY + \sigma^2 \right\} + \frac{p_n}{n} \\ &= \log \sigma^2 + \log \left\{ -\frac{1}{n} X^2 + 1 \right\} + \frac{p_n}{n}. \end{aligned} \tag{A.1.12}$$

Therefore, by equation (A.1.11) and (A.1.12), we have

$$IC_1 - IC_0 \Rightarrow \log \left\{ 1 - \frac{1}{n} X^2 \right\} - \log \left\{ 1 + \frac{1}{4n\rho_n^{-2n}} Y^2 + \frac{1}{n\rho_n^{-n}} XY \right\} + \frac{p_n}{n}.$$

Note $X^2, Y^2 \sim \chi^2(1)$ and $\rho_n^{-n} = o\left(c_n^{-1}\right)$. If $\lim_{n\to\infty} \frac{p_n}{\rho_n^{2n}} = \pi$,

$$\frac{n}{k_n} \left(IC_1 - IC_0 \right) \Rightarrow \begin{cases} \pi - \frac{1}{4}\chi^2(1), & \text{if } \pi \in [0, \infty) \\ 1, & \text{if } \pi = \infty \end{cases},$$

where

$$k_n = \begin{cases} \rho_n^{2n}, & \text{if } \pi \in [0, \infty) \\ p_n, & \text{if } \pi = \infty \end{cases}.$$

A.1.4 Proof of Theorem 2.3.4

When the true DGP is EX model, we have

$$\begin{aligned} IC_0 &= \log \widehat{\sigma}_0^2 = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} \left(X_t - X_{t-1} \right)^2 \right\} \\ &= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\left(\rho - 1 \right) X_{t-1} + u_t \right]^2 \right\} \\ &= \log \left\{ \frac{1}{n} \left(\rho - 1 \right)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} \left(\rho - 1 \right) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\}. \end{aligned}$$

By results established in Anderson (1959), we know

$$\frac{1}{\rho^{2n}} \sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \frac{\sigma^2 Y^2}{\left(\rho^2 - 1 \right)^2}, \tag{A.1.13}$$

$$\frac{1}{\rho^{n}} \sum_{t=1}^{n} X_{t-1} u_t \Rightarrow \frac{\sigma^2 XY}{\rho^2 - 1}, \tag{A.1.14}$$

$$\frac{\rho^n}{\rho^2 - 1} \left(\widehat{\rho} - \rho \right) \Rightarrow \mathcal{C}, \tag{A.1.15}$$

where $X, Y \overset{iid}{\sim} \mathcal{N}\left(0, 1 \right)$ and $\mathcal{C}$ is a standard Cauchy variate. Then we have

$$\begin{aligned} IC_0 &= \log \left\{ \frac{\sigma^2 \rho^{2n}}{n \left(\rho + 1 \right)^2} X^2 + \frac{2\sigma^2 \rho^n}{n \left(\rho + 1 \right)} XY + \sigma^2 \right\} \\ &= \log \sigma^2 + \log \left\{ \frac{\rho^{2n}}{n \left(\rho + 1 \right)^2} X^2 + \frac{2\rho^n}{n \left(\rho + 1 \right)} XY + 1 \right\}. \end{aligned} \tag{A.1.16}$$

For the OLS estimator for the general explosive series, we have

$$\begin{aligned} IC_1 &= \log \widehat{\sigma}_1^2 + \frac{p_n}{n} \\ &= \log \left\{ n^{-1} \sum_{t=1}^{n} \left(X_t - \widehat{\rho} X_{t-1} \right)^2 \right\} + \frac{p_n}{n} \\ &= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\left(\rho - \widehat{\rho} \right) X_{t-1} + u_t \right]^2 \right\} + \frac{p_n}{n} \\ &= \log \left\{ \frac{1}{n} \left(\rho - \widehat{\rho} \right)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} \left(\rho - \widehat{\rho} \right) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n}. \end{aligned}$$

By equation (A.1.13) to (A.1.15), we have

$$IC_1 = \log \sigma^2 + \log \left\{ 1 - \frac{1}{n} X^2 \right\} + \frac{p_n}{n}. \tag{A.1.17}$$

Now, by equation (A.1.16) and (A.1.17), we obtain

$$IC_1 - IC_0 = \log \left\{ 1 - \frac{1}{n} X^2 \right\} - \log \left\{ 1 + \frac{2\rho^n}{n(\rho+1)} XY + \frac{\rho^{2n}}{n(\rho+1)^2} X^2 \right\} + \frac{p_n}{n}.$$

Since $\lim_{n\to\infty} \frac{p_n}{\rho^{2n}} = \pi$, we have

$$\frac{n}{k_n}(IC_1 - IC_0) \Rightarrow \begin{cases} \pi - \frac{1}{(1+\rho)^2}\chi^2(1), & \text{if } \pi \in [0, \infty) \\ 1, & \text{if } \pi = \infty \end{cases},$$

where

$$k_n = \begin{cases} \rho^{2n}, & \text{if } \pi \in [0, \infty) \\ p_n, & \text{if } \pi = \infty \end{cases}.$$

A.1.5 Proof of Proposition 2.3.1

When the true DGP is ME model, we have $0 < c < \infty$, and

$$\begin{aligned} IC_0 = \log \widehat{\sigma}_0^2 &= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} (X_t - X_{t-1})^2 \right\} \\ &= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[(\rho_n - 1) X_{t-1} + u_t \right]^2 \right\} \\ &= \log \left\{ \frac{1}{n} (\rho_n - 1)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} (\rho_n - 1) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\}. \end{aligned}$$

When the process is initialized at X_0, by Lemma 5 in Magdalinos (2012), we know

$$\frac{c_n^2 \rho_n^{-2n}}{\omega^2 n^2} \sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \frac{\sigma^2}{4} Z^2, \tag{A.1.18}$$

and

$$\frac{c_n \rho_n^{-n}}{\omega^2 n} \sum_{t=1}^{n} X_{t-1} u_t \Rightarrow \frac{\sigma^2}{2} YZ, \tag{A.1.19}$$

where by Lemma 2 in Magdalinos (2012), we know Y and Z are independent $\mathcal{N}(0,1)$ variates with $\omega^2 = \left(\sum_{j=0}^{\infty} F_j\right)^2$.

Therefore, by Equation (A.1.18) and (A.1.19) we have

$$\begin{aligned} IC_0 &\Rightarrow \log\left\{\frac{\omega^2\sigma^2}{4n\rho_n^{-2n}} Z^2 + \frac{\omega^2\sigma^2}{n\rho_n^{-n}} YZ + \sigma^2 + o_p(n^{-1})\right\} \\ &= \log\sigma^2 + \log\left\{1 + \frac{\omega^2}{n\rho_n^{-n}} YZ + \frac{\omega^2}{4n\rho_n^{-2n}} Z^2 + o_p(n^{-1})\right\}. \end{aligned} \tag{A.1.20}$$

We also know from Magdalinos (2012) that

$$\frac{n\rho_n^n}{2c_n}\left(\widehat{\rho}_n - \rho_n\right) \Rightarrow \mathcal{C}. \tag{A.1.21}$$

Hence,

$$\begin{aligned} IC_1 &= \log\widehat{\sigma}_1^2 + \frac{p_n}{n} \\ &= \log\left\{n^{-1}\sum_{t=1}^{n}\left(X_t - \widehat{\rho}_n X_{t-1}\right)^2\right\} + \frac{p_n}{n} \\ &= \log\left\{\frac{1}{n}\sum_{t=1}^{n}\left[\left(\rho_n - \widehat{\rho}_n\right) X_{t-1} + u_t\right]^2\right\} + \frac{p_n}{n} \\ &= \log\left\{\frac{1}{n}\left(\rho_n - \widehat{\rho}_n\right)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n}\left(\rho_n - \widehat{\rho}_n\right)\sum_{t=1}^{n} X_{t-1}u_t + \frac{1}{n}\sum_{t=1}^{n} u_t^2\right\} + \frac{p_n}{n} \\ &\Rightarrow \log\left\{-\frac{\omega^2\sigma^2}{n} Y^2 + \sigma^2\right\} + \frac{p_n}{n} \\ &= \log\sigma^2 + \log\left\{1 - \frac{\omega^2}{n} Y^2\right\} + \frac{p_n}{n}. \end{aligned} \tag{A.1.22}$$

Therefore, by Equation (A.1.20) and (A.1.22), we have

$$IC_1 - IC_0 \Rightarrow \log\left\{1 - \frac{\omega^2}{n} Y^2\right\} - \log\left\{1 + \frac{\omega^2}{n\rho_n^{-n}} YZ + \frac{\omega^2}{4n\rho_n^{-2n}} Z^2\right\} + \frac{p_n}{n}.$$

Note $Y^2, Z^2 \sim \chi^2(1)$ and $\rho_n^{-n} = o\left(c_n^{-1}\right)$. If $\lim_{n\to\infty} \dfrac{p_n}{\rho_n^{2n}} = \pi$,

$$\frac{n}{k_n}\left(IC_1 - IC_0\right) \Rightarrow \begin{cases} \pi - \dfrac{\omega^2}{4}\chi^2(1), & \text{if } \pi \in [0,\infty) \\ 1, & \text{if } \pi = \infty \end{cases},$$

where

$$k_n = \begin{cases} \rho_n^{2n}, & \text{if } \pi \in [0,\infty) \\ p_n, & \text{if } \pi = \infty \end{cases}.$$

A.1.6 Proof of Theorem 2.4.1

When the true DGP is the UR model, we have

$$IC_0 = \log \breve{\sigma}_0^2 = \log\left\{\frac{1}{n}\sum_{t=1}^{n} u_t^2\right\} = \log \sigma^2.$$

Also, we have

$$\begin{aligned} IC_1 &= \log \breve{\sigma}_1^2 + \frac{p_n}{n} = \log\left\{\frac{1}{n}\sum_{t=1}^{n}\left(X_t - \breve{\rho}X_{t-1}\right)^2\right\} + \frac{p_n}{n} \\ &= \log\left\{\frac{1}{n}\sum_{t=1}^{n}\left[\left(1-\breve{\rho}\right)X_{t-1} + u_t\right]^2\right\} + \frac{p_n}{n} \\ &= \log\left\{\frac{1}{n}\left(1-\breve{\rho}\right)^2\sum_{t=1}^{n}X_{t-1}^2 + \frac{2}{n}\left(1-\breve{\rho}\right)\sum_{t=1}^{n}X_{t-1}u_t + \frac{1}{n}\sum_{t=1}^{n}u_t^2\right\} + \frac{p_n}{n}. \end{aligned}$$

According to Phillips (2012), we have

$$\breve{\rho} - 1 \Rightarrow \begin{cases} \dfrac{1}{n}h^{-1}\left(\dfrac{\int_0^1 B dB}{\int_0^1 B^2}\right), & \text{if } \tau = 0 \\ \dfrac{1}{n}h^{-1}\left(\dfrac{\int_0^1 B_\tau dB}{\int_0^1 B_\tau^2}\right), & \text{if } \tau \in (0,\infty) \\ \dfrac{1}{\sqrt{n^2/c_n}}h^{-1}\left(\mathcal{C}\right), & \text{if } \tau = \infty \end{cases},$$

where $h(c)$ was defined in Section 4.

According to Phillips and Magdalinos (2009), we have

$$\frac{1}{n^2}\sum_{t=1}^{n} X_{t-1}^2 \Rightarrow \begin{cases} \sigma^2 \int_0^1 B^2, & \text{if } \tau = 0 \\ \sigma^2 \int_0^1 B_\tau^2, & \text{if } \tau \in (0,\infty) \\ \sigma^2 B_0(1)^2/c_n, & \text{if } \tau = \infty \end{cases},$$

$$\frac{1}{n}\sum_{t=1}^{n} X_{t-1}u_t \Rightarrow \begin{cases} \sigma^2 \int_0^1 B dB, & \text{if } \tau = 0 \\ \sigma^2 \int_0^1 B_\tau dB, & \text{if } \tau \in (0,\infty) \\ \sqrt{1/c_n}\sigma^2 B(1)B_0(1), & \text{if } \tau = \infty \end{cases}.$$

Therefore, we have

$$IC_0 - IC_1 \Rightarrow \begin{cases} -\log\left\{\frac{\int_0^1 B^2}{n} h^{-1}\left(\frac{\int_0^1 B dB}{\int_0^1 B^2}\right)^2 - \frac{2\left(\int_0^1 B dB\right)}{n} h^{-1}\left(\frac{\int_0^1 B dB}{\int_0^1 B^2}\right) + 1\right\} - \frac{p_n}{n} \\ -\log\left\{\frac{\int_0^1 B_\tau^2}{n} h^{-1}\left(\frac{\int_0^1 B_\tau dB}{\int_0^1 B_\tau^2}\right)^2 - \frac{2\left(\int_0^1 B_\tau dB\right)}{n} h^{-1}\left(\frac{\int_0^1 B_\tau dB}{\int_0^1 B_\tau^2}\right) + 1\right\} - \frac{p_n}{n} \\ -\log\left\{\frac{1}{n} h^{-1}\left(\mathcal{C}\right)^2 B_0(1)^2 - \frac{2}{n} h^{-1}\left(\mathcal{C}\right) B(1)B_0(1) + 1\right\} - \frac{p_n}{n} \end{cases}.$$

A.1.7 Proof of Theorem 2.4.2

When the true DGP is the LTUE model, we have $0 < c < \infty$. There is no difference between IC_0 based on the OLS estimator and that based on the indirect inference estimator. For IC_1, we have

$$\begin{aligned} IC_1 &= \log \breve{\sigma}_1^2 + \frac{p_n}{n} \\ &= \log\left\{n^{-1}\sum_{t=1}^{n}\left(X_t - \breve{\rho}_n X_{t-1}\right)^2\right\} + \frac{p_n}{n} \\ &= \log\left\{\frac{1}{n}\sum_{t=1}^{n}\left[\left(\rho_n - \breve{\rho}_n\right)X_{t-1} + u_t\right]^2\right\} + \frac{p_n}{n} \\ &= \log\left\{\frac{1}{n}\left(\rho_n - \breve{\rho}_n\right)^2\sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n}\left(\rho_n - \breve{\rho}_n\right)\sum_{t=1}^{n} X_{t-1}u_t + \frac{1}{n}\sum_{t=1}^{n} u_t^2\right\} + \frac{p_n}{n}. \end{aligned}$$

By the limit theory for the indirect inference estimator developed in Phillips (2012), we have

$$n\left(\breve{\rho}_n - \rho_n\right) \Rightarrow h^{-1}\left(\frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c\right) - c. \tag{A.1.23}$$

By equation (A.1.1), (A.1.2) and (A.1.23), we have

$$IC_1 \Rightarrow \log\left\{1 - \frac{2}{n}\left[h^{-1}\left(\frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c\right) - c\right]\int_0^1 J_c dB + \frac{1}{n}\left[h^{-1}\left(\frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c\right) - c\right]^2 \int_0^1 J_c^2\right\}$$
$$+ \log\sigma^2 + \frac{p_n}{n}. \tag{A.1.24}$$

Therefore, by equation (A.1.11) and (A.1.23), we have

$$IC_1 - IC_0 \Rightarrow \log\left\{1 - \frac{2\int_0^1 J_c dB}{n}\left[h^{-1}\left(\frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c\right) - c\right] + \frac{\int_0^1 J_c^2}{n}\left[h^{-1}\left(\frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c\right) - c\right]\right.$$
$$\left. - \log\left\{1 + \frac{2c}{n}\int_0^1 J_c dB + \frac{c^2}{n}\int_0^1 J_c^2\right\} + \frac{p_n}{n}.\right.$$

When $p_n = 2$, as $n \to \infty$ we have

$$n\left(IC_1 - IC_0\right) \Rightarrow 2 - \vartheta^2.$$

where

$$\vartheta^2 \equiv 2h^{-1}\left(\frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c\right)\left(\int_0^1 J_c dB + c\int_0^1 J_c^2\right) - h^{-1}\left(\frac{\int_0^1 J_c dB}{\int_0^1 J_c^2} + c\right)^2 \int_0^1 J_c^2.$$

