OLS Asymptotics Along Drifting Parameter Sequences for AR(2) Processes*

Kyle Shin

Due 3 April 2024

Abstract

This paper solves OLS asymptotic behavior for an AR(2) model with drifting autoregressive parameter sequences. We make extensive use of results regarding AR(1) asymptotics under similar drifting parameter sequences. Then, we attempt to use our results, follow Andrews et al. (2020), and construct a uniform confidence set for the AR(2) parameters that has correct asymptotic size regardless of the true value of the parameters, making them robust in finite samples. However, we show the failure of key continuity assumptions that bars us from using this method.

1 Introduction

AR(p) processes help model important economic variables with dependence on its past values, such as inflation and stock prices. However, inference for parameters is quite tricky, even in the simplest AR(1) case, since we have structured dependence among our observations. The basic AR(1) model with autoregressive parameter ρ is as follows:

$$X_t = \rho X_{t-1} + \varepsilon_t.$$

Results for the stationary case ($|\rho| < 1$) and the explosive case ($\rho > 1$) have been known since at latest Anderson (1959). Phillips (1987a) solved the unit root case ($\rho = 1$) using the functional central limit theorem à la Donsker (1951).

Even before the unit root case was solved, many researchers noticed the poor finite-sample behavior of the OLS estimator when ρ is close to 1, including Phillips (1977), who demonstrated poor finite-sample approximation of the limit distribution when $\rho > 0.8$. This behavior becomes apparent when looking at the OLS distributions for $\rho < 1$ and $\rho = 1$: in the stationary case, the OLS estimator converges with rate $\sqrt{n}$ to a standard normal distribution, whereas in the unit root case, we have n-convergence to the non-Gaussian unit root distribution, which is a functional of Brownian motion. To address this issue, Phillips

*Thank you to my advisor Max Cytrynbaum for guidance and support. Also, thank you to Tassos Magdalinos for providing the inspiration and the technical tools necessary for this topic in Econ553, Don Andrews for helpful conversations and insight, and Peter Phillips for additional help. All errors are my own.

(1987b) studied local-to-unity processes, where the parameter ρ_n drifts to 1 as a function of the sample size at rate n, to understand finite-sample behavior near 1 better. Studying drifting parameter sequences of AR(1) with different convergence rates continued from Chan and Wei (1987) through Giraitis and Phillips (2006) and Phillips and Magdalinos (2007).

The problem of poor finite-sample behavior for ρ near 1 falls under the general umbrella of uniform inference. The goal is to find a confidence set C such that C has the desired $1-\alpha$ asymptotic coverage probability uniformly across the parameter space: $\lim_{n\to\infty} \inf_{\rho} Pr(\rho \in C) \geq 1-\alpha$. A uniform confidence set is difficult to construct for the AR(1) parameter because the limit distribution when $\rho = 1$ differs vastly from the stationary and explosive limit distribution. Early attempts at a uniform confidence set for ρ come from Stock (1991) and Andrews (1993), of which a modified version of the former was proved to be correct by Mikusheva (2007). More recently, Andrews et al. (2020) provided a general method for constructing uniform confidence sets directly using the drifting parameter sequence asymptotics. In particular, for AR(1), since the discontinuity in the limit distribution occurs when $\rho = 1$, one must look at drifting parameter sequences ρ_n converging to 1 at different rates: these asymptotics continuously bridge the unit root distribution to the stationary distribution. Then, the critical values between the distributions vary continuously, which is the key requirement for the method to work.

This paper tries to use the method in Andrews et al. (2020) to build a uniform confidence set for the autoregressive parameters ϕ_1 and ϕ_2 in the AR(2) model:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t.$$

We study limit theory under drifting parameter sequences ϕ_{1n} and ϕ_{2n}. To our knowledge, no work has been done on AR(2) with these drifting parameter sequences. We use the convergence results in Phillips (1987b), Giraitis and Phillips (2006), and Phillips and Magdalinos (2007) for AR(1) and extend it to the AR(2) case by using either a diagonalization or Jordan form argument. When we can diagonalize, our AR(2) process becomes two AR(1) processes, but when we use the Jordan form, we obtain one AR(1) process and another AR(1) process with an AR(1) innovation. However, we show that the asymptotics do not continuously bridge distributions between different cases as in the AR(1) case. Hence, we cannot use Andrews et al. (2020) to build a uniform confidence set.

Somewhat related work for inference of AR(2) parameters can be found in $I(2)$ literature, such as in Dickey and Fuller (1979), Haldrup and Lildholdt (2002), and Haldrup (2002). However, tests like augmented Dickey-Fuller only test for the presence of a unit root, and they don't provide the full asymptotics of different cases of drifting parameter sequences as we do.

The outline of the paper is as follows. In Section 2, we review the martingale central limit theorem, the functional central limit theorem, and asymptotics for the AR(1) model. In Section 3, we solve asymptotics for different cases of drifting parameters in the AR(2) case, depending on whether we diagonalize or use a Jordan form. Finally, we discuss implications for uniform inference in Section 4. The Appendix contains some proof details.

2 Preliminaries

2.1 Central Limit Theorems

We state different central limit theorems that we will use. First, we have the vector martingale CLT, which can be found in most measure-theoretic probability textbooks like Pollard (1984). The vector martingale CLT will help us prove joint convergence of many different quantities.

Theorem 2.1. (Vector Martingale Central Limit Theorem)
Let $\{\xi_{n,j} : 1 \leq j \leq k_n\}$ be an $\mathcal{F}_{n,j}$-martingale difference array in $\mathbb{R}^d$ that satisfies the following Lindeberg condition:

$$L_n(\delta) := \sum_{j=1}^{k_n} \mathbb{E}(\|\xi_{n,j}\|^2 \mathbf{1}\{\|\xi_{n,j}\| > \delta\}) \to 0$$

for all $\delta > 0$ as $n \to \infty$. Then, if

$$\sum_{j=1}^{k_n} \mathbb{E}(\xi_{n,j}\xi_{n,j}' | \mathcal{F}_{n,j-1}) \to_p H$$

as $n \to \infty$, and H is almost surely positive definite, we have that

$$\sum_{j=1}^{k_n} \xi_{n,j} \to_d H^{1/2} Z,$$

where $Z \sim \mathcal{N}(0, I_d)$ is independent of H.

Donsker (1951) proved the original functional central limit theorem, and Phillips (1987a) used the functional limit theorem as described in Herrndorf (1984). For additional references, see Pollard (1984) and Billingsley (1999).

Theorem 2.2. (Functional Central Limit Theorem)
Let $(\varepsilon_j)_{j\in\mathbb{N}} \overset{\text{iid}}{\sim} (0, \sigma^2)$, $S_k = \sum_{j=1}^{k} \varepsilon_j$, and $W(t)$ be a standard Brownian motion on $[0, 1]$.

(i) There exists a sequence $(\tilde{S}_k)_{k\in\mathbb{N}}$ on the same probability space as $W(t)$ such that $\tilde{S}_k =_d S_k$ for all k, and letting $\tilde{S}_n(t) := \sigma^{-1} n^{-1/2} \tilde{S}_{\lfloor nt \rfloor}$, we have:

$$\sup_{t\in[0,1]} |\tilde{S}_n(t) - W(t)| \to_p 0 \quad \text{as} \quad n \to \infty.$$

(ii) Let $S_n(t) := \sigma^{-1} n^{-1/2} S_{\lfloor nt \rfloor}$. We have $S_n(t) \to_d W(t)$ on $D[0, 1]$.

(iii) Let $\Psi : D[0, 1] \to \mathbb{R}$ be continuous. Then $\Psi(S_n(t)) \to_d \Psi(W(t))$ on $\mathbb{R}$.

The following is a corollary of the above, found in Jacod and Shiryaev (2003).

Theorem 2.3. (Vector Martingale Functional Central Limit Theorem)

Let $\{\xi_{n,j} : 1 \leq j \leq k_n\}$ be an $\mathcal{F}_{n,j}$-martingale difference array in $\mathbb{R}^d$ that satisfies the following Lindeberg condition:

$$L_n(\delta) := \sum_{j=1}^{k_n} \mathbb{E}(\|\xi_{n,j}\|^2 \mathbf{1}\{\|\xi_{n,j}\| > \delta\}) \to 0$$

for all $\delta > 0$ as $n \to \infty$. Then, if

$$\sum_{j=1}^{\lfloor k_n t \rfloor} \mathbb{E}(\xi_{n,j}\xi_{n,j}' | \mathcal{F}_{n,j-1}) \to_p tH$$

for all t as $n \to \infty$, and H is almost surely positive definite non-random matrix, we have that

$$\sum_{j=1}^{\lfloor k_n t \rfloor} \xi_{n,j} \to_d H^{1/2} W(t),$$

on $D[0,1]$, where $W(t) = (W_1(t), \ldots, W_d(t))'$, and $\{W_1(t), \ldots, W_d(t)\}$ are independent standard Brownian motions on $[0,1]$.

Again, the vector martingale FCLT helps us with joint convergence.

2.2 AR(1) Results

Crucial to our analysis is a complete understanding of the AR(1) asymptotic behavior under drifting parameter sequences. Suppose we have an AR(1) process

$$X_t = \rho_n X_{t-1} + \varepsilon_t,$$

with a drifting, sample-size dependent parameter sequence $(\rho_n)_{n\in\mathbb{N}}$, $(\varepsilon_t) \overset{\text{iid}}{\sim} (0, \sigma^2)$, $X_0 = 0$, and $t \in \{1, \ldots, n\}$. We employ the same Assumption 1 as Magdalinos and Petrova (2022), except restricting to $\rho_n \to 1$. The stationary case ($\rho_n \to \rho < 1$) and the explosive case ($\rho_n \to \rho > 1$) can be solved similarly to the mildly stationary case and the mildly explosive case, respectively.

Assumption 1. *The sequence $\{\rho_n\}_{n\in\mathbb{N}}$ satisfies $\rho_n \to 1$. Furthermore, assume that the following limit exists:*

$$c := \lim_{n\to\infty} n(\rho_n - 1) \in [-\infty, \infty].$$

Make the same categorizations as Magdalinos and Petrova (2022) as well. $\{\rho_n\}_{n\in\mathbb{N}}$ is:

- $C(i)$: *mildly stationary* if $c = -\infty$
- $C(ii)$: *local-to-unity* if $c \in \mathbb{R}$

- $C(iii)$: *mildly explosive* if $c = +\infty$.

We let $\rho_n \in C(i)$ to mean that ρ_n is mildly stationary, and similarly for the other cases.

We have the following AR(1) theorems. The first result for the mildly explosive case is due to Giraitis and Phillips (2006). The rate of convergence for the OLS estimator is $(1-\rho_n^2)^{-1/2}\sqrt{n}$, which is between $\sqrt{n}$, the convergence rate for the stationary case, and n, the convergence rate for the unit root case.

Theorem 2.4. *If $\rho_n \in C(i)$, we have:*

$$(1-\rho_n^2)\frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2 \to_p \sigma^2,$$
$$(1-\rho_n^2)^{1/2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{t-1}\varepsilon_t \to_d \mathcal{N}(0,\sigma^4),$$
$$(1-\rho_n^2)^{-1/2}\sqrt{n}(\hat{\rho}_n - \rho_n) \to_d \mathcal{N}(0,1).$$

Phillips (1987b) proved the local-to-unity case has OLS rate of convergence n.

Theorem 2.5. *If $\rho_n \in C(ii)$, we have:*

$$\begin{bmatrix} n^{-2}\sum_{t=1}^{n} X_{t-1}^2 \\ n^{-1}\sum_{t=1}^{n} X_{t-1}\varepsilon_t \end{bmatrix} \to_d \begin{bmatrix} \sigma^2\int_0^1 J_c(t)^2 dt \\ \sigma^2\int_0^1 J_c(t) dW(t) \end{bmatrix},$$
$$n(\hat{\rho}_n - \rho_n) \to_d \frac{\int_0^1 J_c(t)dW(t)}{\int_0^1 J_c(t)^2 dt}.$$

where we have the Ornstein-Uhlenbeck process

$$J_c(t) = \int_0^t e^{c(t-r)} dW(r).$$

Finally, for the mildly explosive case, we turn to Phillips and Magdalinos (2007), who proved a convergence rate $(\rho_n^2 - 1)^{-1}\rho_n^n$, which can be shown to be faster than n (see Lemma C.3).

Theorem 2.6. *If $\rho_n \in C(iii)$, we have:*

$$\begin{bmatrix} (\rho_n^2-1)^2\rho_n^{-2n}\sum_{t=1}^{n} X_{t-1}^2 \\ (\rho_n^2-1)\rho_n^{-n}\sum_{t=1}^{n} X_{t-1}\varepsilon_t \end{bmatrix} \to_d \begin{bmatrix} Z^2 \\ YZ \end{bmatrix},$$
$$(\rho_n^2-1)^{-1}\rho_n^n(\hat{\rho}_n - \rho_n) \to_d \textit{Cauchy},$$

where

$$Z = \lim_{n\to\infty} Z_n, \quad Z_n = -(\rho_n^2-1)^{1/2}\sum_{j=1}^{n}\rho_n^{-j}\varepsilon_j = -(\rho_n^2-1)^{1/2}\rho_n^{-n}X_n,$$
$$Y = \lim_{n\to\infty} Y_n, \quad Y_n = -(\rho_n^2-1)^{1/2}\sum_{j=1}^{n}\rho_n^{-(n-j+1)}\varepsilon_j,$$

and Y and Z are independent and $Y =_d Z =_d \mathcal{N}(0,\sigma^2)$.

The beauty of these results is that we can see the continuity of the limit distributions. Let our starting point be the local-to-unity case. As $c \to -\infty$, ρ_n tends to a mildly stationary sequence having a standard normal limit distribution. The local-to-unity distribution also converges to a standard normal distribution as $c \to -\infty$ by Phillips (1987b). Similarly, as $c \to \infty$, ρ_n tends to a mildly explosive sequence having a Cauchy limit distribution. The local-to-unity distribution also converges to a Cauchy distribution as $c \to \infty$ by Phillips (1987b) again. In both cases, the critical values as $c \to \pm\infty$ vary continuously as we travel from one parameter regime to another. Andrews et al. (2020) take advantage of this continuity to form a uniform confidence set by inverting a test: see (2.17) and Assumptions C1 and C2 in the paper.