When $p_n \to \infty$ and $\frac{p_n}{n} \to 0$, we have

$$\frac{n}{p_n}\left(IC_1 - IC_0\right) \Rightarrow 1.$$

A.1.8 Proof of Theorem 2.4.3

When the true DGP is the ME model, we have

$$\begin{aligned} IC_0 &= \log \breve{\sigma}_0^2 = \log \left\{ \frac{1}{n} \sum_{t=1}^{n} \left(X_t - X_{t-1} \right)^2 \right\} \\ &= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\left(\rho_n - 1 \right) X_{t-1} + u_t \right]^2 \right\} \\ &= \log \left\{ \frac{1}{n} \left(\rho_n - 1 \right)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} \left(\rho_n - 1 \right) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\}. \end{aligned}$$

By equation (A.1.8) and (A.1.9) we have

$$\begin{aligned} IC_0 &\Rightarrow \log \left\{ \frac{\sigma^2}{4n\rho_n^{-2n}} Y^2 + \frac{\sigma^2}{n\rho_n^{-n}} XY + \sigma^2 \right\} \\ &= \log \sigma^2 + \log \left\{ \frac{1}{4n\rho_n^{-2n}} Y^2 + \frac{1}{n\rho_n^{-n}} XY + 1 \right\}. \end{aligned} \tag{A.1.25}$$

Similarly, for IC_1 based on the indirect inference estimator, we have

$$\begin{aligned} IC_1 &= \log \breve{\sigma}_1^2 + \frac{p_n}{n} \\ &= \log \left\{ n^{-1} \sum_{t=1}^{n} \left(X_t - \breve{\rho}_n X_{t-1} \right)^2 \right\} + \frac{p_n}{n} \\ &= \log \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[\left(\rho_n - \breve{\rho}_n \right) X_{t-1} + u_t \right]^2 \right\} + \frac{p_n}{n} \\ &= \log \left\{ \frac{1}{n} \left(\rho_n - \breve{\rho}_n \right)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n} \left(\rho_n - \breve{\rho}_n \right) \sum_{t=1}^{n} X_{t-1} u_t + \frac{1}{n} \sum_{t=1}^{n} u_t^2 \right\} + \frac{p_n}{n} \end{aligned}$$

Using the results in Phillips (2012) , equation (A.1.8) and (A.1.9), we obtain

$$\begin{aligned} IC_1 &\Rightarrow \log \left\{ \frac{4c_n^2}{n^3 \rho_n^{2n}} \left(\mathcal{C} + O\left(\frac{1}{2c_n} \right) \right)^2 \frac{n^2 \sigma^2}{4c_n^2 \rho_n^{-2n}} Y^2 - \frac{4c_n}{n^2 \rho_n^n} \left(\mathcal{C} + O\left(\frac{1}{2c_n} \right) \right) \frac{n\sigma^2}{2c_n \rho_n^{-n}} XY + \sigma^2 \right\} + \frac{1}{} \\ &\Rightarrow \log \sigma^2 + \log \left\{ 1 - \frac{1}{n} X^2 + O\left(\frac{1}{c_n n} \right) \right\} + \frac{p_n}{n}. \end{aligned} \tag{A.1.26}$$

Therefore, the similar results to those in Theorem 2.3.3 are obtained.

A.1.9 Proof of Theorem 2.4.4

When the true DGP is the EX model, for the indirect inference estimator, we know that for IC_0, it is the same as OLS estimator. Therefore, we only need to derive the IC_1. Note that for IC_1, we have

$$IC_1 = \log\left\{\frac{1}{n}\left(\rho - \breve{\rho}\right)^2 \sum_{t=1}^{n} X_{t-1}^2 + \frac{2}{n}\left(\rho - \breve{\rho}\right)\sum_{t=1}^{n} X_{t-1}u_t + \frac{1}{n}\sum_{t=1}^{n} u_t^2\right\} + \frac{p_n}{n}.$$

According to the results in Phillips (2012), for $|\rho| > 1$, we know the binding function for ρ is

$$b_n(\rho) = \rho + O(\rho^{-n}).$$

Therefore, we obtain

$$\begin{aligned} IC_1 &= \log\sigma^2 + \log\left\{\frac{1}{n}\left(\mathcal{C} + O\left(\frac{1}{\rho^2-1}\right)\right)^2 Y^2 - \frac{2}{n}\left(\mathcal{C} + O\left(\frac{1}{\rho^2-1}\right)\right)XY + 1\right\} + \frac{p_n}{n} \\ &= \log\sigma^2 + \log\left\{1 - \frac{1}{n}X^2 + O\left(\frac{1}{n\left(\rho^2-1\right)}\right)\right\} + \frac{p_n}{n}. \end{aligned} \tag{A.1.27}$$

Now, by equation (A.1.16) and (A.1.27), we obtain

$$IC_1 - IC_0 = \log\left\{1 - \frac{1}{n}X^2 + O\left(\frac{1}{n\left(\rho^2-1\right)}\right)\right\} - \log\left\{1 + \frac{2\rho^n}{n\left(\rho+1\right)}XY + \frac{\rho^{2n}}{n\left(\rho+1\right)^2}X^2\right\} + \frac{p_n}{n}.$$

Since $\lim_{n\to\infty}\frac{p_n}{\rho^{2n}} = \pi$, we have

$$\frac{n}{k_n}\left(IC_1 - IC_0\right) \Rightarrow \begin{cases} \pi - \dfrac{1}{(1+\rho)^2}\chi^2(1), & \text{if } \pi \in [0,\infty) \\ 1, & \text{if } \pi = \infty \end{cases},$$

where

$$k_n = \begin{cases} \rho^{2n}, & \text{if } \pi \in [0,\infty) \\ p_n, & \text{if } \pi = \infty \end{cases}.$$

A.2 Proof of Chapter 3

A.2.1 List of Symbols and Notations

The symbols and notations used in Chapter 3 are summarized in the following table.

Symbol	Meaning	Definition
$\widetilde{\mu}$	Drift parameter of the random coefficient	Equation (3.2.2)
$\widetilde{\sigma}$	Diffusion parameter of the random coefficient	Equation (3.2.2)
σ	Diffusion parameter of the Ornstein-Uhlenbeck process in a random environment	Equation (3.2.2)
B_u, B_ε	Standard Brownian motion that generates randomness in the drift component/the diffusion component	Equation (3.2.2)
γ	Covariance between the Brownian motions B_u and B_ε	$\gamma = \frac{d\langle B_u, B_\varepsilon\rangle_t}{dt}$
ω	Covariance between the drift process and diffusion process	$\omega = \gamma\widetilde{\sigma}\sigma$
Δ, h	Sampling interval/block	
M	Number of observations within each sampling block	$M = h/\Delta$
T	Time span	
N	Number of sampling blocks	$N = T/h$
$\rho_{t\Delta}$	Random autoregressive coefficient	Equation (3.2.8)
$\eta_{t\Delta}$	Innovation of the random coefficient autoregression	Equation (3.2.9)
ρ	Expectation of the random autoregressive coefficient	$\rho = E(\rho_{t\Delta}) = e^{\widetilde{\mu}\Delta}$
κ	Key parameter controls the unstable behavior	$\kappa = \widetilde{\mu} + \widetilde{\sigma}^2/2$
ϕ	Key parameter controls the explosive behavior	$\phi = \widetilde{\mu} - \widetilde{\sigma}^2/2$
β_κ, β_ϕ	Exponential transformation of κ and ϕ	$\beta_\kappa = e^{\kappa\Delta}$, $\beta_\phi = e^{\phi\Delta}$
$[y]_a^b$	Quadratic variation process of y_t for $t \in [a, b]$	Equation (3.3.1)
$Q_\Delta(\cdot)$	Objective function of the first-stage estimation	Equation (3.3.3)
$\ell_{ALF}(\cdot)$	Approximate log-likelihood function in the second stage	Equation (3.3.5)

The proofs of Theorem 3.4.1 and 3.5.1 follow directly from Phillips and Yu (2009) and are omitted.

A.2.2 Proof of Theorem 3.4.2

To show consistency of $\widehat{\widetilde{\mu}}$, by the Taylor expansion of $\exp\{\widetilde{\mu}_0\Delta\}$ in the observationally equivalent model (3.2.11), we have

$$\begin{aligned} y_{t\Delta} - y_{(t-1)\Delta} &= \exp\{\widetilde{\mu}_0\Delta\}\, y_{(t-1)\Delta} - y_{(t-1)\Delta} + \sqrt{(\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2)\Delta}\cdot v_{t\Delta} \\ &= (\widetilde{\mu}_0\Delta + o(\Delta))y_{(t-1)\Delta} + \sqrt{(\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2)\Delta}\cdot v_{t\Delta}, \end{aligned} \tag{A.2.1}$$

where $v_{t\Delta} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$. Let $\zeta_{(t-1)\Delta} := \dfrac{1}{\widehat{\widetilde{\sigma}^2} y^2_{(t-1)\Delta} + \widehat{\sigma^2}} - \dfrac{1}{\widetilde{\sigma}_0^2 y^2_{(t-1)\Delta} + \sigma_0^2}$, then according to (3.3.6), we have

$$\begin{aligned}
\widehat{A}_N &= \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}\left(y_{t\Delta} - y_{(t-1)\Delta}\right)}{\widehat{\widetilde{\sigma}^2} y^2_{(t-1)\Delta} + \widehat{\sigma^2}} \\
&= \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}\left(y_{t\Delta} - y_{(t-1)\Delta}\right)}{\widetilde{\sigma}_0^2 y^2_{(t-1)\Delta} + \sigma_0^2} + \sum_{t=1}^{M\times N} y_{(t-1)\Delta}\left(y_{t\Delta} - y_{(t-1)\Delta}\right)\zeta_{(t-1)\Delta} \\
&:= A_N + R_A,
\end{aligned}$$

and

$$\widehat{B}_N = \sum_{t=1}^{M\times N} \frac{y^2_{(t-1)\Delta}}{\widehat{\widetilde{\sigma}^2} y^2_{(t-1)\Delta} + \widehat{\sigma^2}} = \sum_{t=1}^{M\times N} \frac{y^2_{(t-1)\Delta}}{\widetilde{\sigma}_0^2 y^2_{(t-1)\Delta} + \sigma_0^2} + \sum_{t=1}^{M\times N} y^2_{(t-1)\Delta}\zeta_{(t-1)\Delta} := B_N + R_B.$$

Note that, by Theorem 3.4.1, we have

$$\zeta_{(t-1)\Delta} = \frac{(\widetilde{\sigma}_0^2 - \widehat{\widetilde{\sigma}^2}) y^2_{(t-1)\Delta} + (\sigma_0^2 - \widehat{\sigma^2})}{(\widehat{\widetilde{\sigma}^2} y^2_{(t-1)\Delta} + \widehat{\sigma^2})(\widetilde{\sigma}_0^2 y^2_{(t-1)\Delta} + \sigma_0^2)} = O_p\left(\frac{\sqrt{\Delta}}{y^2_{(t-1)\Delta}}\right).$$

This implies that

$$\begin{aligned}
R_A &= \widetilde{\mu}_0\Delta \sum_{t=1}^{M\times N} y^2_{(t-1)\Delta}\zeta_{(t-1)\Delta} + \sqrt{\Delta} \sum_{t=1}^{M\times N} y_{(t-1)\Delta} v_{t\Delta} \sqrt{\widetilde{\sigma}_0^2 y^2_{(t-1)\Delta} + \sigma_0^2}\zeta_{(t-1)\Delta} \\
&= O_p(MN\Delta^{3/2}) + O_p(\sqrt{MN}\Delta) = O_p(T\sqrt{\Delta}) + O_p(\sqrt{T\Delta}) = O_p(T\sqrt{\Delta}),
\end{aligned}$$

and

$$R_B = \sum_{t=1}^{M\times N} y^2_{(t-1)\Delta}\zeta_{(t-1)\Delta} = O_p(MN\sqrt{\Delta}) = O_p(T/\sqrt{\Delta}).$$

Now, first solve

$$\begin{aligned}
A_N &= \widetilde{\mu}_0 \sum_{t=1}^{M\times N} \frac{y^2_{(t-1)\Delta}\cdot\Delta}{\widetilde{\sigma}_0^2 y^2_{(t-1)\Delta} + \sigma_0^2} + \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta} v_{t\Delta}\cdot\sqrt{\Delta}}{\sqrt{\widetilde{\sigma}_0^2 y^2_{(t-1)\Delta} + \sigma_0^2}} + \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}\varsigma_{t\Delta}}{\widetilde{\sigma}_0^2 y^2_{(t-1)\Delta} + \sigma_0^2} \\
&= \widetilde{\mu}_0\Delta B_N + \sqrt{\Delta} C_N + D_N,
\end{aligned}$$

where

$$C_N := \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta} v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2}},$$

$$D_N := \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta} \varsigma_{t\Delta}}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2}$$

$$= \frac{\Delta}{2}\widetilde{\sigma}^2 \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}^2 \left(u_{t\Delta}^2 - 1\right)}{\widetilde{\sigma}^2 y_{(t-1)\Delta}^2 + \sigma^2} + \sum_{t=1}^{M\times N} \frac{\sigma y_{(t-1)\Delta} w_{t\Delta} o(\Delta^{1/2})}{\widetilde{\sigma}^2 y_{(t-1)\Delta}^2 + \sigma^2} + \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}^2 O(\Delta^{3/2})}{\widetilde{\sigma}^2 y_{(t-1)\Delta}^2 + \sigma^2}.$$

Denote $U_{t\Delta} := \dfrac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2}$. Clearly $\{U_{t\Delta}\}$ is bounded for all t. Let us look at the asymptotic behavior of B_N. First consider the case $\phi < 0$ where $\{y_{t\Delta}\}$ is asymptotically stationary and ergodic as shown in Föllmer and Schweizer (1993). In this case $\{U_{t\Delta}\}$ is also asymptotically stationary and ergodic. By the ergodic theorem we then have,

$$\frac{1}{MN} \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2} \xrightarrow{a.s.} E\left(\frac{y_t^2}{\widetilde{\sigma}_0^2 y_t^2 + \sigma_0^2}\right).$$

In the case $\phi \geq 0$ where $\{y_t\}$ is nonstationary, by equation (3.2.9),

$$y_{(t-1)\Delta} = \rho_{1\Delta} y_0 + \sigma_0 \left(\eta_{(t-1)\Delta} + \rho_{(t-1)\Delta}\eta_{(t-2)\Delta} + \cdots + \rho_{(t-1)\Delta}\rho_{(t-2)\Delta}\cdots\rho_{2\Delta}\eta_{1\Delta}\right).$$

Hence, we have

$$E\left(\left|U_{t\Delta} - \widetilde{\sigma}_0^{-2}\right| \middle| \rho_{(t-1)\Delta}, \cdots, \rho_{2\Delta}\right) = E\left[\left|\frac{\sigma_0^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2}\right| \middle| \rho_{(t-1)\Delta}, \cdots, \rho_{2\Delta}\right]$$
$$= \int_0^\infty \frac{\sigma_0^2}{\widetilde{\sigma}_0^2 y_0^2 + \widetilde{\sigma}_0^2 \lambda^2 x^2 + \sigma_0^2} d\Phi(x),$$

where $\Phi(\cdot)$ is the cdf of $\eta_{t\Delta} \sim N(0, \gamma_\Delta^2)$ and $\lambda^2 = 1 + \rho_{(t-1)\Delta}^2 + \cdots + \rho_{(t-1)\Delta}^2\rho_{(t-2)\Delta}^2\cdots\rho_{2\Delta}^2 \xrightarrow{p} \infty$ as $t \to \infty$. Thus, $E\left(\left|U_{t\Delta} - \widetilde{\sigma}_0^{-2}\right| \middle| \rho_{(t-1)\Delta}, \cdots, \rho_{2\Delta}\right) \xrightarrow{p} 0$. Since $\{U_{t\Delta}\}$ is bounded, $E\left|U_{t\Delta} - \widetilde{\sigma}_0^{-2}\right| \xrightarrow{p} 0$ as $t \to \infty$. Consequently,

$$\frac{1}{MN} \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2} \xrightarrow{p} \widetilde{\sigma}_0^{-2}.$$

To unify the results for both stationary and nonstationary cases, we write

$$\frac{1}{MN}\sum_{t=1}^{M\times N}\frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2+\sigma_0^2}\xrightarrow{p}V_1, \tag{A.2.2}$$

where V_1 is defined in (3.4.1). This result implies that $\Delta B_N = O_p(MN\Delta) = O_p(T)$.