3 Main Results

3.1 AR(2) Model

Now, consider an AR(2) process

$$y_t = \phi_{1n} y_{t-1} + \phi_{2n} y_{t-2} + \varepsilon_t, \tag{1}$$

with drifting parameter sequences $(\phi_{1n})_{n\in\mathbb{N}}$ and $(\phi_{2n})_{n\in\mathbb{N}}$, $(\varepsilon_t) \overset{\text{iid}}{\sim} (0, \sigma^2)$, $y_{-1} = y_0 = 0$, and $t \in \{1, \ldots, n\}$. Written in companion form, we have

$$x_t = R_n x_{t-1} + u_t, \tag{2}$$

where

$$x_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}, \; R_n = \begin{bmatrix} \phi_{1n} & \phi_{2n} \\ 1 & 0 \end{bmatrix}, \; u_t = \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}.$$

Let the eigenvalues of R_n be ρ_{1n} and ρ_{2n}. We impose an assumption:

Assumption 2. *R_n has real eigenvalues.*

Simple calculations reveal that

$$\rho_{1n} + \rho_{2n} = \phi_{1n}, \quad \rho_{1n}\rho_{2n} = -\phi_{2n}. \tag{3}$$

Assume that $\{\rho_{1n}\}_{n\in\mathbb{N}}$ and $\{\rho_{2n}\}_{n\in\mathbb{N}}$ satisfy Assumption 1 with limits c_1 and c_2. We are interested in estimating $\phi_n = (\phi_{1n}, \phi_{2n})' = R_n' e_1$, where $e_1 = (1, 0)'$. Naive OLS gives us

$$\hat{R}_n = \big(\sum_{t=1}^{n} x_t x_{t-1}'\big)\big(\sum_{t=1}^{n} x_{t-1} x_{t-1}'\big)^{-1},$$

so noting that $u_t' e_1 = \varepsilon_t$, we have:

$$\begin{aligned} \hat{\phi}_n - \phi_n &= (\hat{R}_n - R_n)' e_1 \\ &= \big(\sum_{t=1}^{n} x_{t-1} x_{t-1}'\big)^{-1}\big(\sum_{t=1}^{n} x_{t-1}\varepsilon_t\big). \end{aligned} \tag{4}$$

We introduce a rotation matrix to use our complete knowledge of AR(1) asymptotics. We consider two cases. On one hand, we have distinct eigenvalues, so we can diagonalize R_n. Else, we have repeated eigenvalues, so we can use a Jordan form on R_n.

3.2 Distinct Eigenvalues

In the distinct eigenvalues case, we can diagonalize R_n: let

$$R_n = T_n \Lambda_n T_n^{-1},$$

where T_n contains the linearly independent eigenvectors, and Λ_n is a diagonal matrix storing the eigenvalues ρ_{1n} and ρ_{2n}:

$$\begin{aligned} T_n &= \frac{1}{\rho_{1n} - \rho_{2n}} \begin{bmatrix} \rho_{1n} & \rho_{2n} \\ 1 & 1 \end{bmatrix}, \\ T_n^{-1} &= \begin{bmatrix} 1 & -\rho_{2n} \\ -1 & \rho_{1n} \end{bmatrix}, \\ \Lambda_n &= \begin{bmatrix} \rho_{1n} & 0 \\ 0 & \rho_{2n} \end{bmatrix}. \end{aligned}$$

We can calculate T_n easily using (3). Do the transformation $X_t = T_n^{-1} x_t$ and $U_t = T_n^{-1} u_t$:

$$X_t = \Lambda_n X_{t-1} + U_t, \tag{5}$$

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \rho_{1n} & 0 \\ 0 & \rho_{2n} \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ -\varepsilon_t \end{bmatrix}.$$

We see that X_{1n} and X_{2n} are just AR(1) processes, which greatly aids our analysis. Then,

$$T_n'(\hat{\phi}_n - \phi_n) = \big(\sum_{t=1}^{n} X_{t-1} X_{t-1}' \big)^{-1} \big(\sum_{t=1}^{n} X_{t-1} \varepsilon_t \big). \tag{6}$$

We will have a normalizing diagonal matrix D_n, so we can write the OLS as follows:

$$D_n T_n'(\hat{\phi}_n - \phi_n) = \big(D_n^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}' D_n^{-1} \big)^{-1} \big(D_n^{-1} \sum_{t=1}^{n} X_{t-1} \varepsilon_t \big). \tag{7}$$

The appropriate normalization will allow us to use Theorems 2.4, 2.5, and 2.6 on the diagonal terms of the matrix and the terms of the vector on the RHS of (7). For the off-diagonal terms of the matrix, we have a very helpful decomposition:

$$(1 - \rho_{1n}\rho_{2n}) \sum_{t=1}^{n} X_{1,t-1} X_{2,t-1} = -X_{1,n} X_{2,n} - \rho_{1n} \sum_{t=1}^{n} X_{1,t-1} \varepsilon_t + \rho_{2n} \sum_{t=1}^{n} X_{2,t-1} \varepsilon_t - \sum_{t=1}^{n} \varepsilon_t^2. \tag{8}$$

We presently begin solving asymptotics. There are six primary cases that depend on the convergence behavior of ρ_{1n} and ρ_{2n} outlined in Assumption 1 (we have six rather than nine cases since we account for symmetry).

3.2.1 $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(i)$, $q \neq 1$

We need an additional specification here. Let

$$q_n := \frac{1-\rho_{1n}^2}{1-\rho_{2n}^2}, \quad q := \lim_{n\to\infty} q_n \in [0,\infty] \tag{9}$$

and

$$Q_{n,i} := \frac{(1-\rho_{1n}^2)^{1/2}(1-\rho_{1n}^2)^{1/2}}{1-\rho_{1n}\rho_{2n}}, \quad Q_i := \lim_{n\to\infty} Q_{n,i} \in [0,\infty). \tag{10}$$

Some calculations reveal that

$$Q_{n,i} = (1+\rho_{2n})\frac{q_n^{1/2}}{1+\frac{1+\rho_{1n}}{1+\rho_{2n}}q_n\rho_{2n}} \implies Q_i = (1+\rho_2)\frac{q^{1/2}}{1+q\frac{1+\rho_2}{1+\rho_1}\rho_2}.$$

Thus, if $q = 0$ or $q = \infty$, $Q_i = 0$. This occurs ρ_{1n} and ρ_{2n} converge at different rates. Otherwise, if $q \neq 0, \infty$, then $Q_i = 1$ iff $q = 1$: they converge at very similar rates. In this section, we require that they converge at fairly different rates ($q \neq 1$).

Theorem 3.1. *Suppose $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, and $\rho_{2n} \in C(i)$. Let $D_{n,(i,i)}$ be the following diagonal matrix:*

$$D_{n,(i,i)} = \begin{bmatrix} (1-\rho_{1n}^2)^{-1/2}n^{1/2} & 0 \\ 0 & (1-\rho_{2n}^2)^{-1/2}n^{1/2} \end{bmatrix}.$$

Define

$$V_{(i,i)} = \begin{bmatrix} 1 & -Q_i \\ -Q_i & 1 \end{bmatrix}.$$

Then, when $q \neq 1$,

$$D_{n,(i,i)}^{-1}\sum_{t=1}^{n} X_{t-1}X_{t-1}' D_{n,(i,i)}^{-1} \to_p \sigma^2 V_{(i,i)}, \tag{11}$$

$$D_{n,(i,i)}^{-1}\sum_{t=1}^{n} X_{t-1}\varepsilon_t \to_d \mathcal{N}(0, \sigma^4 V_{(i,i)}), \tag{12}$$

$$D_{n,(i,i)}T_n'(\hat{\phi}_n - \phi_n) \to_d \mathcal{N}(0, V_{(i,i)}^{-1}). \tag{13}$$

Diagonalization yields two mildly stationary AR(1) processes. It is no surprise that the diagonal entries of $D_{n,(i,i)}$ have the mildly stationary normalizations from Theorem 2.4. Also, when ρ_{1n} and ρ_{2n} converge at different rates, $Q_i = 0$, so the components of the asymptotic distribution are standard normal distributions. The AR(1) mildly stationary case also has a Gaussian limit distribution, so we can imagine this case as being two independent mildly stationary AR(1) processes.

Proof. We have:

$$D_{n,(i,i)}^{-1}\sum_{t=1}^{n}X_{t-1}X_{t-1}'D_{n,(i,i)}^{-1}=\begin{bmatrix}(1-\rho_{1n}^2)n^{-1}\sum_{t=1}^{n}X_{1,t-1}^2 & *\\ (1-\rho_{1n}^2)^{1/2}(1-\rho_{2n}^2)^{1/2}n^{-1}\sum_{t=1}^{n}X_{1,t-1}X_{2,t-1} & (1-\rho_{2n}^2)n^{-1}\sum_{t=1}^{n}X_{2,t-1}^2\end{bmatrix},$$

Theorem 2.4 takes care of the diagonal elements. We focus on the off-diagonal elements. Using (8), we have:

$$\begin{aligned}(1-\rho_{1n}^2)^{1/2}(1-\rho_{1n}^2)^{1/2}\frac{1}{n}\sum_{t=1}^{n}X_{1,t-1}X_{2,t-1}=Q_{n,i}\Big(&-\frac{1}{n}X_{1,n}X_{2,n}-\frac{\rho_{1n}}{n}\sum_{t=1}^{n}X_{1,t-1}\varepsilon_t\\ &+\frac{\rho_{2n}}{n}\sum_{t=1}^{n}X_{2,t-1}\varepsilon_t-\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2\Big)\end{aligned}$$

By Giraitis and Phillips (2006), we know that $n^{-1/2}X_{1,n}\to_{L_1}0$ and $n^{-1/2}X_{2,n}\to_{L_1}0$, so the first term in the brackets on the RHS is $o_p(1)$. Next, for the second term, since $n(1-\rho_{1n})\to\infty$ by assumption, then $n(1-\rho_{1n}^2)\to\infty$, so using Theorem 2.4, we have that

$$\frac{1}{n}\sum_{t=1}^{n}X_{1,t-1}\varepsilon_t=\Big(\frac{1}{(1-\rho_{1n}^2)^{1/2}\sqrt{n}}\Big)(1-\rho_{1n}^2)^{1/2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}X_{1,t-1}\varepsilon_t\to_p 0.$$

The third term goes to 0 in a similar fashion. Finally, for the last term, the WLLN implies $n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\to_p\sigma^2$. Thus the off-diagonal elements converge to $-Q_i\sigma^2$ in probability. Then,

$$D_{n,(i,i)}^{-1}\sum_{t=1}^{n}X_{t-1}X_{t-1}'D_{n,(i,i)}^{-1}\to_p\sigma^2\begin{bmatrix}1 & -Q_i\\ -Q_i & 1\end{bmatrix}=\sigma^2V_{(i,i)}.$$

Note that $V_{(i,i)}>0$ since we assume $q\neq 1\implies Q_i\neq 1$.

Next, we have

$$D_{n,(i,i)}^{-1}\sum_{t=1}^{n}X_{t-1}\varepsilon_t=\begin{bmatrix}(1-\rho_{1n}^2)^{1/2}n^{-1/2}\sum_{t=1}^{n}X_{1,t-1}\varepsilon_t\\ (1-\rho_{2n}^2)^{1/2}n^{-1/2}\sum_{t=1}^{n}X_{2,t-1}\varepsilon_t\end{bmatrix}.$$

We do a martingale CLT for convergence. We have the following martingale difference array with $\mathcal{F}_{n,t}=\sigma(\varepsilon_t,\dots,\varepsilon_1)$ (from now on, the filtration will be self-evident, and the martingale difference property is obvious):

$$\xi_{n,t}=\begin{bmatrix}(1-\rho_{1n}^2)^{1/2}n^{-1/2}X_{1,t-1}\varepsilon_t\\ (1-\rho_{2n}^2)^{1/2}n^{-1/2}X_{2,t-1}\varepsilon_t\end{bmatrix}=\begin{bmatrix}(1-\rho_{1n}^2)^{1/2}n^{-1/2}X_{1,t-1}\\ (1-\rho_{2n}^2)^{1/2}n^{-1/2}X_{2,t-1}\end{bmatrix}\varepsilon_t \tag{14}$$

Then, we have

$$\begin{aligned}&\sum_{t=1}^{n}\mathbb{E}(\xi_{n,t}\xi_{n,t}'|\mathcal{F}_{n,t-1})\\ &=\sigma^2\sum_{t=1}^{n}\begin{bmatrix}(1-\rho_{1n}^2)n^{-1}\mathbb{E}(X_{1,t-1}^2|\mathcal{F}_{n,t-1}) & *\\ (1-\rho_{1n}^2)^{1/2}(1-\rho_{2n}^2)^{1/2}n^{-1}\mathbb{E}(X_{1,t-1}X_{2,t-1}|\mathcal{F}_{n,t-1}) & (1-\rho_{2n}^2)n^{-1}\mathbb{E}(X_{2,t-1}^2|\mathcal{F}_{n,t-1})\end{bmatrix}\\ &=\sigma^2\begin{bmatrix}(1-\rho_{1n}^2)n^{-1}\sum_{t=1}^{n}X_{1,t-1}^2 & *\\ (1-\rho_{1n}^2)^{1/2}(1-\rho_{2n}^2)^{1/2}n^{-1}\sum_{t=1}^{n}X_{1,t-1}X_{2,t-1} & (1-\rho_{2n}^2)n^{-1}\sum_{t=1}^{n}X_{2,t-1}^2\end{bmatrix}\\ &\to_p\sigma^4V_{(i,i)}\end{aligned}$$

where we used our work from before. The Lindeberg condition holds by Lemma A.1. We conclude that

$$D_{n,(i,i)}^{-1} \sum_{t=1}^{n} X_{t-1}\varepsilon_t \to_d \mathcal{N}(0, \sigma^4 V_{(i,i)}).$$

The continuous mapping theorem with the items just proved complete the final statement of the theorem by (7). $\square$

3.2.2 $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(ii)$

Theorem 3.2. *Suppose $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, and $\rho_{2n} \in C(ii)$. Let $D_{n,(i,ii)}$ be the following diagonal matrix:*

$$D_{n,(i,ii)} = \begin{bmatrix} (1-\rho_{1n}^2)^{-1/2} n^{1/2} & 0 \\ 0 & n \end{bmatrix}.$$

Then,

$$\left[D_{n,(i,ii)}^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}' D_{n,(i,ii)}^{-1}, D_{n,(i,ii)}^{-1} \sum_{t=1}^{n} X_{t-1} \varepsilon_t \right]' \to_d \left[\sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & \int_0^1 J_{c_2}(t)^2 dt \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathcal{N}(0,1) \\ \int_0^1 J_{c_2}(t) dW(t) \end{bmatrix} \right]', \tag{15}$$

$$D_{n,(i,ii)} T_n' (\hat{\phi}_n - \phi_n) \to_d \begin{bmatrix} \mathcal{N}(0,1) \\ (\int_0^1 J_{c_2}(t)^2 dt)^{-1} \int_0^1 J_{c_2}(t) dW(t) \end{bmatrix}. \tag{16}$$

Diagonalization yields a mildly stationary AR(1) process and a local-to-unity AR(1) process, and the components of the AR(2) limit distribution are the respective AR(1) limit distributions. The diagonal entries of $D_{n,(i,ii)}$ are the normalizations from Theorem 2.4 and Theorem 2.5.