To examine the asymptotic behavior of C_N, denote

$$\xi_t := \frac{y_{(t-1)\Delta}v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2+\sigma_0^2}},$$

and note that ξ_t is a martingale difference sequence with respect to the filtration $\mathcal{F}_t := \sigma(B_v(s) : 0 \le s \le t)$ because

$$E\left(\xi_t|\mathcal{F}_{t-1}\right) = E\left(\left.\frac{y_{(t-1)\Delta}v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2+\sigma_0^2}}\right|\mathcal{F}_{t-1}\right) = \frac{y_{(t-1)\Delta}}{\sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2+\sigma_0^2}}E\left(v_{t\Delta}|\,\mathcal{F}_{t-1}\right) = 0.$$

To apply the martingale CLT to $\{\xi_t\}$, we need to check the stability and Lindeberg conditions. For the stability condition, the conditional variance of the standardized martingale is

$$\left\langle\frac{1}{\sqrt{MN}}\sum_{t=1}^{M\times N}\xi_t\right\rangle = \frac{1}{MN}\sum_{t=1}^{M\times N}E\left(\xi_t^2|\mathcal{F}_{t-1}\right) = \frac{1}{MN}\sum_{t=1}^{M\times N}\frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2+\sigma_0^2}\xrightarrow{a.s.}V_1.$$

For the Lindeberg condition, we have for any $\delta > 0$,

$$\begin{aligned}&E\left[\frac{1}{MN}\sum_{t=1}^{M\times N}E\left\{\xi_t^2\mathbf{1}\left(|\xi_t|>\sqrt{MN}\delta\right)\middle|\,\mathcal{F}_{t-1}\right\}\right]\\ &\le \sup_t E\left\{\xi_t^2\mathbf{1}\left\{\frac{y_{(t-1)\Delta}^2 v_{1\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2+\sigma_0^2}>MN\delta^2\right\}\right\}\to 0,\end{aligned}$$

since ξ_t^2 is uniformly integrable. By the martingale CLT, as $T\to\infty$, we deduce that

$$\frac{1}{\sqrt{MN}}\sum_{t=1}^{M\times N}\xi_t\xrightarrow{L}\mathcal{N}\left(0,V_1\right),$$

which implies $\sqrt{\Delta}C_N = O_p(\sqrt{MN\Delta}) = O_p(\sqrt{T})$.

Similarly, applying martingale CLT to the first term of D_N and note that $\frac{y_{(t-1)\Delta}w_{t\Delta}}{\widetilde{\sigma}^2 y_{(t-1)\Delta}^2 + \sigma^2}$ is bounded, we then have, as $T\Delta \to 0$,

$$\begin{aligned} D_N &= O_p(\sqrt{MN}\Delta) + o_p(\sqrt{MN\Delta}) + O_p(MN\Delta^{3/2}) \\ &= O_p(\sqrt{T\Delta}) + o_p(\sqrt{T}) + O_p(\sqrt{T}\cdot\sqrt{T\Delta}) \\ &= o_p(\sqrt{T}). \end{aligned}$$

This leads to

$$\breve{\widetilde{\mu}} := \Delta^{-1}\frac{A_N}{B_N} = \widetilde{\mu}_0 + \frac{C_N}{\sqrt{\Delta}B_N} + o_p\left(\frac{1}{\sqrt{T}}\right). \tag{A.2.3}$$

and hence,

$$\sqrt{T}\left(\breve{\widetilde{\mu}} - \widetilde{\mu}_0\right) = \frac{\frac{1}{\sqrt{MN}}\sum_{t=1}^{M\times N}\frac{y_{(t-1)\Delta}v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2}}}{\frac{1}{MN}\sum_{t=1}^{M\times N}\frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2}} + o_p(1). \tag{A.2.4}$$

It follows from this result, (A.2.4) and (A.2.2) that

$$\sqrt{T}\left(\breve{\widetilde{\mu}} - \widetilde{\mu}_0\right) \xrightarrow{L} \mathcal{N}\left(0, V_1^{-1}\right).$$

Then, by the fact that $A_N = O_p(\Delta B_N)$, $R_A = O_p(\Delta R_B)$ and $B_N + R_B = O_p(T/\Delta)$, we know

$$\begin{aligned} \widehat{\widetilde{\mu}} &= \Delta^{-1}\frac{\widehat{A}_N}{\widehat{B}_N} = \Delta^{-1}\frac{A_N + R_A}{B_N + R_B} = \Delta^{-1}\frac{A_N}{B_N} + \Delta^{-1}\left(\frac{A_N + R_A}{B_N + R_B} - \frac{A_N}{B_N}\right) \\ &= \breve{\widetilde{\mu}} + \Delta^{-1}\frac{R_A B_N - R_B A_N}{B_N(B_N + R_B)} = \breve{\widetilde{\mu}} + O_p\left(\frac{R_B}{B_N + R_B}\right) = \breve{\widetilde{\mu}} + O_p(\sqrt{\Delta}). \end{aligned}$$

Therefore, by assuming $T\Delta \to 0$, we have

$$\sqrt{T}\left(\widehat{\widetilde{\mu}} - \widetilde{\mu}_0\right) = \sqrt{T}\left(\breve{\widetilde{\mu}} - \widetilde{\mu}_0\right) + \sqrt{T}\left(\widehat{\widetilde{\mu}} - \breve{\widetilde{\mu}}\right) \xrightarrow{L} \mathcal{N}\left(0, V_1^{-1}\right). \tag{A.2.5}$$

Consistency of $\widehat{\widetilde{\mu}}$ follows naturally from the results above.

A.2.3 Proof of Theorem 3.4.3

Similar to the previous proof, by equation (3.2.11) and consistency of $\widehat{\boldsymbol{\theta}}$, we have

$$\begin{aligned}
\widehat{\rho} - \rho_0 &= \frac{\sum_{t=1}^{T/\Delta} \frac{y_{(t-1)\Delta}(y_{t\Delta} - \rho_0 y_{(t-1)\Delta})}{\widehat{\widetilde{\sigma}^2} y_{(t-1)\Delta}^2 + \widehat{\sigma^2}}}{\sum_{t=1}^{T/\Delta} \frac{y_{(t-1)\Delta}^2}{\widehat{\widetilde{\sigma}^2} y_{(t-1)\Delta}^2 + \widehat{\sigma^2}}} \\
&= \frac{\sqrt{\Delta} \sum_{t=1}^{T/\Delta} \frac{y_{(t-1)\Delta} v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2}} + \sqrt{\Delta} \sum_{t=1}^{T/\Delta} y_{(t-1)\Delta} v_{t\Delta} \sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2} \zeta_{(t-1)\Delta}}{\sum_{t=1}^{T/\Delta} \frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2} + R_B} \\
&= \frac{\sqrt{\Delta} C_N + R_C}{B_N + R_B},
\end{aligned}$$

which leads to the decomposition

$$\frac{1}{\Delta}(\widehat{\rho} - \rho_0) = \frac{\sqrt{\Delta} C_N + R_C}{\Delta B_N + \Delta R_B},$$

where B_N and C_N are defined in the proof of Theorem 3.4.2, and

$$R_C := \sqrt{\Delta} \sum_{t=1}^{T/\Delta} y_{(t-1)\Delta} v_{t\Delta} \sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2} \zeta_{(t-1)\Delta} = O_p(\sqrt{\Delta}). \tag{A.2.6}$$

when T is finite.

In the previous proof, we have shown that $C_N = O_p(\sqrt{\Delta} B_N)$ and one can easily check that $R_C = O_p(\Delta R_B)$ given the order of R_B in the previous proof. Therefore, we have

$$\frac{1}{\Delta}(\widehat{\rho} - \rho_0) = \frac{\sqrt{\Delta} C_N + R_C}{\Delta B_N + \Delta R_B} = \frac{\sqrt{\Delta} C_N}{\Delta B_N} + O_p\left(\frac{R_B}{B_N + R_B}\right) = \frac{\sqrt{\Delta} C_N}{\Delta B_N} + O_p(\sqrt{\Delta}).$$

By the similar arguments to those in the proof of Theorem 3.4.2,

$$\frac{\Delta}{T}B_N = \frac{\Delta}{T}\sum_{t=1}^{T/\Delta}\frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2} \xrightarrow{p} V_1, \text{ i.e. } \Delta B_N \xrightarrow{p} TV_1.$$

Further, by the martingale CLT,

$$\sqrt{\frac{\Delta}{T}}C_N = \sqrt{\frac{\Delta}{T}}\sum_{t=1}^{T/\Delta}\frac{y_{(t-1)\Delta}v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + \sigma_0^2}} \xrightarrow{L} \mathcal{N}(0, V_1),$$

when $\Delta \to 0$. This is equivalent to $\sqrt{\Delta}C_N \xrightarrow{L} \mathcal{N}(0, TV_1)$. Combining these results gives

$$\frac{1}{\Delta}(\widehat{\rho} - \rho_0) = \frac{\sqrt{\Delta}\cdot C_N}{\Delta \cdot B_N} \xrightarrow{L} \mathcal{N}(0, (TV_1)^{-1}).$$

A.2.4 Proof of Proposition 3.4.1

Under the assumption that $\Delta \to 0$ with fixed T, we have

$$\begin{aligned}
\widehat{\beta}_\kappa - \beta_\kappa^0 &= \exp\left(\widehat{\widetilde{\mu}}\Delta + \frac{\widehat{\widetilde{\sigma}}^2\Delta}{2}\right) - \exp\left(\widetilde{\mu}_0\Delta + \frac{\widetilde{\sigma}_0^2\Delta}{2}\right) \\
&= \left(\widehat{\widetilde{\mu}}\Delta - \widetilde{\mu}_0\Delta\right) + \frac{1}{2}\left(\widehat{\widetilde{\sigma}}^2\Delta - \widetilde{\sigma}_0^2\Delta\right) + O_p(\Delta^2) \\
&= \left(\widehat{\widetilde{\mu}}\Delta - \widetilde{\mu}_0\Delta\right) + \frac{\Delta^{3/2}}{2}\left\{\frac{1}{\sqrt{\Delta}}\left(\widehat{\widetilde{\sigma}}^2 - \widetilde{\sigma}_0^2\right)\right\} + O_p(\Delta^2) \\
&= \left(\widehat{\widetilde{\mu}}\Delta - \widetilde{\mu}_0\Delta\right) + O_p(\Delta^{3/2}).
\end{aligned}$$

By Theorem 3.4.3,

$$\frac{1}{\Delta}(\widehat{\rho} - \rho_0) \xrightarrow{L} \mathcal{N}(0, (TV_1)^{-1}). \tag{A.2.7}$$

Then, by the Taylor expansion, we obtain

$$\widehat{\rho} - \rho_0 = \widehat{\widetilde{\mu}}\Delta - \widetilde{\mu}_0\Delta + O_p(\Delta^2). \tag{A.2.8}$$

Therefore, by Theorem 3.4.1 and 3.4.3 we have

$$\frac{1}{\Delta}\left(\widehat{\beta}_{\kappa}-\beta_{\kappa}^{0}\right)=\frac{1}{\Delta}\left(\widehat{\rho}-\rho_{0}\right)+O_{p}(\sqrt{\Delta})\xrightarrow{L}\mathcal{N}(0,(TV_{1})^{-1}). \tag{A.2.9}$$

The same argument yields the asymptotic result for $\widehat{\beta}_{\phi}$. Details of the proof are omitted.

A.2.5 Proof of Modified LBI Test Statistic $\widetilde{Z}_N$

Under the null, by Chan et al. (2012), we have the following asymptotic distribution result for $\widetilde{\rho}$,

$$\left(\sum_{t=1}^{M\times N}\frac{y_{(t-1)\Delta}^{2}}{\delta+y_{(t-1)\Delta}^{2}}\right)^{-1/2}\left(\sum_{t=1}^{M\times N}\frac{y_{(t-1)\Delta}^{2}}{(\delta+y_{(t-1)\Delta}^{2})^{1/2}}\right)\times(\widetilde{\rho}-\rho_{0})\xrightarrow{L}\mathcal{N}\left(0,\,Var(\varepsilon_{t\Delta})\right). \tag{A.2.10}$$

Then, we know for $y_{(t-1)\Delta}$, no matter it is stationary or nonstationary, we have

$$\widetilde{\varepsilon}_{t\Delta}-\varepsilon_{t\Delta}=(\widetilde{\rho}-\rho_{0})y_{(t-1)\Delta}=o_{p}(1).$$

Note $\widetilde{y}_{t\Delta}$ is always stationary, by WLLN and the ergodic theorem, we can easily show that, for any $p\in Z_{+}$ such that $p\leq 4$,

$$\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{\varepsilon}_{t\Delta}^{p}\xrightarrow{p}E(\varepsilon_{t\Delta}^{p}),\quad \frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{y}_{t\Delta}^{p}\xrightarrow{a.s.}E(\widetilde{y}_{t\Delta}^{p}).$$

Therefore, for the denominator, we have

$$\sqrt{\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{\varepsilon}_{t\Delta}^{4}-\left(\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{\varepsilon}_{t\Delta}^{2}\right)^{2}}\xrightarrow{p}\sqrt{E(\varepsilon_{t\Delta}^{4})-E(\varepsilon_{t\Delta}^{2})^{2}}=Std(\varepsilon_{t\Delta}^{2}),$$

$$\sqrt{\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{y}_{t\Delta}^{4}-\left(\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{y}_{t\Delta}^{2}\right)^{2}}\xrightarrow{a.s.}\sqrt{E(\widetilde{y}_{t\Delta}^{4})-E(\widetilde{y}_{t\Delta}^{2})^{2}}=Std(\widetilde{y}_{t\Delta}^{2}).$$

For the numerator, denote $\tilde{\xi}_{t\Delta}=\widetilde{\varepsilon}_{t\Delta}^{2}-\left(\frac{1}{MN}\sum_{t=1}^{M\times N}\widetilde{\varepsilon}_{t\Delta}^{2}\right)$ and $\xi_{t\Delta}=\varepsilon_{t\Delta}^{2}-$

$\left(\frac{1}{MN}\sum_{t=1}^{M\times N}\varepsilon_{t\Delta}^2\right)$. We know $E(\tilde{\xi}_{t\Delta})=0=E(\xi_{t\Delta})$ and $Var(\tilde{\xi}_{t\Delta})=Var(\widetilde{\varepsilon}_{t\Delta}^2)\xrightarrow{p}Var(\varepsilon_{t\Delta}^2)$.

$$\frac{1}{\sqrt{MN}}\sum_{t=1}^{M\times N}\widetilde{y}_{(t-1)\Delta}^2\tilde{\xi}_{t\Delta}=\frac{1}{\sqrt{MN}}\sum_{t=1}^{M\times N}\widetilde{y}_{(t-1)\Delta}^2(\tilde{\xi}_{t\Delta}-\xi_{t\Delta})+\frac{1}{\sqrt{MN}}\sum_{t=1}^{M\times N}\widetilde{y}_{(t-1)\Delta}^2\xi_{t\Delta}.$$

By equation (3.3) in Lee (1998), one can easily show

$$\frac{1}{\sqrt{MN}}\sum_{t=1}^{M\times N}\widetilde{y}_{(t-1)\Delta}^2(\tilde{\xi}_{t\Delta}-\xi_{t\Delta})=o_p(1),$$

and by applying the martingale CLT (cf. Hall and Heyde, 1980), we have

$$\frac{1}{\sqrt{MN}}\sum_{t=1}^{M\times N}\widetilde{y}_{(t-1)\Delta}^2\xi_{t\Delta}\xrightarrow{L}\mathcal{N}\left(0,\,Var(\widetilde{y}_{t\Delta}^2)\,Var(\varepsilon_{t\Delta}^2)\right).$$

Then, by combining the results above, we can derive the asymptotic distribution of $\widetilde{Z}_N$ under $H_0:\widetilde{\sigma}_0^2=0$, i.e.,

$$\frac{1}{\sqrt{MN}}\widetilde{Z}_N\xrightarrow{L}\mathcal{N}(0,1).$$

Lastly, under the alternative, one just need to note that $Cov(\varepsilon_{t\Delta}^2,y_{(t-1)\Delta}^2)$ diverges when $\widetilde{\sigma}_0^2\neq 0$, and this leads to the divergence of $\widetilde{Z}_N$.

A.2.6 Proof of Proposition 3.5.1

It has been proved in Föllmer et al. (1994) that y_t is strictly stationary and ergodic when $\widetilde{\mu}-\frac{1}{2}\widetilde{\sigma}^2<0$. This means that we can still characterize explosiveness using $\phi:=\widetilde{\mu}-\frac{1}{2}\widetilde{\sigma}^2$. However, for characterizing instability, we need to calculate the second moment of y_t.

The expectation of $J(t)$ is

$$\begin{aligned} EJ(t) &= -\omega \int_0^t E\left(\exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(t-s) + \widetilde{\sigma}\left(B_u(t) - B_u(s)\right)\right\}\right) ds \\ &= -\omega \int_0^t \exp\left\{\widetilde{\mu}(t-s)\right\} ds = \frac{\omega}{\widetilde{\mu}}(1 - \exp(\widetilde{\mu}t)), \end{aligned}$$

and so $EJ(t)$ is finite as $t \to \infty$ if and only if $\widetilde{\mu} < 0$. Further, to figure out the order of $Var(J(t))$, we apply the Cauchy-Schwartz inequality to $EJ(t)^2$, giving

$$EJ(t)^2 = E[K(t) - L(t)]^2 \leq 2EK(t)^2 + 2EL(t)^2$$

This inequality indicate that we only need to calculate $EK(t)^2$ and $EL(t)^2$ to evaluate the asymptotic order of $EJ(t)^2$. By Itō's isometry

$$EK(t)^2 = \sigma^2 \int_0^t E\left(\exp\left\{\left(2\widetilde{\mu} - \widetilde{\sigma}^2\right)(t-s) + 2\widetilde{\sigma}\left(B_u(t) - B_u(s)\right)\right\}\right) ds = \sigma^2 \frac{e^{2\kappa t} - 1}{2\kappa},$$

$$\begin{aligned} EL(t)^2 &= \omega^2 E\left(\int_0^t \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(t-s) + \widetilde{\sigma}\left(B_u(t) - B_u(s)\right)\right\} ds\right)^2 \\ &= \omega^2 E\left(\int_0^t \int_0^t \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(2t - s - r) + \widetilde{\sigma}\left(2B_u(t) - B_u(s) - B_u(r)\right)\right\} dsdr\right) \\ &= \omega^2 \int_0^t \int_0^t \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(2t - s - r) + \frac{1}{2}\widetilde{\sigma}^2\left(2t - s - r + 2\min\{t-s, t-r\}\right)\right\} dsdr \\ &= \omega^2 \int_0^t \int_0^r \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(2t - s - r) + \frac{1}{2}\widetilde{\sigma}^2\left(4t - s - 3r\right)\right\} dsdr \\ &\quad + \omega^2 \int_0^t \int_r^t \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(2t - s - r) + \frac{1}{2}\widetilde{\sigma}^2\left(4t - 3s - r\right)\right\} dsdr \\ &= \omega^2 \left(\int_0^t \frac{(e^{\widetilde{\mu}r} - 1)e^{2\kappa(t-r)}}{\widetilde{\mu}} dr + \int_0^t \frac{e^{\widetilde{\mu}(t-r)}(e^{(\widetilde{\mu}+\widetilde{\sigma}^2)(t-r)} - 1)}{\widetilde{\mu} + \widetilde{\sigma}^2} dr\right) \\ &= \omega^2 \left(\frac{\widetilde{\mu}(e^{2\kappa t} - 2e^{\widetilde{\mu}t} + 1) - \widetilde{\sigma}^2(e^{\widetilde{\mu}t} - 1)}{2\widetilde{\mu}\kappa(\widetilde{\mu} + \widetilde{\sigma}^2)} + \frac{1 - e^{\widetilde{\mu}t}}{\widetilde{\mu}(\widetilde{\mu} + \widetilde{\sigma}^2)} + \frac{e^{2\kappa t} - 1}{2\kappa(\widetilde{\mu} + \widetilde{\sigma}^2)}\right) \\ &= \omega^2 \left(\frac{\widetilde{\mu}e^{2\kappa t} - 2\kappa e^{\widetilde{\mu}t} + \widetilde{\mu} + \widetilde{\sigma}^2}{\widetilde{\mu}\kappa(\widetilde{\mu} + \widetilde{\sigma}^2)}\right) \end{aligned}$$

Note that for $\kappa < 0$

$$\lim_{t\to\infty} EJ(t)^2 \leq \lim_{t\to\infty} 2EK(t)^2 + 2EL(t)^2 = \frac{2\omega^2}{\widetilde{\mu}\kappa} - \frac{\sigma^2}{\kappa} < \infty, \tag{A.2.11}$$

showing that, when $\kappa < 0$, $J(t)$ has finite second-order moments as $t \to \infty$. Further, for $\kappa \to 0$, by L'Hôpital's rule, we have

$$\lim_{\kappa\to 0} EK(t)^2 = \lim_{\kappa\to 0} \sigma^2 \frac{2te^{2\kappa t}}{2} = \sigma^2 t, \tag{A.2.12}$$

$$\begin{aligned}\lim_{\kappa\to 0} EL(t)^2 &= \lim_{\widetilde{\mu}\to -\widetilde{\sigma}^2/2} \omega^2 \left(\frac{\widetilde{\mu}e^{2\kappa t} - 2\kappa e^{\widetilde{\mu}t} + \widetilde{\mu} + \widetilde{\sigma}^2}{\widetilde{\mu}\kappa(\widetilde{\mu} + \widetilde{\sigma}^2)}\right) \\ &= 4\gamma^2\sigma^2 t + \frac{2(e^{-\frac{1}{2}\widetilde{\sigma}^2 t} - 1)}{\widetilde{\sigma}^2}.\end{aligned} \tag{A.2.13}$$

Combining results (A.2.12) and (A.2.13), we obtain $EJ(t)^2 = O_p(t)$, as $t \to \infty$.

Lastly, for $\kappa > 0$, note that $\kappa = \widetilde{\mu} + \frac{1}{2}\widetilde{\sigma}^2 \geq \widetilde{\mu}$, we then have

$$EK(t)^2 = \sigma^2 \frac{e^{2\kappa t} - 1}{2\kappa} = O_p(e^{2\kappa t}) \tag{A.2.14}$$

$$EL(t)^2 = \omega^2 \left(\frac{\widetilde{\mu}e^{2\kappa t} - 2\kappa e^{\widetilde{\mu}t} + \widetilde{\mu} + \widetilde{\sigma}^2}{\widetilde{\mu}\kappa(\widetilde{\mu} + \widetilde{\sigma}^2)}\right) = O_p(e^{2\kappa t}) \tag{A.2.15}$$

which leads to $EJ(t)^2 = O_p(e^{2\kappa t})$.

From the results above we know that $EJ(t)$ is finite if and only if $\widetilde{\mu} < 0$, and $EJ(t)^2$ is finite if and only if $\kappa < 0$. Since $\kappa < 0$ implies $\widetilde{\mu} < 0$, $Var(J(t)) < \infty$ if and only if $\kappa < 0$. We can now work out the first two moments of y_t,

$$Ey_t = E\left[\exp\left(\widetilde{\sigma}B_u(t) + \left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)t\right)\right] y_0 + EJ(t) = e^{\widetilde{\mu}t}y_0 + \frac{\omega}{\widetilde{\mu}}(1 - \exp(\widetilde{\mu}t)),$$

$$
\begin{aligned}
Ey_t^2 &= E\left[\exp\left(2\widetilde{\sigma}B_u(t) + \left(2\widetilde{\mu} - \widetilde{\sigma}^2\right)t\right)\right] Ey_0^2 \\
&\quad + 2Ey_0 E\left[\exp\left(\widetilde{\sigma}B_u(t) + \left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)t\right) J(t)\right] + EJ(t)^2 \\
&= e^{2\kappa t}Ey_0^2 - 2Ey_0 E\left(\omega\int_0^t \exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(2t-s) + \widetilde{\sigma}\left(2B_u(t) - B_u(s)\right)\right\} ds\right) + EJ(t)^2 \\
&= e^{2\kappa t}Ey_0^2 - 2\omega\int_0^t E\left(\exp\left\{\left(\widetilde{\mu} - \frac{1}{2}\widetilde{\sigma}^2\right)(2t-s) + \widetilde{\sigma}\left(2B_u(t) - B_u(s)\right)\right\}\right) ds Ey_0 + EJ(t)^2 \\
&= e^{2\kappa t}y_0^2 - 2\omega\int_0^t \exp\left(2\kappa t - \left(\widetilde{\mu} + \widetilde{\sigma}^2\right)s\right) ds Ey_0 + EJ(t)^2 \\
&= e^{2\kappa t}Ey_0^2 - 2\omega\frac{e^{2\kappa t} - e^{\widetilde{\mu}t}}{\widetilde{\mu} + \widetilde{\sigma}^2} Ey_0 + EJ(t)^2.
\end{aligned}
$$

Evidently from these expressions Ey_t is asymptotically finite if and only if $\widetilde{\mu} < 0$, and Ey_t^2 is asymptotically finite if and only if $\kappa < 0$. This indicates that $Var(y_t) < \infty$ if and only if $\kappa < 0$. Therefore, we can still characterize instability with $\kappa \geq 0$ and locally explosiveness with $\widetilde{\mu} \geq 0$.

A.2.7 Proof of Remark 3.5.2

Denote $X_n = \dfrac{\log RV_n - \log[y]_{(n-1)h}^{nh} + \frac{1}{2}s_n^2}{s_n}$, where $s_n = \max\left\{\sqrt{2\Delta\dfrac{RQ_n}{RV_n^2}}, \sqrt{\dfrac{2}{M}}\right\}$. According to Barndorff-Nielsen and Shephard (2005), $\{X_n\}_{n=1}^N \xrightarrow{L} \mathcal{N}(0,1)$ as $\Delta \to 0$. Note $N = \frac{T}{M\Delta}$, so when $\Delta \to 0$ with T, M being finite, we have $N \to \infty$. Therefore, the log-likelihood function for $\theta = (\widetilde{\sigma}^2, \gamma, \sigma^2)$ is given by

$$
\ell_{ur}(\theta) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\sum_{n=1}^N X_n(\theta)^2 + O(\Delta). \tag{A.2.16}
$$

As $\ell_{ur}(\theta)$ is based on the standard normal distribution, Wilks's theorem applies in this case, i.e. under $\mathcal{H}_0 : \gamma_0 = 0$, as $N \to \infty$,

$$
\begin{aligned}
LR = -2\left(\ell_r - \ell_{ur}\right) &= \sum_{n=1}^N X_n(\theta_0)^2 - \sum_{n=1}^N X_n(\theta)^2 + o_p(1) \\
&= \Delta^{-1}\left(Q_\Delta(\theta_0) - Q_\Delta(\theta)\right) + o_p(1) \xrightarrow{L} \chi^2(1),
\end{aligned}
$$

where $\theta_0 = (\widetilde{\sigma}_0^2, \gamma_0, \sigma_0^2)$.

A.2.8 Proof of Theorem 3.5.2

The dependence of B_u and B_ε leads to a complex relationship among $y_{(t-1)\Delta}$, $\rho_{t\Delta}$ and $J_{t\Delta}$ in model (3.5.4). Without loss of generality, we know that y_t can also be viewed as generated from model (3.5.3) by virtue of the observational equivalence of these mechanisms. Note that the approximate representation of the exact discretized model of (3.5.3) is

$$y_{t\Delta} - y_{(t-1)\Delta} = \widetilde{\mu}_0 \Delta y_{(t-1)\Delta} \cdot + \sqrt{(\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + 2\omega_0 y_{(t-1)\Delta} + \sigma_0^2)\Delta} \cdot v_{t\Delta} + \widetilde{\varsigma}_{t\Delta}, \tag{A.2.17}$$

where $v_{t\Delta} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$ and $\widetilde{\varsigma}_{t\Delta}$ is the discretization error.

Similar to the proof of Theorem 3.4.2, the limiting distribution of the feasible estimator will converge to that of the infeasible estimator given $T\Delta \to 0$. Therefore, we only need to figure out the distribution of the infeasible estimator. According to (3.5.10) and with a similar derivation of the order the discretization error in the proof of Theorem 3.4.2, we have

$$\begin{aligned} A_N^* &= \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}\left(y_{t\Delta} - y_{(t-1)\Delta}\right)}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + 2\omega_0 y_{(t-1)\Delta} + \sigma_0^2} \\ &= \widetilde{\mu}_0 \Delta \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + 2\omega_0 y_{(t-1)} + \sigma_0^2} + \sqrt{\Delta} \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta} v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + 2\omega_0 y_{(t-1)\Delta} + \sigma_0^2}} + o_p(\sqrt{T}) \\ &= \widetilde{\mu}_0 \Delta B_N^* + \sqrt{\Delta} C_N^* + o_p(\sqrt{T}), \end{aligned}$$

where $C_N^* = \sum_{t=1}^{M\times N} \frac{y_{(t-1)\Delta} v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + 2\omega_0 y_{(t-1)\Delta} + \sigma_0^2}}$. This leads to

$$\breve{\widetilde{\mu}} = \widetilde{\mu}_0 + \frac{C_N^*}{\sqrt{\Delta} B_N^*} + o_p\left(\frac{1}{\sqrt{T}}\right), \tag{A.2.18}$$

and hence,

$$\sqrt{T}\left(\breve{\widetilde{\mu}}-\widetilde{\mu}_0\right)=\frac{\frac{1}{\sqrt{MN}}\sum_{t=1}^{M\times N}\frac{y_{(t-1)\Delta}v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2y_{(t-1)\Delta}^2+2\omega_0y_{(t-1)\Delta}+\sigma_0^2}}}{\frac{1}{MN}\sum_{t=1}^{M\times N}\frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2y_{(t-1)\Delta}^2+2\omega_0y_{(t-1)\Delta}+\sigma_0^2}}+o_p(1). \quad \text{(A.2.19)}$$

Note that $\frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2y_{(t-1)\Delta}^2+2\omega_0y_{(t-1)\Delta}+\sigma_0^2}$ is bounded above by $\widetilde{\sigma}_0^{-2}$. Similarly, by the same argument as in the proof of Theorem 3.4.2, we have

$$\frac{1}{MN}\sum_{t=1}^{M\times N}\frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2y_{(t-1)\Delta}^2+2\omega_0y_{(t-1)\Delta}+\sigma_0^2}\xrightarrow{p}V_2,$$

where V_2 is defined in (3.5.12).

Further, denote

$$\xi_t:=\frac{y_{(t-1)\Delta}v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2y_{(t-1)\Delta}^2+2\omega_0y_{(t-1)\Delta}+\sigma_0^2}}$$

and observe that ξ_t is a martingale difference sequence with respect to the filtration $\mathcal{F}_t:=\sigma(B_v(s):0\le s\le t)$ as

$$\begin{aligned}E\left(\xi_t|\mathcal{F}_{t-1}\right)&=E\left(\frac{y_{(t-1)\Delta}v_{t\Delta}}{\sqrt{\widetilde{\sigma}_0^2y_{(t-1)\Delta}^2+2\omega_0y_{(t-1)\Delta}+\sigma_0^2}}\middle|\mathcal{F}_{t-1}\right)\\&=\frac{y_{(t-1)\Delta}}{\sqrt{\widetilde{\sigma}_0^2y_{(t-1)\Delta}^2+2\omega_0y_{(t-1)\Delta}+\sigma_0^2}}E\left(v_{t\Delta}|\,\mathcal{F}_{t-1}\right)=0.\end{aligned}$$

To apply the martingale CLT, we check the stability condition and the Lindeberg condition. For the stability condition, we have

$$\left\langle\frac{1}{\sqrt{MN}}\sum_{t=1}^{M\times N}\xi_t\right\rangle=\frac{1}{MN}\sum_{t=1}^{M\times N}E\left(\frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2y_{(t-1)\Delta}^2+2\omega_0y_{(t-1)\Delta}+\sigma_0^2}\middle|\mathcal{F}_{t-1}\right)\xrightarrow{a.s.}V_2.$$

For the Lindeberg condition, we have for any $\delta > 0$

$$\frac{1}{MN}\sum_{t=1}^{M\times N} E\left\{\xi_t^2 \mathbf{1}\left(|\xi_t| > \sqrt{MN}\delta\right)\middle|\mathcal{F}_{t-1}\right\}$$
$$\leq \sup_t E\left\{\xi_t^2 \mathbf{1}\left\{\frac{y_{(t-1)\Delta}^2 v_{1\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + 2\omega_0 y_{(t-1)\Delta} + \sigma_0^2} > MN\delta^2\right\}\middle|\mathcal{F}_{t-1}\right\} \to 0,$$

since ξ_t^2 is uniformly integrable and $MN \to \infty$. From the martingale CLT, as $T \to \infty$,

$$\sqrt{T}\left(\breve{\widetilde{\mu}} - \widetilde{\mu}_0\right) \xrightarrow{L} \mathcal{N}\left(0, V_2^{-1}\right),$$

and hence,

$$\sqrt{T}\left(\widehat{\widetilde{\mu}} - \widetilde{\mu}_0\right) \xrightarrow{L} \mathcal{N}\left(0, V_2^{-1}\right).$$

Consistency of $\widehat{\widetilde{\mu}}$ follows naturally from the results above.