Proof. We have:

$$D_{n,(i,ii)}^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}' D_{n,(i,ii)}^{-1} = \begin{bmatrix} (1-\rho_{1n}^2) n^{-1} \sum_{t=1}^{n} X_{1,t-1}^2 & * \\ (1-\rho_{1n}^2)^{1/2} n^{-3/2} \sum_{t=1}^{n} X_{1,t-1} X_{2,t-1} & n^{-2} \sum_{t=1}^{n} X_{2,t-1}^2 \end{bmatrix}.$$

The diagonal elements of the matrix are taken care of by Theorem 2.4 and 2.5. We focus on the off-diagonal elements. Using (8), we have:

$$\begin{aligned} (1-\rho_{1n}^2)^{1/2} \frac{1}{n^{3/2}} \sum_{t=1}^{n} X_{1,t-1} X_{2,t-1} = (1-\rho_{1n}^2)^{1/2} \frac{1}{n^{3/2}} \frac{1}{1-\rho_{1n}\rho_{2n}} \Big(&- X_{1,n} X_{2,n} - \rho_{1n} \sum_{t=1}^{n} X_{1,t-1} \varepsilon_t \\ &+ \rho_{2n} \sum_{t=1}^{n} X_{2,t-1} \varepsilon_t - \sum_{t=1}^{n} \varepsilon_t^2 \Big) \end{aligned}$$

By Giraitis and Phillips (2006), we know that $n^{-1/2} X_{1,n} \to_{L_1} 0$, and by Phillips (1987b), we know that $\sigma^{-1} n^{-1/2} X_n \to_d J_{c_2}(1)$. Thus, the first term in the brackets on the RHS is $O_p(n^{1/2})$. For the second term, Theorem 2.4 implies that the convergence rate is $(1-\rho_{1n}^2)^{-1/2} n^{1/2} \in (n^{1/2}, n)$. The third term is $O_p(n)$ by Theorem 2.5. The last term is $O_p(n)$ by WLLN. Note that $1-\rho_{1n}\rho_{2n} \sim 1-\rho_{1n}^2$ since ρ_{1n} converges slower than ρ_{2n} by assumption. Thus

$$\begin{aligned} (1-\rho_{1n}^2)^{1/2} \frac{1}{n^{3/2}} \sum_{t=1}^{n} X_{1,t-1} X_{2,t-1} &= (1-\rho_{1n}^2)^{1/2} \frac{1}{n^{3/2}} \frac{1}{1-\rho_{1n}\rho_{2n}} O_p(n) \\ &= O_p((1-\rho_{1n}^2)^{-1/2} n^{-1/2}) \\ &= o_p(1) \end{aligned}$$

Thus the off-diagonal elements converge to 0 in probability. Since the upper-left element converges in probability, we conclude that

$$D_{n,(i,ii)}^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}' D_{n,(i,ii)}^{-1} \to_d \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & \int_0^1 J_{c_2}(t)^2 dt \end{bmatrix}.$$

Clearly, this is positive-definite.

Next, we have:

$$D_{n,(i,ii)}^{-1} \sum_{t=1}^{n} X_{t-1} \varepsilon_t = \begin{bmatrix} (1-\rho_{1n}^2)^{1/2} n^{-1/2} \sum_{t=1}^{n} X_{1,t-1}\varepsilon_t \\ n^{-1} \sum_{t=1}^{n} X_{2,t-1}\varepsilon_t \end{bmatrix}.$$

We must use the vector martingale FCLT for joint convergence of all terms in (7). Let

$$\xi_{n,j} = \begin{bmatrix} n^{-1/2}(1-\rho_{1n}^2)^{1/2} X_{1,j-1}\varepsilon_j \\ n^{-1/2}\varepsilon_j \end{bmatrix} = \frac{1}{\sqrt{n}} \begin{bmatrix} (1-\rho_{1n}^2)^{1/2} X_{1,j-1} \\ 1 \end{bmatrix} \varepsilon_j. \tag{17}$$

Then, we have

$$\sum_{j=1}^{\lfloor nt \rfloor} \mathbb{E}(\xi_{n,j}\xi_{n,j}' | \mathcal{F}_{n,j-1}) = \sigma^2 \begin{bmatrix} (1-\rho_{1n}^2) n^{-1} \sum_{j=1}^{\lfloor nt \rfloor} X_{1,j-1}^2 & * \\ (1-\rho_{1n}^2)^{1/2} n^{-1} \sum_{j=1}^{\lfloor nt \rfloor} X_{1,j-1} & n^{-1}\lfloor nt \rfloor \end{bmatrix}.$$

The top-left element goes to $t\sigma^2$ by a re-scaling of Theorem 2.4. Clearly the bottom-right element goes to t. We show that the off-diagonal elements go to 0. We have

$$\begin{aligned} X_{1,t} &= \rho_{1n} X_{1,t-1} + \varepsilon_t \\ \implies \sum_{j=1}^{\lfloor nt \rfloor} X_{1,j} &= \rho_{1n} \sum_{j=1}^{\lfloor nt \rfloor} X_{1,j-1} + \sum_{j=1}^{\lfloor nt \rfloor} \varepsilon_j \\ \implies (1-\rho_{1n}) \sum_{j=1}^{\lfloor nt \rfloor} X_{1,j} &= -X_{1,\lfloor nt \rfloor} + \sum_{j=1}^{\lfloor nt \rfloor} \varepsilon_j \\ \implies (1-\rho_{1n}) \sum_{j=1}^{\lfloor nt \rfloor} X_{1,j} &= O_p(n^{1/2}) \end{aligned}$$

since $n^{-1/2} X_{1,nt} \to_{L_1} 0$ and the FCLT implies $\sigma^{-1} n^{-1/2} \sum_{j=1}^{\lfloor nt \rfloor} \to_d W(t)$ on $D[0,1]$. Then, since $1-\rho_{1n} \sim 1-\rho_{1n}^2$, we have:

$$\begin{aligned} (1-\rho_{1n}^2)^{1/2} n^{-1} \sum_{j=1}^{\lfloor nt \rfloor} X_{1,j-1} &= O_p\Big((1-\rho_{1n}^2)^{1/2} n^{-1} n^{1/2} (1-\rho_{1n})^{-1}\Big) \\ &= O_p\Big((1-\rho_{1n}^2)^{-1/2} n^{-1/2}\Big) \\ &= o_p(1). \end{aligned}$$

Thus,

$$\sum_{t=1}^{\lfloor nt \rfloor} \mathbb{E}(\xi_{n,j}\xi'_{n,j}|\mathcal{F}_{n,j-1}) = \sigma^2 \begin{bmatrix} \sigma^2 t & 0 \\ 0 & t \end{bmatrix} = t \begin{bmatrix} \sigma^4 & 0 \\ 0 & \sigma^2 \end{bmatrix}.$$