A.2.9 Proof of Theorem 3.5.3

Similar to the proof of Theorem 3.4.3 by substituting equation (A.2.17) into $\widehat{\rho}$, we obtain

$$\widehat{\rho} - \rho_0 = \widehat{\widetilde{\mu}}\Delta - \widetilde{\mu}_0\Delta + o_p(\Delta) = \frac{A_N^*}{B_N^*} - \widetilde{\mu}_0\Delta + o_p(\Delta) = \frac{\sqrt{\Delta}C_N^*}{B_N^*} + o_p(\Delta).$$

Then, by the same arguments as those in the proof of Theorem 3.4.2, we have

$$\frac{\Delta}{T}B_N^* = \frac{\Delta}{T}\sum_{t=1}^{T/\Delta}\frac{y_{(t-1)\Delta}^2}{\widetilde{\sigma}_0^2 y_{(t-1)\Delta}^2 + 2\omega_0 y_{(t-1)\Delta} + \sigma_0^2} \xrightarrow{p} V_2, \text{ i.e. } \Delta B_N^* \xrightarrow{p} TV_2,$$

Further, as proved in the previous section, $\sqrt{\Delta/T}C_N^* \xrightarrow{L} \mathcal{N}(0, V_2)$ by the martingale CLT, which gives $\sqrt{\Delta}C_N^* \xrightarrow{L} \mathcal{N}(0, TV_2)$. Combining these results gives

$$\frac{1}{\Delta}\left(\widehat{\rho} - \rho_0\right) = \frac{\sqrt{\Delta}C_N^*}{\Delta B_N^*} + o_p(1) \xrightarrow{L} \mathcal{N}(0, (TV_2)^{-1}).$$

A.3 Proof of Chapter 4

The notations that have been frequently used in the proofs are as follows: $[n] \stackrel{\text{def}}{=} \{1, 2, \cdots, n\}$ for any positive integer n, $\mathcal{M}_{m,n}$ be the set of all $m \times n$ matrices which have exactly one 1 and $n-1$ 0's in each row. $\mathbb{R}^{m\times n}$ denotes the set of all $m \times n$ real matrices. $\|\cdot\|$ is used to denote Euclidean ℓ_2-norm for vectors in $\mathbb{R}^{m\times 1}$ and the spectral norm for matrices on $\mathbb{R}^{m\times n}$. $\|\cdot\|_\infty$ denotes the largest element of the matrix in absolute value. $\|\cdot\|_F$ is the Frobenius norm on $\mathbb{R}^{m\times n}$, namely $\|M\|_F \stackrel{\text{def}}{=} \sqrt{\text{tr}(M^\top M)}$. $\|\cdot\|_{\phi_2}$ is the sub-Gaussian norm such that for any random variable x, there is $\|x\|_{\phi_2} \stackrel{\text{def}}{=} \sup_{\kappa\geq 1} \kappa^{-1/2}(E|x|^\kappa)^{1/\kappa}$. $\mathbf{1}_{m,n} \in \mathbb{R}^{m\times n}$ consists of all 1's, ι_n denotes the column vector with n elements of all 1's. $\mathbb{1}_A$ denotes the indicator function of the event A.

A.3.1 Preliminary Lemmas

Lemma A.3.1 *Suppose A_t and X are the adjacency matrix and the node covariate matrices sampled from the SC-DCBM/SC-DCcBM. Recall W_t and $\mathcal{W}_t$ are empirical and population weight matrices. Then, we have*

$$\sup_t \|W_t - \mathcal{W}_t\|_\infty = \mathcal{O}_p(\xi),$$

where $\xi = \max(\sigma^2\|L_\tau\|_F\sqrt{\log(TR)}, \sigma^2\|L_\tau\|\log(TR), NRJ^2/\underline{\delta})$ *and* $\underline{\delta} = \inf_t\{\min_i \mathcal{D}_{\tau,t}(i,i)\}$.

Proof. Define $\mathcal{I}_t = \mathcal{X}L_{\tau,t}\mathcal{X}$. Then we have

$$\sup_t \|W_t - \mathcal{W}_t\|_\infty \leq \sup_t \|W_t - \mathcal{I}_t\|_\infty + \sup_t \|\mathcal{I}_t - \mathcal{W}_t\|_\infty.$$

For the first part, define $L_\tau = \sup_t L_{\tau,t}$ and $\zeta = \max(\sigma^2\|L_\tau\|_F\sqrt{\log(TR)}, \sigma^2\|L_\tau\|\log(TR))$, then by Hansen-Wright inequality (c.f., Theorem 1.1 of Rudelson and Vershynin

(2013)), we have

$$\begin{aligned}\Pr(\sup_t \|X^\top L_{\tau,t} X - \mathcal{X}^\top L_{\tau,t}\mathcal{X}\| > \zeta) &\leq \sum_{t=1}^{T} \Pr(\|X^\top L_\tau X - \mathcal{X}^\top L_\tau \mathcal{X}\| > \zeta) \\ &\leq 2T \exp\left\{-c \min\left(\frac{\zeta^2}{\sigma^4 \|L_\tau\|_F^2}, \frac{\zeta}{\sigma^2 \|L_\tau\|}\right)\right\} \\ &= \mathcal{O}(1/R).\end{aligned}$$

Next, denote $\mathcal{C}_t = \mathcal{D}_{\tau,t}^{-1/2} A_t \mathcal{D}_{\tau,t}^{-1/2}$, then we can decompose the second part into two parts:

$$\sup_t \|\mathcal{I}_t - \mathcal{W}_t\|_\infty = \sup_t \|\mathcal{X}(L_{\tau,t} - \mathcal{L}_{\tau,t})\mathcal{X}\|_\infty \leq \sup_t \|\mathcal{X}(L_{\tau,t} - \mathcal{C}_t)\mathcal{X}\|_\infty + \sup_t \|\mathcal{X}(\mathcal{C}_t - \mathcal{L}_{\tau,t})\mathcal{X}\|_\infty.$$

Then, for part one, we have

$$\begin{aligned}&\sup_t \|\mathcal{X}(L_{\tau,t} - \mathcal{C}_t)\mathcal{X}\|_\infty \\ &= \sup_t \max_{s,r} \left| \sum_{i,j} \mathcal{X}_{is}\mathcal{X}_{jr} \frac{A_t(i,j)}{\sqrt{\mathcal{D}_{\tau,t}(i,i)\mathcal{D}_{\tau,t}(j,j)}} \left(\frac{\sqrt{\mathcal{D}_{\tau,t}(i,i)\mathcal{D}_{\tau,t}(j,j)}}{\sqrt{D_{\tau,t}(i,i)D_{\tau,t}(j,j)}} - 1 \right) \right| \\ &\leq \frac{1}{\delta} \max_{s,r} \sum_{i,j} |\mathcal{X}_{is}\mathcal{X}_{jr}| \sup_t \left\{ \max\left(\left| \frac{\mathcal{D}_{\tau,t}(i,i)}{D_{\tau,t}(i,i)} - 1 \right|, \left| \frac{\mathcal{D}_{\tau,t}(j,j)}{D_{\tau,t}(j,j)} - 1 \right| \right) \right\} \\ &= \max_{s,r} \sum_{i,j} |\mathcal{X}_{is}\mathcal{X}_{jr}| \mathcal{O}_p(\underline{\delta}^{-3/2} \log(TR)) \\ &= \mathcal{O}_p\left(\frac{NRJ^2}{\underline{\delta}^{3/2}} \log(TR) \right),\end{aligned}$$

where the second to the last equality comes from the following proof. For any $i \in \{1, \cdots, N\}$ and $\varsigma = \underline{\delta}^{-1/2} \log(TR)$, from Bernstein inequality,

$$\begin{aligned}\Pr\left(\sup_t \left| \frac{D_{\tau,t}(i,i)}{\mathcal{D}_{\tau,t}(i,i)} - 1 \right| > \varsigma \right) &\leq \sum_{t=1}^{T} \Pr\left(\left| \frac{D_{\tau,t}(i,i)}{\mathcal{D}_{\tau,t}(i,i)} - 1 \right| > \varsigma \right) \\ &\leq 2T \exp\left\{ -\frac{\varsigma^2 \mathcal{D}_{\tau,t}(i,i)}{2 + \frac{2}{3}\varsigma} \right\} \\ &\leq 2T \exp\left\{ -\frac{\varsigma^2 \underline{\delta}}{2 + \frac{2}{3}\varsigma} \right\} \\ &= \mathcal{O}(1/R).\end{aligned}$$

For part two, similarly, we have

$$\begin{aligned}\sup_t \|\mathcal{X}(\mathcal{C}_t - \mathcal{L}_{\tau,t})\mathcal{X}\|_\infty &= \sup_t \max_{s,r} \left| \sum_{i,j} \mathcal{X}_{is}\mathcal{X}_{jr} \frac{A_t(i,j) - \mathcal{A}_t(i,j)}{\sqrt{\mathcal{D}_{\tau,t}(i,i)\mathcal{D}_{\tau,t}(j,j)}} \right| \\ &\leq \max_{s,r} \left| \sum_{i,j} \mathcal{X}_{is}\mathcal{X}_{jr} \right| \sup_t \max_{i,j} \left| \frac{A_t(i,j) - \mathcal{A}_t(i,j)}{\sqrt{\mathcal{D}_{\tau,t}(i,i)\mathcal{D}_{\tau,t}(j,j)}} \right| \\ &= \mathcal{O}_p\left(\frac{NRJ^2}{\underline{\delta}}\right).\end{aligned}$$

Note that $\varsigma \to 0$ as $\underline{\delta}, R \to \infty$, we then know

$$\sup_t \|\mathcal{I}_t - \mathcal{W}_t\|_\infty = \mathcal{O}_p\left(\frac{NRJ^2}{\underline{\delta}}\right).$$

Thus, by union bounds, we obtain

$$\sup_t \|W_t - \mathcal{W}_t\|_\infty = \mathcal{O}_p\left(\zeta + \frac{NRJ^2}{\underline{\delta}}\right) = \mathcal{O}_p(\xi).$$

■

Lemma A.3.2 *Under Assumption 4, for any $\epsilon > 0$, we have*

$$\sup_t \|S_t - \mathcal{S}_t\| \leq (4 + c_w)\left\{\frac{3\log(8NT/\epsilon)}{\underline{\delta}}\right\}^{1/2}, \tag{A.3.1}$$

with probability at least $1 - \epsilon$.

Proof. Note by triangular inequality, we have

$$\sup_t \|S_t - \mathcal{S}_t\| \leq \sup_t \left\|\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right\| \tag{A.3.2}$$

$$+ \sup_t \left\|\mathcal{D}_{\tau,t}^{-1/2} A_t \mathcal{D}_{\tau,t}^{-1/2} - \mathcal{D}_{\tau,t}^{-1/2}\mathcal{A}_t\mathcal{D}_{\tau,t}^{-1/2}\right\| \tag{A.3.3}$$

$$+ \sup_t \left\|D_{\tau,t}^{-1/2} A_t D_{\tau,t}^{-1/2} - \mathcal{D}_{\tau,t}^{-1/2} A_t \mathcal{D}_{\tau,t}^{-1/2}\right\|. \tag{A.3.4}$$

For equation (A.3.2), we have,

$$\begin{aligned}
&\sup_t \left\|\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right\| \\
&= \sup_t \left\|\alpha_t X(W_t - \mathcal{W}_t)X^\top\right\| + \sup_t \left\|\alpha_t X\mathcal{W}_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right\| \\
&\leq \alpha_{\max} NRJ^2 \sup_t \|W_t - \mathcal{W}_t\| + 2\alpha_{\max} NRJ^2 \sup_t \|\mathcal{W}_t\| \\
&= \mathcal{O}_p(\alpha_{\max} NRJ^2\xi).
\end{aligned}$$

So, by Assumption 4 we know, for large enough N, with probability at least $1-\epsilon/2$,

$$\sup_t \left\|\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right\| \leq c_w a$$

For equation (A.3.3), let $Y_t(i,j) = \mathcal{D}_{\tau,t}^{-1/2}[(A_t(i,j) - p_t(i,j))E_{ij}]\mathcal{D}_{\tau,t}^{-1/2}$ with $E_{ij} \in \mathbb{R}^{N\times N}$ being the matrix with 1 in ij and ji'th positions and 0 everywhere else. Then we know

$$\sup_t \|Y_t(i,j)\| \leq \sup_t \sqrt{\mathcal{D}_{\tau,t}(i,i)\mathcal{D}_{\tau,t}(j,j)} \leq \frac{1}{\underline{\delta}}, \quad v^2 = \sup_t \|\sum E(Y_t^2(i,j))\| \leq \frac{1}{\underline{\delta}}.$$

So, denote $a = \left\{\dfrac{3\log(8NT/\epsilon)}{\underline{\delta}}\right\}^{1/2}$, which is smaller than 1 by assumption, and by matrix Bernstein inequality, we have

$$\begin{aligned}
&\Pr(\sup_t \|\mathcal{D}_{\tau,t}^{-1/2}[A_t(i,j) - \mathcal{A}_t(i,j)]\mathcal{D}_{\tau,t}^{-1/2}\| > a) \\
&\leq \sum_{t=1}^{T} \Pr(\|\mathcal{D}_{\tau,t}^{-1/2}[A_t(i,j) - \mathcal{A}_t(i,j)]\mathcal{D}_{\tau,t}^{-1/2}\| > a) \\
&\leq 2NT\exp\left(-\frac{a^2}{2/\underline{\delta} + 2a/3\underline{\delta}}\right) \\
&\leq 2NT\exp\left(-\frac{3\log(8NT/\epsilon)}{3}\right) \\
&= \epsilon/4.
\end{aligned}$$

Hence, with probability at least $1 - \epsilon/4$,

$$\sup_t \|\mathcal{D}_{\tau,t}^{-1/2} A_t \mathcal{D}_{\tau,t}^{-1/2} - \mathcal{D}_{\tau,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{\tau,t}^{-1/2}\| \leq a \tag{A.3.5}$$

Lastly, for equation (A.3.4), by Qin and Rohe (2013) and setting $\lambda = a\mathcal{D}_{\tau,t}(i,i)$ we have

$$\begin{aligned}
\Pr(|D_{\tau,t}(i,i) - \mathcal{D}_{\tau,t}(i,i)| \geq \lambda) &\leq \exp\left\{-\frac{\lambda^2}{2\mathcal{D}_{\tau,t}(i,i)}\right\} + \exp\left\{-\frac{\lambda^2}{2\mathcal{D}_{\tau,t}(i,i) + \frac{2}{3}\lambda}\right\} \\
&\leq 2\exp\left\{-\frac{\lambda^2}{2\mathcal{D}_{\tau,t}(i,i) + \frac{2}{3}\lambda}\right\} \\
&= 2\exp\left\{-\frac{a^2\mathcal{D}_{\tau,t}(i,i)}{2 + \frac{2}{3}a}\right\} \\
&\leq 2\exp\left\{-\log(8NT/\epsilon) \times \frac{\mathcal{D}_{\tau,t}(i,i)}{\underline{\delta}}\right\} \\
&\leq \frac{\epsilon}{4NT}.
\end{aligned}$$

Further note that

$$\begin{aligned}
\Pr\left(\sup_t \|\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\| \geq a\right) &\leq \sum_{t=1}^{T} \Pr\left(\|\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\| \geq a\right) \\
&\leq \sum_{t=1}^{T} \Pr\left(\max_i \left|\frac{D_{\tau,t}(i,i)}{\mathcal{D}_{\tau,t}(i,i)} - 1\right| \geq a\right) \\
&\leq \sum_{t=1}^{T}\sum_{i=1}^{N} \Pr\left(|D_{\tau,t}(i,i) - \mathcal{D}_{\tau,t}(i,i)| \geq a\mathcal{D}_{\tau,t}(i,i)\right) \\
&\leq NT \times \frac{\epsilon}{4NT} \\
&= \epsilon/4.
\end{aligned}$$

Therefore, with probability at least $1-\epsilon/4$, we have

$$\begin{aligned}
&\sup_t \|D_{\tau,t}^{-1/2} A_t D_{\tau,t}^{-1/2} - \mathcal{D}_{\tau,t}^{-1/2} A_t \mathcal{D}_{\tau,t}^{-1/2}\| \\
&\quad = \sup_t \|L_{\tau,t} - \mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} L_{\tau,t} D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2}\| \\
&\quad = \sup_t \|(I - \mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2}) L_{\tau,t} D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2} + L_{\tau,t}(I - D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2})\| \\
&\quad \leq \sup_t \|\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\| \sup_t \|\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2}\| + \sup_t \|\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\| \\
&\quad \leq a^2 + 2a
\end{aligned}$$

where the second last inequality comes from the fact that $\sup_t \|L_{\tau,t}\| \leq 1$.

Therefore, joining the results for these three equations, we have, with probability at least $1-\epsilon$,

$$\sup_t \|S_t - \mathcal{S}_t\| \leq a^2 + 3a + c_w a \leq (4 + c_w) a = (4 + c_w) \left\{ \frac{3 \log(8NT/\epsilon)}{\underline{\delta}} \right\}^{1/2}. \tag{A.3.6}$$

■

Lemma A.3.3 *Under the dynamic SC-DCBM with K blocks, define $\Gamma_{\tau,t} \in \mathbb{R}^{N \times K}$ with columns containing the top K eigenvectors of $\mathcal{S}_t$. Then, under Assumption 4, there exists an orthogonal matrix U_t depending on τ_t for each $t = 1, \cdots, T$, such that for any $i, j = 1, \cdots, N$,*

$$\Gamma_{\tau,t} = \Psi_{\tau,t}^{1/2} Z_t (Z_t^\top \Psi_{\tau,t} Z_t)^{-1/2} U_t \text{ and } \Gamma_{\tau,t}^*(i,*) = \Gamma_{\tau,t}^*(j,*) \Longleftrightarrow Z_t(i,*) = Z_t(j,*),$$

where $\Gamma_{\tau,t}^*(i,*) = \Gamma_{\tau,t}(i,*)/\|\Gamma_{\tau,t}(i,*)\|$.

Proof. Denote $D_{B,t}$ as a diagonal matrix with entries $D_{B,t}(i,i) = \sum_{j=1}^K B_t(i,j)$, and $\Psi_{\tau,t} = \text{Diag}(\psi_{\tau,t})$ with $\psi_{\tau,t}(i) = \psi_t \frac{\mathcal{D}_t(i,i)}{\mathcal{D}_{\tau,t}(i,i)}$. Then, Under the dynamic SC-DCBM, we have the decomposition below

$$\mathcal{L}_{\tau,t} = \mathcal{D}_{\tau,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{\tau,t}^{-1/2} = \Psi_{\tau,t}^{1/2} Z_t B_{L,t} Z_t^\top \Psi_{\tau,t}^{1/2},$$

where $B_{L,t} = D_{B,t}^{-1/2} B_t D_{B,t}^{-1/2}$.

Define M_t such that $\mathcal{X} = E(X) = \Psi_{\tau,t}^{1/2} Z_t M_t$, and $\Omega_t = B_{L,t} + \alpha_t M_t \mathcal{W}_t M_t^\top$, then we know

$$\mathcal{S}_t = \Psi_{\tau,t}^{1/2} Z_t \Omega_t Z_t^\top \Psi_{\tau,t}^{1/2}. \tag{A.3.7}$$

Now, denote $Y_{\tau,t} = Z_t^\top \Psi_{\tau,t} Z_t$, and let $H_{\tau,t} = Y_{\tau,t}^{1/2} \Omega_t Y_{\tau,t}^{1/2}$. Then, by eigen-decomposition, we have $H_{\tau,t} = U_t \Lambda_t U_t^\top$. Define $\mathit{\Gamma}_{\tau,t} = \Psi_{\tau,t}^{1/2} Z_t Y_{\tau,t}^{-1/2} U_t$, then

$$\begin{aligned}
\mathit{\Gamma}_{\tau,t}^\top \mathit{\Gamma}_{\tau,t} &= U_t^\top Y_{\tau,t}^{-1/2} Z_t^\top \Psi_{\tau,t}^{1/2} \Psi_{\tau,t}^{1/2} Z_t Y_{\tau,t}^{-1/2} U_t \\
&= U_t^\top Y_{\tau,t}^{-1/2} Y_{\tau,t} Y_{\tau,t}^{-1/2} U_t \\
&= U_t^\top U_t = I,
\end{aligned}$$

and we have

$$\begin{aligned}
\mathcal{S}_t \mathit{\Gamma}_{\tau,t} &= (\Psi_{\tau,t}^{1/2} Z_t \Omega_t Z_t^\top \Psi_{\tau,t}^{1/2}) \Psi_{\tau,t}^{1/2} Z_t (Z_t^\top \Psi_{\tau,t} Z_t)^{-1/2} U_t \\
&= \Psi_{\tau,t}^{1/2} Z_t \Omega_t Y_{\tau,t}^{1/2} U_t \\
&= \left\{ \Psi_{\tau,t}^{1/2} Z_t Y_{\tau,t}^{-1/2} \left(Y_{\tau,t}^{1/2} \Omega_t Y_{\tau,t}^{1/2} \right) \right\} U_t \\
&= \Psi_{\tau,t}^{1/2} Z_t Y_{\tau,t}^{-1/2} (U_t \Lambda_t U_t^\top) U_t \\
&= \mathit{\Gamma}_{\tau,t} \Lambda_t.
\end{aligned}$$

Following Qin and Rohe (2013), it is obvious that

$$\mathit{\Gamma}_{\tau,t}^*(i,*) = \frac{\mathit{\Gamma}_{\tau,t}(i,*)}{\|\mathit{\Gamma}_{\tau,t}(i,*)\|} = Z_{i,t} U_t.$$

Then, by directly applying the Lemma 1 in Binkiewicz et al. (2017), we complete the proof. ■

Lemma A.3.4 *Under Assumption 4', for any $\epsilon > 0$, we have*

$$\sup_t \|\mathit{sym}\,(S_t - \mathcal{S}_t)\| \leq \delta_{\max} \{3 \log(16NT/\epsilon)\}^{1/2}, \tag{A.3.8}$$

with probability at least $1-\epsilon$

Proof. By triangular inequality, we have

$$\sup_t \|\text{sym}\,(S_t - \mathcal{S}_t)\| \le \sup_t \left\|\text{sym}\left(\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right)\right\| \tag{A.3.9}$$

$$+ \sup_t \left\|\text{sym}\left(\mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2}\mathcal{A}_t\mathcal{D}_{C,t}^{-1/2}\right)\right\| \tag{A.3.10}$$

$$+ \sup_t \left\|\text{sym}\left(D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2}\right)\right\|. \tag{A.3.11}$$

For equation (A.3.9), by similar results in proof of Lemma 2, the spectral norm of the symmetrized $\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top$ is bounded by

$$\begin{aligned}
&\sup_t \left\|\text{sym}\left(\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right)\right\| \\
&\quad = \alpha_{\max} \sup_t \left\|\text{sym}\left(X(W_t - \mathcal{W}_t)X^\top\right)\right\| + \alpha_{\max} \sup_t \left\|\text{sym}\left(X\mathcal{W}_t X^\top - \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right)\right\| \\
&\quad \le \alpha_{\max} NRJ^2 \sup_t \|\text{sym}\,(W_t - \mathcal{W}_t)\| + 2\alpha_{\max} NRJ^2 \sup_t \|\text{sym}\,(\mathcal{W}_t)\| \\
&\quad = \mathcal{O}_p(\alpha_{\max} NRJ^2 \xi).
\end{aligned}$$

So, by Assumption 4', we know that for large enough N, with probability at least $1-\epsilon/2$,

$$\sup_t \left\|\text{sym}\left(\alpha_t X W_t X^\top - \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^\top\right)\right\| \le c'_w a.$$

For equation (A.3.10), by Assumption 4' and matrix Bernstein inequality, we know under assumption $\underline{\delta}' > 3\log(16NT/\epsilon)$, $a < 1$. Therefore, similar to proof of

Lemma 2, we have

$$\begin{aligned}
&\Pr\left(\sup_t \left\|\text{sym}\left(\mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{C,t}^{-1/2}\right)\right\| > a\right) \\
&\leq \sum_{t=1}^{T} \Pr\left(\left\|\text{sym}\left(\mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2} \mathcal{A}_t \mathcal{D}_{C,t}^{-1/2}\right)\right\| > a\right) \\
&\leq 4NT \exp\left(-\frac{3\log(16NT/\epsilon)/\underline{\delta}'}{2/\underline{\delta}' + 2a/(3\underline{\delta}')}\right) \\
&\leq 4NT \exp\left(-\log(16NT/\epsilon)\right) \\
&= \epsilon/4.
\end{aligned}$$

Lastly, for equation (A.3.11), by Rohe et al. (2016), we know with probability at least $1 - \epsilon/2$,

$$\begin{aligned}
&\sup_t \left\|\text{sym}\left(D_{R,t}^{-1/2} A_t D_{C,t}^{-1/2} - \mathcal{D}_{R,t}^{-1/2} A_t \mathcal{D}_{C,t}^{-1/2}\right)\right\| \\
&\quad = \sup_t \left\|\text{sym}\left(L_{\tau,t} - \mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} L_{\tau,t} D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2}\right)\right\| \\
&\quad = \sup_t \left\|\text{sym}\left((I - \mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2}) L_{\tau,t} D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2} + L_{\tau,t}(I - D_{\tau,t}^{1/2} \mathcal{D}_{\tau,t}^{-1/2})\right)\right\| \\
&\quad \leq \sup_t \left\|\text{sym}\left(\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\right)\right\| \sup_t \left\|\text{sym}\left(\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2}\right)\right\| + \sup_t \left\|\text{sym}\left(\mathcal{D}_{\tau,t}^{-1/2} D_{\tau,t}^{1/2} - I\right)\right\| \\
&\quad \leq a^2 + 2a
\end{aligned}$$

Therefore, combine the results above, we obtain the upper bound for $\|\text{sym}\,(S_t - \mathcal{S}_t)\|$, i.e., with probability at least $1 - \epsilon$,

$$\sup_t \|\text{sym}\,(S_t - \mathcal{S}_t)\| \leq a^2 + 3a + c'_w a \leq (4 + c'_w)a = (4 + c'_w)\left\{\frac{3\log(16N/\epsilon)}{\underline{\delta}'}\right\}^{1/2}. \tag{A.3.12}$$

■

Lemma A.3.5 *Under the dynamic SC-DCcBM with K_R row blocks and K_C column blocks, define $\Gamma_{R,t} \in \mathbb{R}^{N \times K_R}$ with columns containing the top K_R left singular vectors of $\mathcal{S}_t$ and $\Gamma_{C,t} \in \mathbb{R}^{N \times K_C}$ with columns containing the top K_C right singular vectors of $\mathcal{S}_t$. Then, under Assumption 4', there exist orthogonal matrices $U_{R,t}$ and*

$U_{C,t}$ depending on τ_t for each $t = 1, \cdots, T$, such that for any $i, j = 1, \cdots, N$,

$$\Gamma_{p,t} = {\Psi^p_{\tau,t}}^{1/2} Z_{p,t} (Z^\top_{p,t} {\Psi^p_{\tau,t}}^{1/2} Z_{p,t})^{-1/2} U_{p,t}$$

and

$$\Gamma^*_{p,t}(i,*) = \Gamma^*_{p,t}(j,*) \Longleftrightarrow Z_{p,t}(i,*) = Z_{p,t}(j,*).$$

where $\Gamma^*_{p,t}(i,*) = \Gamma_{p,t}(i,*)/\|\Gamma_{p,t}(i,*)\|$ with $p \in \{R, C\}$.

Proof. Define $D^R_{B,t}$ and $D^C_{B,t}$ are diagonal matrices with entries $D^R_{B,t}(i,i) = \sum_{j=1}^K B_t(i,j)$ and $D^C_{B,t}(i,i) = \sum_j^K B_t(j,i)$, and $\Psi^p_{\tau,t} = \text{Diag}(\psi^p_{\tau,t})$ with $\psi^p_{\tau,t}(i) = \psi^p_i \dfrac{\mathcal{D}_{p,t}(i,i)}{\mathcal{D}_{p,t}(i,i) + \tau_{p,t}}$ for $p \in \{R, C\}$. Then under dynamic SC-DCcBM, we have the decomposition below,

$$\mathcal{L}_{\tau,t} = \mathcal{D}^{-1/2}_{R,t} \mathcal{A}_t \mathcal{D}^{-1/2}_{C,t} = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} B_{L,t} Z^\top_{C,t} {\Psi^C_{\tau,t}}^{1/2},$$

where $B_{L,t} = \left(D^R_{B,t}\right)^{-1/2} B_t \left(D^C_{B,t}\right)^{-1/2}$.

Define $M_{R,t}$ and $M_{C,t}$ such that $\mathcal{X} = E(X) = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} M_{R,t} = {\Psi^C_{\tau,t}}^{1/2} Z_{C,t} M_{C,t}$, and $\Omega_t = B_{L,t} + \alpha_t M_{R,t} \mathcal{W}_t M^\top_{C,t}$, then we know

$$\mathcal{S}_t = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} \Omega_t Z^\top_{C,t} {\Psi^C_{\tau,t}}^{1/2}. \tag{A.3.13}$$

Now, denote $Y_{R,t} = Z^\top_{R,t} \Psi^R_{\tau,t} Z_{R,t}$ and $Y_{C,t} = Z^\top_{C,t} \Psi^C_{\tau,t} Z_{C,t}$, and let $H_{\tau,t} = Y^{1/2}_{R,t} \Omega_t Y^{1/2}_{C,t}$. Then, by singular value decomposition, we have $H_{\tau,t} = U_{R,t} \Lambda_t U^\top_{C,t}$. Define $\Gamma_{R,t} = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} Y^{-1/2}_{R,t} U_{R,t}$ and $\Gamma_{C,t} = {\Psi^C_{\tau,t}}^{1/2} Z_{C,t} Y^{-1/2}_{C,t} U_{C,t}$, then, for $p \in \{R, C\}$,

$$\begin{aligned}
\Gamma^\top_{p,t} \Gamma_{p,t} &= U^\top_{p,t} Y^{-1/2}_{p,t} Z^\top_{p,t} {\Psi^p_{\tau,t}}^{1/2} {\Psi^p_{\tau,t}}^{1/2} Z_{p,t} Y^{-1/2}_{p,t} U_{p,t} \\
&= U^\top_{p,t} Y^{-1/2}_{p,t} Y_{p,t} Y^{-1/2}_{p,t} U_{p,t} \\
&= U^\top_{p,t} U_{p,t} = I,
\end{aligned}$$

and we have

$$\begin{aligned}
&\Gamma_{R,t}\Lambda_t\Gamma_{C,t} = \Psi_{\tau,t}^{R\,1/2} Z_{R,t} Y_{R,t}^{-1/2} U_{R,t}\Lambda_t U_{C,t}^{\top} Y_{C,t}^{-1/2} Z_{C,t}^{\top} \Psi_{\tau,t}^{C\,1/2} \\
&= \Psi_{\tau,t}^{R\,1/2} Z_{R,t} Y_{R,t}^{-1/2} H_{\tau,t} Y_{C,t}^{-1/2} Z_{C,t}^{\top} \Psi_{\tau,t}^{C\,1/2} \\
&= \Psi_{\tau,t}^{R\,1/2} Z_{R,t} Y_{R,t}^{-1/2} \left(Y_{R,t}^{1/2} \Omega_t Y_{C,t}^{1/2} \right) Y_{C,t}^{-1/2} Z_{C,t}^{\top} \Psi_{\tau,t}^{C\,1/2} \\
&= \Psi_{\tau,t}^{R\,1/2} Z_{R,t} \Omega_t Z_{C,t}^{\top} \Psi_{\tau,t}^{C\,1/2} = \mathcal{S}_t
\end{aligned}$$

Following Rohe et al. (2016), it is obvious that

$$\Gamma_{p,t}^{*}(i,*) = \frac{\Gamma_{p,t}(i,*)}{\|\Gamma_{p,t}(i,*)\|} = Z_{p,t}(i,*)U_{p,t}, \text{ for } p \in \{R, C\},$$

which completes the proof. ■

A.3.2 Proof of Theorem 4.2.1

Proof. By Binkiewicz et al. (2017) and the solution of $(1+\varepsilon)$-approximate k-means method, we know for each period $t = 1, 2, \cdots, T$, we have

$$\frac{|\mathbb{M}_t|}{N} \leq \frac{2(2+\varepsilon)^2}{m_z^2 N} \|U_t - \mathcal{U}_t\mathcal{O}_t\|_F^2 \tag{A.3.14}$$

and

$$\|U_t - \mathcal{U}_t\mathcal{O}_t\|_F \leq \frac{8K^{1/2}}{\lambda_{K,t}} \left\| \widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t \right\|, \tag{A.3.15}$$

where $m_z \overset{\text{def}}{=} \min_{i,t}\{\min\{\|\Gamma_{\tau,t}(i,*)\|, \|\Gamma_{\tau,t}(i,*)\|\}\}$ with $\Gamma_{\tau,t}$ and $\Gamma_{\tau,t}$ being defined in Lemma 3.