The Lindeberg condition holds by Lemma A.2, so the vector martingale FCLT implies that

$$\sum_{j=1}^{\lfloor nt \rfloor} \xi_{n,j} \to_d \begin{bmatrix} \sigma^2 W_1(t) \\ \sigma W(t) \end{bmatrix}$$

on $D[0,1]$, where $W_1(t)$ and $W(t)$ are independent standard Brownian motions. Next, note that $n^{-1}\sum_{t=1}^{n} X_{2,t-1}\varepsilon_t$ and $n^{-2}\sum_{t=1}^{n} X_{2,t-1}^2$ are continuous functionals of $\sum_{j=1}^{\lfloor nt \rfloor} n^{-1/2}\varepsilon_j$ by Phillips (1987b) using an integral argument. Thus,

$$\begin{bmatrix} (1-\rho_{1n}^2)^{1/2} n^{-1/2} \sum_{t=1}^{n} X_{1,t-1}\varepsilon_t \\ n^{-1}\sum_{t=1}^{n} X_{2,t-1}\varepsilon_t \\ n^{-2}\sum_{t=1}^{n} X_{2,t-1}^2 \end{bmatrix} \to_d \sigma^2 \begin{bmatrix} W_1(1) \\ \int_0^1 J_{c_2}(t)dW(t) \\ \int_0^1 J_{c_2}(t)^2 dt \end{bmatrix}$$

since these are just continuous functionals of $\sum_{j=1}^{\lfloor nt \rfloor} \xi_{n,j}$. Now, every term of (7) is a continuous function of the LHS of the above since $(1-\rho_{1n}^2)n^{-1}\sum_{t=1}^{n} X_{1,t-1}^2$ is a function of $(1-\rho_{1n}^2)^{1/2} n^{-1/2}\sum_{t=1}^{n} X_{1,t-1}\varepsilon_t$ by Giraitis and Phillips (2006). Therefore, the continuous mapping theorem completes the proof:

$$\begin{aligned} D_{n,(i,ii)} T_n'(\hat{\phi}_n - \phi_n) &\to_d \left(\sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & \int_0^1 J_{c_2}(t)^2 dt \end{bmatrix} \right)^{-1} \sigma^2 \begin{bmatrix} \mathcal{N}(0,1) \\ \int_0^1 J_{c_2}(t)dW(t) \end{bmatrix} \\ &= \begin{bmatrix} \mathcal{N}(0,1) \\ (\int_0^1 J_{c_2}(t)^2 dt)^{-1} \int_0^1 J_{c_2}(t)dW(t) \end{bmatrix}. \end{aligned}$$

□

3.2.3 $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(iii)$

Theorem 3.3. *Suppose $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, and $\rho_{2n} \in C(iii)$. Let $D_{n,(i,iii)}$ be the following diagonal matrix:*

$$D_{n,(i,iii)} = \begin{bmatrix} (1-\rho_{1n}^2)^{-1/2} n^{1/2} & 0 \\ 0 & (\rho_{2n}^2 - 1)^{-1} \rho_{2n}^n \end{bmatrix}.$$

Then,

$$\left[D_{n,(i,iii)}^{-1} \sum_{t=1}^n X_{t-1} X_{t-1}' D_{n,(i,iii)}^{-1}, D_{n,(i,iii)}^{-1} \sum_{t=1}^n X_{t-1} \varepsilon_t \right]' \to_d \left[\begin{bmatrix} \sigma^2 & 0 \\ 0 & Z_2^2 \end{bmatrix}, \begin{bmatrix} \sigma^2 \mathcal{N}(0,1) \\ Y_2 Z_2 \end{bmatrix} \right]', \tag{18}$$

$$D_{n,(i,iii)} T_n' (\hat{\phi}_n - \phi_n) \to_d \begin{bmatrix} \mathcal{N}(0,1) \\ Z_2^{-1} Y_2 \end{bmatrix}, \tag{19}$$

where Z_2 and Y_2 follows the definition in Theorem 2.6 with innovation sequence $(-\varepsilon_t)_{t \in \mathbb{N}}$. Note that $Z_2^{-1} Y_2 =_d$ Cauchy.

Diagonalization again works very well: we get a mildly stationary AR(1) process and a mildly explosive AR(1) process, and the components of the limit distribution are simply the respective AR(1) limit distributions. The diagonal entries of $D_{n,(i,iii)}$ are simply the normalizations from Theorem 2.4 and 2.6 as expected.

Proof. We have:

$$D_{n,(i,iii)}^{-1} \sum_{t=1}^n X_{t-1} X_{t-1}' D_{n,(i,iii)}^{-1} = \begin{bmatrix} (1-\rho_{1n}^2) n^{-1} \sum_{t=1}^n X_{1,t-1}^2 & * \\ (1-\rho_{1n}^2)^{1/2} n^{-1/2} (\rho_{2n}^2 - 1) \rho_{2n}^{-n} \sum_{t=1}^n X_{1,t-1} X_{2,t-1} & (\rho_{2n}^2 - 1)^2 \rho_{2n}^{-2n} \sum_{t=1}^n X_{2,t-1}^2 \end{bmatrix}.$$

Theorem 2.4 and Theorem 2.6 takes care of the diagonal elements. We focus on the off-diagonal elements. Using (8), we have:

$$\begin{aligned} \frac{(1-\rho_{1n}^2)^{1/2}}{\sqrt{n}} (\rho_{2n}^2 - 1) \rho_{2n}^{-n} \sum_{t=1}^n X_{1,t-1} X_{2,t-1} &= \frac{(1-\rho_{1n}^2)^{1/2}}{\sqrt{n}} (\rho_{2n}^2 - 1) \rho_{2n}^{-n} \frac{1}{1 - \rho_{1n}\rho_{2n}} \\ &\quad \cdot \Big(- X_{1,n} X_{2,n} - \rho_{1n} \sum_{t=1}^n X_{1,t-1} \varepsilon_t \\ &\quad + \rho_{2n} \sum_{t=1}^n X_{2,t-1} \varepsilon_t - \sum_{t=1}^n \varepsilon_t^2 \Big). \end{aligned}$$

We know from previously that $X_{1,n} = o_p(n^{1/2})$, hence $X_{1,n} = O_p((1-\rho_{1n}^2)^{-1/2})$. Also, $X_{2,n} = O_p((\rho_{2n}^2 - 1)^{-1/2} \rho_{2n}^n)$ from Phillips and Magdalinos (2007). Then, $X_{1,n} X_{2,n} = O_p((1 - \rho_{1n}^2)^{-1/2} (\rho_{2n}^2 - 1)^{-1/2} \rho_{2n}^n)$. We know that the second term is $O_p(n)$. The third term is $O_p((\rho_{2n}^2 - 1)^{-1} \rho_{2n}^n)$ by Theorem 2.6. The last term is $O_p(n)$ by WLLN. n is a faster rate than $(1-\rho_{1n}^2)^{-1/2}$, so we can say that the rate of the bracketed term is

$$O_p\Big((\rho_{2n}^2 - 1)^{-1} \rho_{2n}^n + (1 - \rho_{1n}^2)^{-1/2} (\rho_{2n} - 1)^{-1/2} \rho_{2n}^n \Big).$$

Thus,

$$\frac{(1-\rho_{1n}^2)^{1/2}}{\sqrt{n}}(\rho_{2n}^2-1)\rho_{2n}^{-n}\sum_{t=1}^{n}X_{1,t-1}X_{2,t-1} = \frac{(1-\rho_{1n}^2)^{1/2}}{\sqrt{n}}(\rho_{2n}^2-1)\rho_{2n}^{-n}\frac{1}{1-\rho_{1n}\rho_{2n}}$$

$$\cdot O_p\Big((\rho_{2n}^2-1)^{-1}\rho_{2n}^{n} + (1-\rho_{1n}^2)^{-1/2}(\rho_{2n}-1)^{-1/2}\rho_{2n}^{n}\Big)$$

$$= O_p\Big((1-\rho_{1n}\rho_{2n})^{-1}n^{-1/2}[(1-\rho_{1n}^2)^{1/2}+(1-\rho_{2n}^2)^{1/2}]\Big)$$

$$= O_p\Big((1-\rho_{1n}\rho_{2n})^{-1/2}n^{-1/2}\Big)$$

$$= o_p(1),$$

where we use that $(1-\rho_{1n}\rho_{2n})^{1/2}$ and $(1-\rho_{1n}^2)^{1/2}+(1-\rho_{2n}^2)^{1/2}$ are the same rate (each is dominated by the slower one). Thus the diagonal terms converges to 0, so

$$D_{n,(i,iii)}^{-1}\sum_{t=1}^{n}X_{t-1}X_{t-1}'D_{n,(i,iii)}^{-1} = \begin{bmatrix}\sigma^2 & 0\\ 0 & Z_{2n}^2\end{bmatrix} + o_p(I_2),$$

where the bottom-right element is by Phillips and Magdalinos (2007).