Then, we have

$$\sup_t \frac{|\mathbb{M}_t|}{N} \leq \frac{2^9(2+\varepsilon)^2 K}{m_z^2 N \lambda_{K,\max}^2} \sup_t \left\| \widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t \right\|^2. \tag{A.3.16}$$

Then, for $\mathcal{S}_t$, we have the following representation:

$$\mathcal{S}_t = \mathcal{D}_{\tau,t}^{-1/2} \Psi Z_t B_t Z_t^{\top} \Psi \mathcal{D}_{\tau,t}^{-1/2} + \alpha_t \mathcal{X}\mathcal{W}_t\mathcal{X}^{\top}, \tag{A.3.17}$$

To figure out the upper bound of the estimation error, we have to evaluate the error bound $\sup_t \left\| \widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t \right\|$. Define

$$\mathcal{S}_{t,r} = \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,l}(i) \mathcal{S}_{t+i}, \tag{A.3.18}$$

then by triangle inequality, we have

$$\Delta(r) = \sup_t \left\| \widehat{\mathcal{S}}_{t,r} - \mathcal{S}_t \right\| \leq \sup_t \left\| \widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r} \right\| + \sup_t \left\| \mathcal{S}_{t,r} - \mathcal{S}_t \right\| = \Delta_1(r) + \Delta_2(r). \tag{A.3.19}$$

For $\Delta_1(r)$, by Lemma A.3.2, we have

$$\begin{aligned} \Delta_1(r) &= \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,l}(i) \sup_t \left\| S_{t+i} - \mathcal{S}_{t+i} \right\| \\ &\leq \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,l}(i) \left\{ (4 + c_w) \left[\frac{3 \log(8NT/\epsilon)}{\underline{\delta}} \right]^{1/2} \right\} \\ &\leq W_{\max}(4 + c_w) \left\{ \frac{3 \log(8NT/\epsilon)}{\underline{\delta}} \right\}^{1/2}. \end{aligned} \tag{A.3.20}$$

For $\Delta_2(r)$, we have the following decomposition

$$\Delta_2(r) = \sup_t \left\| \mathcal{S}_{t,r} - \mathcal{S}_t \right\| \leq \sup_t \left\| \mathcal{S}_{t,r} - \widetilde{\mathcal{S}}_{t,r} \right\| + \sup_t \left\| \widetilde{\mathcal{S}}_{t,r} - \mathcal{S}_t \right\| = \Delta_{21}(r) + \Delta_{22}(r), \tag{A.3.21}$$

where

$$\widetilde{\mathcal{S}}_{t,r} = \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,l}(i) \left(\mathcal{D}_{\tau,t}^{-1/2} \Psi Z_t B_{t+i} Z_t^\top \Psi \mathcal{D}_{\tau,t}^{-1/2} + \alpha_{t+i} \mathcal{X} \mathcal{W}_{t+i} \mathcal{X}^\top \right). \tag{A.3.22}$$

Then for Δ_{21}, we have

$$
\begin{aligned}
&\Delta_{21}(r) \\
&\leq W_{\max}\frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r}\sup_t\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i}B_{t+i}Z_{t+i}^{\top}\Psi\mathcal{D}_{\tau,t+i}^{-1/2}-\mathcal{D}_{\tau,t}^{-1/2}\Psi Z_tB_{t+i}Z_t^{\top}\Psi\mathcal{D}_{\tau,t}^{-1/2}\right\| \\
&\leq W_{\max}\frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r}\sup_t\left\{\left(\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i}\right\|+\left\|\mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t\right\|\right)\|B_{t+i}\|\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i}-\mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t\right\|\right\} \\
&\leq W_{\max}\frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r}\sup_t\left\{\left(\left\|\mathcal{D}_{\tau,t}^{-1/2}\right\|\|Z_{t+i}\|+\left\|\mathcal{D}_{\tau,t}^{-1/2}\right\|\|Z_t\|\right)\|B_{t+i}\|\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i}-\mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t\right\|\right\},
\end{aligned}
$$

where the last inequality comes from the fact that $\|\Psi\| = \max_i\left|\sqrt{\psi_i}\right| \leq 1$.

Then, observe that $\sup_t\left\|\mathcal{D}_{\tau,t}^{-1/2}\right\| \leq \underline{\delta}^{-1/2}$, $\sup_t\|Z_t\| \leq P_{\max}^{1/2}$, $\sup_t\|B_t\| \leq K$, we then have

$$
\sup_t\left\{\left\|\mathcal{D}_{\tau,t}^{-1/2}\right\|\|Z_{t+i}\|+\left\|\mathcal{D}_{\tau,t}^{-1/2}\right\|\|Z_t\|\right\} \leq 2\underline{\delta}^{-1/2}P_{\max}^{1/2}. \tag{A.3.23}
$$

Further, note that

$$
\begin{aligned}
&\sup_t\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i}-\mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t\right\| \qquad \text{(A.3.24)}\\
&\quad\leq \sup_t\left\{\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_{t+i}-\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_t\right\|+\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\Psi Z_t-\mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t\right\|\right\} \\
&\quad\leq \sup_t\left\{\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\right\|\|\Psi\|\|Z_{t+i}-Z_t\|+\left(\left\|\mathcal{D}_{\tau,t+i}^{-1/2}\right\|\|\Psi\|+\left\|\mathcal{D}_{\tau,t}^{-1/2}\right\|\|\Psi\|\right)\|Z_t\|\right\} \\
&\quad\leq \sqrt{\frac{2|r|s}{\underline{\delta}}}+\sqrt{\frac{4P_{\max}}{\underline{\delta}}}.
\end{aligned}
$$

Then, combine the results above with the assumption $\underline{\delta} > 3\log(8NT/\epsilon)$ in Lemma A.3.2, we have

$$
\Delta_{21}(r) \leq \frac{2W_{\max}K}{\underline{\delta}}(\sqrt{2P_{\max}rs}+2P_{\max}). \tag{A.3.25}
$$

Lastly, for $\Delta_{22}(r)$, for notational simplicity, denote $Y_{\tau,t} \overset{\text{def}}{=} \mathcal{D}_{\tau,t}^{-1/2}\Psi Z_t$. Then,

apply the results in Pensky and Zhang (2017) and proof of Lemma A.3.2, we obtain

$$\begin{aligned}
\Delta_{22}(r) &= \sup_t \left\| \widetilde{\mathcal{S}}_{t,r} - \mathcal{S}_t \right\| \\
&= \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,l}(i) \sup_t \left(Y_{\tau,t} \left\| B_{t+i} - B_t \right\| Y_{\tau,t}^\top + \left\| \alpha_{t+i} \mathcal{X} \mathcal{W}_{t+i} \mathcal{X}^\top - \alpha_t \mathcal{X} \mathcal{W}_t \mathcal{X}^\top \right\| \right) \\
&\leq \sup_t \left\{ \max_{1 \leq j' \leq N} \sum_{j=1}^{N} \left| (Y_{\tau,t} Q_{r,t} Y_{\tau,t}^\top)(j, j') \right| \right\} + 2\alpha_{\max} W_{\max} N R J^2 \sup_t \left\| \mathcal{W}_t \right\| \\
&\leq \sup_t \left\{ \max_{k,k'} |Q_{r,t}| \max_{1 \leq j' \leq N} \sum_{k=1}^{K} \sum_{k'=1}^{K} \left[\sum_{j \in \mathcal{G}_{t,k}} Y_{\tau,t}(j,k) \right] Y_{\tau,t}(j',k') \right\} \qquad \text{(A.3.26)} \\
&\quad + 2W_{\max} \left\{ \frac{3 \log(8NT/\epsilon)}{\underline{\delta}} \right\}^{1/2} \\
&\leq \frac{NLW_{\max}}{\underline{\delta} \cdot l!} \left(\frac{r}{T} \right)^{\beta} + 2W_{\max} \left\{ \frac{3 \log(8NT/\epsilon)}{\underline{\delta}} \right\}^{1/2} \qquad \text{(A.3.27)}
\end{aligned}$$

where the second last inequality comes from Assumption 4 and the last inequality come from the fact that $\max_i \psi_i \leq 1$.

Now, combine the results provided by equation (A.3.16), (A.3.20), (A.3.25), and (A.3.27), we derive the upper bound for misclustering rate of dynamic DCBM: with probability at least $1 - \epsilon$,

$$\sup_t \frac{|\mathbb{M}_t|}{N} \leq \frac{c_1(\varepsilon) K W_{\max}^2}{m_z^2 N \lambda_{K,\max}^2} \left\{ (6 + c_w) \frac{b}{\underline{\delta}^{1/2}} + \frac{2K}{\underline{\delta}} (\sqrt{2P_{\max} rs} + 2P_{\max}) + \frac{NL}{\underline{\delta} \cdot l!} \left(\frac{r}{T} \right)^{\beta} \right\}^2.$$

where $b = \sqrt{3 \log(8NT/\epsilon)}$, $\lambda_{K,\max} = \max_t \{\lambda_{K,t}\}$ and $c_1(\varepsilon) = 2^9(2+\varepsilon)^2$. ■

A.3.3 Proof of Lemma 4.2.1

Proof. Firstly, by Lemma B.1 in Supplementary material of Lei and Rinaldo (2015), fix $\eta \in (0, 1)$, we have

$$\left\| \widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r} \right\| \leq (1 - \eta)^{-2} \sup_{x,y \in \mathcal{T}} \left| x^\top (\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}) y \right|, \qquad \text{(A.3.28)}$$

where $\mathcal{T} = \{x = (x_1, \cdots, x_N) \in \mathbb{R}^N, \|x\| = 1, \sqrt{N}x_i/\eta \in \mathbb{Z}, \forall i\}$. Then, let $d = rN\|\mathcal{S}_t\|_\infty$ with $r \geq 1$, we can split the pairs (x_i, y_j) into light pairs

$$\mathscr{L} = \mathscr{L}(x, y) \stackrel{\text{def}}{=} \{(i, j) : |x_i y_j| \leq \sqrt{d}/N\},$$

and into heavy pairs

$$\bar{\mathscr{L}} = \bar{\mathscr{L}}(x, y) \stackrel{\text{def}}{=} \{(i, j) : |x_i y_j| > \sqrt{d}/N\}.$$

For the light pair, first denote

$$u_{ij} = x_i y_j \mathbb{1}_{\{|x_i y_j| \leq \sqrt{d}/N\}} + x_j y_i \mathbb{1}_{\{|x_j y_i| \leq \sqrt{d}/N\}},$$

then we have

$$\begin{aligned}
&\sum_{(i,j)\in\mathscr{L}(x,y)} x_i y_j (\widehat{\mathcal{S}}_{t,r}(i, j) - \mathcal{S}_{t,r}(i, j)) \\
&\qquad = \frac{1}{|\mathcal{F}_r|} \sum_{1\leq i\leq j\leq N} \sum_{k\in\mathcal{F}_r} u_{ij} W_{r,\ell}(k) \left[S_{t+k}(i, j) - \mathcal{S}_{t+k}(i, j)\right].
\end{aligned}$$

Denote $w_{ij} = |\mathcal{F}_r|^{-1} \sum_{k\in\mathcal{F}_r} W_{r,\ell}(k) \left[S_{t+k}(i, j) - \mathcal{S}_{t+k}(i, j)\right]$ and $\xi_{ij} = w_{ij} u_{ij}$, then we have $|w_{ij}| \leq W_{\max}\|\mathcal{S}_t\|_\infty$, and by Pensky and Zhang (2017), it is known that ξ_{ij} is a independent random variable with zero mean and absolute values bounded by $|\xi_{ij}| \leq 2W_{\max}\sqrt{r\|\mathcal{S}_t\|_\infty^3/N}$, using the fact that $|u_{ij}| \leq 2\sqrt{d}/N$.

Now, applying Bernstein inequality, for any $c > 0$, we have

$$\begin{aligned}
&\Pr\left(\sup_{x,y\in\mathcal{T}}\left|\sum_{1\le i\le j\le N}\xi_{ij}\right| \ge \frac{c\sqrt{d}}{r}\right) \\
&\quad\le 2\exp\left(-\frac{\frac{c^2 d}{2r}}{\sum_{1\le i\le j\le N}\mathbb{E}\left(\xi_{ij}^2\right) + \frac{2W_{\max}}{3}\sqrt{\frac{r\|\mathcal{S}_t\|_\infty^3}{N}} \times \frac{c\sqrt{d}}{r}}\right) \\
&\quad\le 2\exp\left(-\frac{\frac{c^2 d}{2r}}{\left(\sum_{1\le i\le j\le N} u_{ij}^2\right) W_{\max}^2\|\mathcal{S}_t\|_\infty^2 + \frac{2W_{\max}}{3}\sqrt{\frac{r\|\mathcal{S}_t\|_\infty^3}{N}} \times \frac{c\sqrt{d}}{r}}\right) \\
&\quad\le 2\exp\left(-\frac{3c^2 N}{12W_{\max}^2\|\mathcal{S}_t\|_\infty + 4cW_{\max}\|\mathcal{S}_t\|_\infty}\right).
\end{aligned}$$

Then, by a standard volume argument, we have the cardinality of $\mathcal{T} \le \exp(N\log(7/\eta))$, and this ensures

$$\begin{aligned}
&\Pr\left(\sup_{x,y\in\mathcal{T}}\left|\sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} x_i y_j(\widehat{\mathcal{S}}_{t,r}(i,j) - \mathcal{S}_{t,r}(i,j))\right| \ge \frac{c\sqrt{d}}{r}\right) \\
&\quad\le \exp\left\{-\left(\frac{3c^2}{12W_{\max}^2\|\mathcal{S}_t\|_\infty + 4cW_{\max}\|\mathcal{S}_t\|_\infty} - 2\log\left(\frac{7}{\eta}\right)\right)N\right\}. \quad \text{(A.3.29)}
\end{aligned}$$

For the heavy pairs, we know

$$\begin{aligned}
&\left|\sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} x_i y_j(\widehat{\mathcal{S}}_{t,r}(i,j) - \mathcal{S}_{t,r}(i,j))\right| \\
&= \left|\frac{1}{|\mathcal{F}_r|}\sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} x_i y_j \sum_{k\in\mathcal{F}_r} W_{r,\ell}(k)(S_{t+k}(i,j) - \mathcal{S}_{t+k}(i,j))\right| \\
&\le \left|\frac{1}{|\mathcal{F}_r|}\sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} \frac{x_i^2 y_j^2}{|x_i y_j|}\sum_{k\in\mathcal{F}_r} W_{r,\ell}(k)(S_{t+k}(i,j) - \mathcal{S}_{t+k}(i,j))\right| \\
&\le \frac{N}{\sqrt{d}} W_{\max}\|\mathcal{S}_t\|_\infty \sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} x_i^2 y_j^2 \\
&= \frac{W_{\max}}{r}\sqrt{d}\sum_{(i,j)\in\bar{\mathscr{L}}(x,y)} x_i^2 y_j^2 \\
&\le \frac{W_{\max}}{r}\sqrt{d}.
\end{aligned}$$

Therefore, choosing $c = W_{\max}$ in equation (A.3.29), we have

$$\Pr\left(\sup_{x,y\in\mathcal{T}}\left|\sum_{1\le i\le j\le N} x_i y_j(\widehat{\mathcal{S}}_{t,r}(i,j) - \mathcal{S}_{t,r}(i,j))\right| \le \frac{W_{\max}\sqrt{d}}{r}\right) \ge 1-\epsilon \quad \text{(A.3.30)}$$

where $\epsilon = N^{\left(\frac{3}{16\|\mathcal{S}_t\|_\infty} - 2\log\left(\frac{7}{\eta}\right)\right)}$.