Next, we have:

$$D_{n,(i,iii)}^{-1}\sum_{t=1}^{n}X_{t-1}\varepsilon_t = \begin{bmatrix}(1-\rho_{1n}^2)^{1/2}n^{-1/2}\sum_{t=1}^{n}X_{1,t-1}\varepsilon_t\\ (\rho_{2n}^2-1)\rho_{2n}^{-n}\sum_{t=1}^{n}X_{2,t-1}\varepsilon_t\end{bmatrix}$$

$$= \begin{bmatrix}(1-\rho_{1n}^2)^{1/2}n^{-1/2}\sum_{t=1}^{n}X_{1,t-1}\varepsilon_t\\ Y_{2n}Z_{2n}\end{bmatrix} + o_p(1),$$

where the second element is by Phillips and Magdalinos (2007). We will do a martingale CLT to prove joint convergence of all terms in (7). Let

$$\xi_{n,t} = \begin{bmatrix}(1-\rho_{1n}^2)^{1/2}n^{-1/2}X_{1,t-1}\\ (\rho_{2n}^2-1)^{1/2}\rho_{2n}^{-t}\\ (\rho_{2n}^2-1)^{1/2}\rho_{2n}^{-(n-t+1)}\end{bmatrix}\varepsilon_t, \implies \sum_{t=1}^{n}\xi_{n,t} = \begin{bmatrix}(1-\rho_{1n}^2)^{1/2}n^{-1/2}\sum_{t=1}^{n}X_{1,t-1}\varepsilon_t\\ Z_{2n}\\ Y_{2n}\end{bmatrix}. \tag{20}$$

Then, we have:

$$\sum_{t=1}^{n}\mathbb{E}(\xi_{n,t}\xi_{n,t}'|\mathcal{F}_{n,t-1}) = \sigma^2$$

$$\cdot\begin{bmatrix}(1-\rho_{1n}^2)n^{-1}\sum_{t=1}^{n}X_{1,t-1}^2 & * & *\\ (1-\rho_{1n}^2)^{1/2}(\rho_{2n}^2-1)^{1/2}n^{-1/2}\sum_{t=1}^{n}\rho_{2n}^{-t}X_{1,t-1} & (\rho_{2n}^2-1)\sum_{t=1}^{n}\rho_{2n}^{-2t} & *\\ (1-\rho_{1n}^2)^{1/2}(\rho_{2n}^2-1)^{1/2}n^{-1/2}\sum_{t=1}^{n}\rho_{2n}^{-(n-t+1)}X_{1,t-1} & (\rho_{2n}^2-1)\sum_{t=1}^{n}\rho_{2n}^{-n-1} & (\rho_{2n}^2-1)\sum_{t=1}^{n}\rho_n^{-2(n-t+1)}\end{bmatrix}$$

The (1,1) element converges in probability to σ^2 by Theorem 2.4. The (2,2) and (3,3) elements converge to 1 since they can be written as geometric series: to demonstrate, for the

(2,2) element, we have

$$(\rho_{2n}^2 - 1)\sum_{t=1}^{n} \rho_{2n}^{-2t} = (\rho_{2n}^2 - 1)\frac{\rho_{2n}^{-2}(1 - \rho_{2n}^{-2(n+1)})}{1 - \rho_{2n}^{-2}} = 1 - \rho_{2n}^{-2(n+1)} \to 1$$

by Lemma C.3. For the (2,1) element, we have:

$$\begin{aligned}
\left\| (1-\rho_{1n}^2)^{1/2}(\rho_{2n}^2-1)^{1/2} n^{-1/2} \sum_{t=1}^{n} \rho_{2n}^{-t} X_{1,t-1} \right\|_{L_1} &\leq (1-\rho_{1n}^2)^{1/2} \max_{1\leq t\leq n} \|X_{1,t-1}\|_{L_2} \\
&\quad \cdot (\rho_{2n}^2-1)^{1/2} n^{-1/2} \sum_{t=1}^{n} \rho_{2n}^{-t} \\
&= O_p(1)(\rho_{2n}^2-1)^{1/2} n^{-1/2} \frac{1-\rho_{2n}^{-n-1}}{\rho_{2n}-1} \\
&= O_p((\rho_{2n}^2-1)^{-1/2} n^{-1/2}) \\
&= o_p(1)
\end{aligned}$$

where the first inequality uses norm relations, the first equality holds since

$$\|X_{1,t}\|_{L_2}^2 = \|\sum_{j=0}^{n-1} \rho_t^j \varepsilon_{t-j+1}\|_{L_2}^2 = \sigma^2 \sum_{j=0}^{t-1} \rho_{2n}^{2j} = O((1-\rho_{1n}^2)^{-1}),$$

the second equality holds since $\rho_{2n}^2 - 1 \sim \rho_{2n} - 1$, and the last equality holds by assumption of ρ_{2n}.

The (3,1) element goes to 0 in the same manner. Finally, the (3,2) element is:

$$(\rho_{2n}^2 - 1)\sum_{t=1}^{n} \rho_{2n}^{-n-1} = n(\rho_{2n}^2 - 1)\rho_{2n}^{-n-1} \to 0$$

by Lemma C.3. Therefore, the conditional variance converges to $\sigma^2 diag(\sigma^2, 1, 1)$. The Lindeberg condition holds by Lemma A.3. So, we conclude that

$$\begin{bmatrix} (1-\rho_{1n}^2)^{1/2} n^{-1/2} \sum_{t=1}^{n} X_{1,t-1}\varepsilon_t \\ Z_{2n} \\ Y_{2n} \end{bmatrix} = \sum_{t=1}^{n} \xi_{n,t} \to_d \mathcal{N}\Big(0, \sigma^2 \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\Big).$$

Now, every term of (7) is a continuous function of the LHS of the above since $(1-\rho_{1n}^2) n^{-1} \sum_{t=1}^{n} X_{1,t-1}^2$ is a function of $(1-\rho_{1n}^2)^{1/2} n^{-1/2} \sum_{t=1}^{n} X_{1,t-1}\varepsilon_t$ by Giraitis and Phillips (2006). Therefore, the continuous mapping theorem completes the proof:

$$\begin{aligned}
D_{n,(i,iii)} T_n'(\hat{\phi}_n - \phi_n) &\to_d \left(\begin{bmatrix} \sigma^2 & 0 \\ 0 & Z_2^2 \end{bmatrix} \right)^{-1} \begin{bmatrix} \sigma^2 \mathcal{N}(0,1) \\ Y_2 Z_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{N}(0,1) \\ Z_2^{-1} Y_2 \end{bmatrix}.
\end{aligned}$$

□

3.2.4 $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(ii)$, $\rho_{2n} \in C(ii)$