In the end, by equation (A.3.28) and (A.3.30), we obtain, with probability $1-\epsilon$,

$$\left\|\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r}\right\| \le (1-\eta)^{-2}\sup_{x,y\in\mathcal{T}}\left|x^\top(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r})y\right| \le (1-\eta)^{-2}\frac{W_{\max}\sqrt{d}}{r}.$$

■

A.3.4 Proof of Theorem 4.2.2

Proof. In this proof, we deal with the clustering of left singular vector and the right singular vectors separately.

(1) Clustering for $Z_{R,t}$. First, by Rohe et al. (2016) and solution of $(1+\varepsilon)$-approximate k-means clustering, for each period $t = 1, \cdots, T$, we have

$$\frac{\left|\mathbb{M}_t^R\right|}{N} \le \frac{8(2+\varepsilon)^2}{m_r^2 N}\left\|U_t - \mathcal{U}_t\mathcal{O}_t\right\|_F^2, \quad \text{(A.3.31)}$$

where denote $m_r \overset{\text{def}}{=} \min_{i,t}\{\min\{\|\Gamma_{R,t}(i,*)\|, \|\mathit{\Gamma}_{R,t}(i,*)\|\}\}$, and by improved version of Davis-Kahn theorem from Lei and Rinaldo (2015), we have

$$\left\|U_t - \mathcal{U}_t\mathcal{O}_t\right\|_F \le \frac{2\sqrt{2K_R}}{\lambda_{K_R,t}}\left\|\text{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t\right)\right\|, \quad \text{(A.3.32)}$$

as $K_R \le K_C$.

Then, base on equation (A.3.31) and (A.3.32), we have

$$\sup_t \frac{\left|\mathbb{M}_t^R\right|}{N} \le \frac{2^6(2+\varepsilon)^2 K_R}{m_r^2 N\lambda_{K_R,\max}^2}\sup_t\left\|\text{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t\right)\right\|^2. \quad \text{(A.3.33)}$$

Then, for $\mathcal{S}_t$, we have the following representation:

$$\mathcal{S}_t = \mathcal{D}_{R,t}^{-1/2} \Psi^R Z_{R,t} B_t Z_{C,t}^\top \Psi^C \mathcal{D}_{C,t}^{-1/2} + \alpha_t \mathcal{X} \mathcal{W}_t \mathcal{X}^\top, \tag{A.3.34}$$

where $\Psi^p = \text{Diag}(\psi^p)$ with $p \in \{R, C\}$. Then, by definition of $\mathcal{S}_{t,r} \stackrel{\text{def}}{=} |\mathcal{F}_r|^{-1} \sum_{i \in \mathcal{F}_r} W_{r,\ell}(i) \mathcal{S}_{t+i}$, we have the decomposition

$$\Delta(r) =\leq \sup_t \left\| \text{sym}\left(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r} \right) \right\| + \sup_t \left\| \text{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t \right) \right\| = \Delta_1(r) + \Delta_2(r). \tag{A.3.35}$$

Now, we evaluate $\Delta_1(r)$ and $\Delta_2(r)$ respectively. For $\Delta_1(r)$, by Lemma 4, we have

$$\begin{aligned} \sup_t \left\| \text{sym}\left(\widehat{\mathcal{S}}_{t,r} - \mathcal{S}_{t,r} \right) \right\| &= \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,\ell}(i) \sup_t \left\| \text{sym}\left(S_{t+i} - \mathcal{S}_{t+i} \right) \right\| \\ &\leq W_{\max}(4 + c'_w) \left\{ \frac{3 \log(16N/\epsilon)}{\underline{\delta}'} \right\}^{1/2}. \end{aligned} \tag{A.3.36}$$

For $\Delta_2(r)$, we first define

$$\widetilde{\mathcal{S}}_{t,r} = \frac{1}{|\mathcal{F}_r|} \sum_{i \in \mathcal{F}_r} W_{r,\ell}(i) \left(Y_{R,t} B_{t+i} Y_{C,t}^\top + \alpha_{t+i} \mathcal{X} \mathcal{W}_{t+i} \mathcal{X} \right). \tag{A.3.37}$$

where $Y_{R,t} \stackrel{\text{def}}{=} \mathcal{D}_{R,t}^{-1/2} \Psi^R Z_{R,t}$ and $Y_{C,t} \stackrel{\text{def}}{=} \mathcal{D}_{C,t}^{-1/2} \Psi^C Z_{C,t}$.

Then, we decompose $\Delta_2(r)$ as

$$\sup_t \left\| \text{sym}\left(\mathcal{S}_{t,r} - \mathcal{S}_t \right) \right\| \leq \sup_t \left\| \text{sym}\left(\mathcal{S}_{t,r} - \widetilde{\mathcal{S}}_{t,r} \right) \right\| + \sup_t \left\| \text{sym}\left(\widetilde{\mathcal{S}}_{t,r} - \mathcal{S}_t \right) \right\| = \Delta_{21}(r) + \Delta_{22}(r). \tag{A.3.38}$$

For notation simplicity, we define $Y_{p,t} \stackrel{\text{def}}{=} \mathcal{D}_{p,t}^{-1/2} \Psi^p Z_{p,t}$ for $p \in \{R, C\}$, and the block diagonal matrix $\mathcal{Y}_t$ such that

$$\mathcal{Y}_t \stackrel{\text{def}}{=} \begin{bmatrix} Y_{R,t} & 0_{N \times K_C} \\ 0_{N \times K_R} & Y_{C,t} \end{bmatrix}. \tag{A.3.39}$$

Note that

$$\begin{aligned}\left\|\text{sym}\left(\mathcal{S}_{t,r}-\widetilde{\mathcal{S}}_{t,r}\right)\right\| &\leq W_{\max}\max_{|i|\leq r}\left\|\text{sym}\left(Y_{R,t+i}B_{t+i}Y_{C,t+i}^{\top}-Y_{R,t}B_{t+i}Y_{C,t}^{\top}\right)\right\| \\ &= W_{\max}\max_{|i|\leq r}\left\|\mathcal{Y}_{t+i}\text{sym}(B_{t+i})\mathcal{Y}_{t+i}^{\top}-\mathcal{Y}_{t}\text{sym}(B_{t+i})\mathcal{Y}_{t}^{\top}\right\| \\ &\leq W_{\max}\max_{|i|\leq r}\left(\|\mathcal{Y}_{t+i}\|+\|\mathcal{Y}_{t}\|\right)\|\text{sym}(B_{t+i})\|\,\|\mathcal{Y}_{t+i}-\mathcal{Y}_{t}\|,\end{aligned}$$

and $\|\text{sym}(B_{t+i})\| \leq K_C$ and $\|\Psi^p\| \leq 1$ for $p \in \{R, C\}$, we then have

$$\begin{aligned}\|\mathcal{Y}_t\| &= \max\{\|\mathcal{D}_{R,t}^{-1/2}\Psi^R Z_{R,t}\|, \|\mathcal{D}_{C,t}^{-1/2}\Psi^C Z_{C,t}\|\} \\ &\leq \max\{\|\mathcal{D}_{R,t}^{-1/2}\|\|\Psi^R\|\|Z_{R,t}\|, \|\mathcal{D}_{C,t}^{-1/2}\|\|\Psi^C\|\|Z_{C,t}\|\} \\ &\leq \underline{\delta'}^{-1/2}P_{\max}^{1/2},\end{aligned}$$

and

$$\begin{aligned}\|\mathcal{Y}_{t+i}-\mathcal{Y}_t\| &= \max\left\{\|Y_{R,t+i}-Y_{R,t}\|, \|Y_{C,t+i}-Y_{C,t}\|\right\} \\ &\leq \max\left\{\left\|\mathcal{D}_{R,t+i}^{-1/2}Z_{R,t+i}-\mathcal{D}_{R,t}^{-1/2}Z_{C,t}\right\|, \left\|\mathcal{D}_{C,t+i}^{-1/2}Z_{C,t+i}-\mathcal{D}_{C,t}^{-1/2}Z_{C,t}\right\|\right\} \\ &\leq \sqrt{\frac{2|r|s}{\underline{\delta'}}}+\sqrt{\frac{4P_{\max}}{\underline{\delta'}}},\end{aligned}$$

where the last inequality comes from the same derivation as equation (A.3.24).

Therefore, we have

$$\begin{aligned}\Delta_{21}(r) &= \sup_t\left\|\text{sym}\left(\mathcal{S}_{t,r}-\widetilde{\mathcal{S}}_{t,r}\right)\right\| \qquad \text{(A.3.40)} \\ &\leq W_{\max}\times(2\underline{\delta'}^{-1/2}P_{\max}^{1/2})\times K_C\times\left(\sqrt{\frac{2|r|s}{\underline{\delta'}}}+\sqrt{\frac{4P_{\max}}{\underline{\delta'}}}\right) \\ &= \frac{2W_{\max}K_C}{\underline{\delta'}}\left(\sqrt{2P_{\max}rs}+2P_{\max}\right).\end{aligned}$$

Lastly, for $\Delta_{22}(r)$, define

$$\mathcal{Z}_t \stackrel{\text{def}}{=} \begin{bmatrix} \Psi_{\tau,t}^{R\,1/2} Z_{R,t} & 0_{N\times K_C} \\ 0_{N\times K_R} & \Psi_{\tau,t}^{C\,1/2} Z_{C,t} \end{bmatrix}. \tag{A.3.41}$$

Then, by Assumption 4', we have

$$\begin{aligned}
\Delta_{22}(r) &= \sup_t \left\| \text{sym}\left(\widetilde{\mathcal{S}}_{t,r} - \mathcal{S}_t\right)\right\| \\
&\le \frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r} W_{r,\ell}(i) \sup_t \left\|\text{sym}\left(Y_{R,t}(B_{t+i}-B_t)Y_{C,t}^\top\right)\right\| \\
&\quad + \frac{1}{|\mathcal{F}_r|}\sum_{i\in\mathcal{F}_r} W_{r,\ell}(i) \sup_t \left\|\text{sym}\left(\alpha_{t+i}\mathcal{X}\mathcal{W}_{t+i}\mathcal{X} - \alpha_t\mathcal{X}\mathcal{W}_t\mathcal{X}\right)\right\| \\
&\le \frac{1}{|\mathcal{F}_{r,j}|}\sum_{i\in\mathcal{F}_{r,j}} W_{r,\ell}^j(i) \left\|\mathcal{Y}_t \text{sym}(B_{t+i}-B_t)\mathcal{Y}_t^\top\right\| + 2\alpha_{\max}W_{\max}NRJ^2 \sup_t \|\text{sym}(\mathcal{W}_t)\| \\
&\le \Delta_{22}^{(1)} + 2W_{\max}\left\{\frac{3\log(16N/\epsilon)}{\underline{\delta}'}\right\}^{1/2}.
\end{aligned}$$

For $\Delta_{22}^{(1)}$, apply the same argument in previous proof for ScBM, we know

$$\begin{aligned}
\Delta_{22}^{(1)} &\le \max_{k,k'}|Q_{r,t}(k,k')| \max_{1\le j'\le 2N} \sum_{k=1}^{K_R+K_C}\sum_{k'=1}^{K_R+K_C}\left[\sum_{j\in\mathcal{G}_{t,k}}\mathcal{Y}_t(j,k)\right]\mathcal{Y}_t(j',k') \\
&\le W_{\max}\frac{NL}{\underline{\delta}'\cdot\ell!}\left(\frac{r}{T}\right)^\beta.
\end{aligned} \tag{A.3.42}$$

Therefore, combine equation (A.3.33), (A.3.35), (A.3.36), (A.3.40), and (A.3.42), we obtain

$$\frac{\left|\mathbb{M}_t^R\right|}{N} \le \frac{c_2(\varepsilon)K_R W_{\max}^2}{m_r^2 N\lambda_{K_R,\max}^2}\left\{(6+c_w')\frac{b'}{\underline{\delta}'^{1/2}} + \frac{2K_C}{\underline{\delta}'}(\sqrt{2P_{\max}rs}+2P_{\max}) + \frac{NL}{\underline{\delta}'\cdot\ell!}\left(\frac{r}{T}\right)^\beta\right\}^2,$$

where $c_2(\varepsilon) = 2^6(2+\varepsilon)^2$, $b' = \{3\log(16NT/\epsilon)\}^{1/2}$ and $\lambda_{K_R,\max} = \max_t\{\lambda_{K_R,t}\}$.

(2) Clustering for $Z_{C,t}$.

As shown in equation (A.3.13), the population regularized graph Laplacian of

dynamic DCcBM has following decomposition

$$\mathcal{S}_t = {\Psi^R_{\tau,t}}^{1/2} Z_{R,t} \Omega_t Z^\top_{C,t} {\Psi^C_{\tau,t}}^{1/2} \tag{A.3.43}$$

Then, let $Y_{R,t} = Z^\top_{R,t} \Psi^R_{\tau,t} Z_{R,t}$ and $Y_{C,t} = Z^\top_{C,t} \Psi^C_{\tau,t} Z_{C,t}$, and

$$H_{\tau,t} = Y^{1/2}_{R,t} \Omega_t Y^{1/2}_{C,t}. \tag{A.3.44}$$

Now, following Rohe et al. (2016), we can define

$$\gamma_c \stackrel{\text{def}}{=} \min_t \{ \min_{i \neq j} \| H_t(*, i) - H_t(*, j) \| \}, \tag{A.3.45}$$

and thus

$$\frac{\left| \mathbb{M}^C_t \right|}{N} \leq \frac{16(2+\varepsilon)^2}{m_c^2 N \gamma_c^2} \left\| U_t - \mathcal{U}_t \mathcal{O}_t \right\|^2_F. \tag{A.3.46}$$

where $m_c \stackrel{\text{def}}{=} \min_{i,t} \{ \min \{ \| \Gamma_{C,t}(i, *) \|, \| \mathit{\Gamma}_{C,t}(i, *) \| \} \}$.

Now, combining equation (A.3.46) with equations (A.3.32), (A.3.35), (A.3.36), (A.3.40), and (A.3.42), we obtain

$$\sup_t \frac{\left| \mathbb{M}^C_t \right|}{N} \leq \frac{c_3(\varepsilon) K_R W^2_{\max}}{m_c^2 N \gamma_c^2 \lambda^2_{K_R,\max}} \left\{ (6 + c'_w) \frac{b'}{\underline{\delta}'^{1/2}} + \frac{2K_C}{\underline{\delta}'} (\sqrt{2 P_{\max} r s} + 2P_{\max}) + \frac{NL}{\underline{\delta}' \cdot \ell!} \left(\frac{r}{T} \right)^\beta \right\}^2$$

where $c_3(\varepsilon) = 2^7 (2+\varepsilon)^2$, $b' = \{ 3 \log(16N/\epsilon) \}^{1/2}$ and $\lambda_{K_R,\max} = \max_t \{ \lambda_{K_R,t} \}$. ■