Theorem 3.4. *Suppose $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(ii)$, and $\rho_{2n} \in C(ii)$. Let $D_{n,(ii,ii)}$ be the following diagonal matrix:*

$$D_{n,(ii,ii)} = \begin{bmatrix} n & 0 \\ 0 & n \end{bmatrix}.$$

Then,

$$\left[D_{n,(ii,ii)}^{-1} \sum_{t=1}^{n} X_{t-1}X_{t-1}' D_{n,(ii,ii)}^{-1}, D_{n,(ii,ii)}^{-1} \sum_{t=1}^{n} X_{t-1}\varepsilon_t\right]' \to_d \left[\begin{bmatrix} \int_0^1 J_{c_1}(t)^2 dt & \int_0^1 J_{c_1}(t)J_{c_2}(t)dt \\ \int_0^1 J_{c_1}(t)J_{c_2}(t)dt & \int_0^1 J_{c_2}(t)^2 dt \end{bmatrix}, \begin{bmatrix} \int_0^1 J_{c_1}(t)dW(t) \\ \int_0^1 J_{c_2}(t)dW(t) \end{bmatrix}\right]', \tag{21}$$

$$D_{n,(ii,ii)} T_n'(\hat{\phi}_n - \phi_n) \to_d \begin{bmatrix} \int_0^1 J_{c_1}(t)^2 dt & \int_0^1 J_{c_1}(t)J_{c_2}(t)dt \\ \int_0^1 J_{c_1}(t)J_{c_2}(t)dt & \int_0^1 J_{c_2}(t)^2 dt \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 J_{c_1}(t)dW(t) \\ \int_0^1 J_{c_2}(t)dW(t) \end{bmatrix}. \tag{22}$$

Diagonalization yields two local-to-unity AR(2) processes. However, rather than getting a clean limit distribution that has AR(1) local-to-unity asymptotics in the components, we get cross-terms. The normalization matrix $D_{n,(ii,ii)}$ has n in the diagonal entries, which is expected by Theorem 2.5.

Proof. We have:

$$D_{n,(ii,ii)}^{-1} \sum_{t=1}^{n} X_{t-1}X_{t-1}' D_{n,(ii,ii)}^{-1} = \begin{bmatrix} n^{-2}\sum_{t=1}^{n} X_{1,t-1}^2 & n^{-2}\sum_{t=1}^{n} X_{1,t-1}X_{2,t-1} \\ n^{-2}\sum_{t=1}^{n} X_{1,t-1}X_{2,t-1} & n^{-2}\sum_{t=1}^{n} X_{2,t-1}^2 \end{bmatrix}.$$

Theorem 2.5 takes care of the matrix's diagonal elements and the vector element. We focus on the off-diagonal elements. Following Phillips (1987b), we know that from the FCLT that

$$\begin{bmatrix} \sigma^{-1}n^{-1/2}X_{1,\lfloor nt \rfloor} \\ \sigma^{-1}n^{-1/2}X_{2,\lfloor nt \rfloor} \end{bmatrix} \to_d \begin{bmatrix} J_{c_1}(t) \\ J_{c_2}(t) \end{bmatrix}.$$

since both terms are continuous functionals of the same random walk $S_n(t) = \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} \varepsilon_j$ in the FCLT, where convergence is on $D[0,1]$. Note that integration is a continuous functional on the space $D[0,1]$. Then, the continuous mapping theorem implies

$$\begin{aligned} n^{-2}\sum_{t=1}^{n} X_{1,t-1}X_{2,t-1} &= n^{-2}\int_0^n X_{1,\lfloor r \rfloor}X_{2,\lfloor r \rfloor}dr \\ &= n^{-1}\int_0^1 X_{1,\lfloor nt \rfloor}X_{2,\lfloor nt \rfloor}dt \\ &= \sigma^2 \int_0^1 \frac{X_{1,\lfloor nt \rfloor}}{\sigma\sqrt{n}} \frac{X_{2,\lfloor nt \rfloor}}{\sigma\sqrt{n}} dt \\ &\to_d \sigma^2 \int_0^1 J_{c_1}(t)J_{c_2}(t)dt. \end{aligned}$$

Thus, we have found the limiting behavior of the off-diagonal elements. The matrix is positive definite by the Cauchy-Schwarz inequality with a strict inequality since $c_1 \neq c_2$.

Next, we have:

$$D_{n,(ii,ii)}^{-1}\sum_{t=1}^{n}X_{t-1}\varepsilon_t=\begin{bmatrix}n^{-1}\sum_{t=1}^{n}X_{1,t-1}\varepsilon_t\\ n^{-1}\sum_{t=1}^{n}X_{2,t-1}\varepsilon_t\end{bmatrix}.$$

We can use the FCLT and the continuous mapping theorem again: since all terms are built from the FCLT result that

$$S_n(t)\to_d W(t)$$

with convergence on $D[0,1]$ (see Phillips (1987b) again), we can conclude that the theorem holds using (7). $\square$

3.2.5 $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(ii)$, $\rho_{2n} \in C(iii)$

Theorem 3.5. *Suppose $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(ii)$, $\rho_{2n} \in C(iii)$. Let $D_{n,(ii,iii)}$ be the following diagonal matrix:*

$$D_{n,(ii,iii)} = \begin{bmatrix} n & 0 \\ 0 & (\rho_{2n}^2 - 1)^{-1}\rho_{2n}^n \end{bmatrix}.$$

Then,

$$\left[D_{n,(ii,iii)}^{-1} \sum_{t=1}^{n} X_{t-1}X_{t-1}' D_{n,(ii,iii)}^{-1}, D_{n,(ii,iii)}^{-1} \sum_{t=1}^{n} X_{t-1}\varepsilon_t \right]' \to_d \left[\begin{bmatrix} \sigma^2 \int_0^1 J_{c_1}(t)^2 dt & 0 \\ 0 & Z_2^2 \end{bmatrix}, \begin{bmatrix} \sigma^2 \int_0^1 J_{c_1}(t) dW(t) \\ Y_2 Z_2 \end{bmatrix} \right]', \tag{23}$$

$$D_{n,(ii,iii)} T_n'(\hat{\phi}_n - \phi_n) \to_d \begin{bmatrix} (\int_0^1 J_{c_1}(t)^2 dt)^{-1} \int_0^1 J_{c_1}(t) dW(t) \\ Z_2^{-1} Y_2 \end{bmatrix}, \tag{24}$$

where Z_2 and Y_2 follows the definition in Theorem 2.6 with innovation sequence $(-\varepsilon_t)_{t\in\mathbb{N}}$. Note that $Z_2^{-1}Y_2 =_d$ Cauchy.

Diagonalization works cleanly in this case: we get a local-to-unity AR(1) process and a mildly explosive AR(1) process, and the components of the limit distribution are simply the respective AR(1) limit distributions. The diagonal entries of $D_{n,(ii,iii)}$ are simply the normalizations from Theorem 2.5 and 2.6.

Proof. We have:

$$D_{n,(ii,iii)}^{-1} \sum_{t=1}^{n} X_{t-1}X_{t-1}' D_{n,(ii,iii)}^{-1} = \begin{bmatrix} n^{-2}\sum_{t=1}^n X_{1,t-1}^2 & * \\ n^{-1}(\rho_{2n}^2 - 1)\rho_{2n}^{-n} \sum_{t=1}^n X_{1,t-1}X_{2,t-1} & (\rho_{2n}^2 - 1)^2 \rho_{2n}^{-2n} \sum_{t=1}^n X_{2,t-1}^2 \end{bmatrix},$$

Theorem 2.5 and 2.6 takes care of the diagonal elements. We focus on the off-diagonal elements. Using (8), we have:

$$\frac{1}{n}(\rho_{2n}^2 - 1)\rho_{2n}^{-n} \sum_{t=1}^{n} X_{1,t-1}X_{2,t-1} = \frac{1}{n}(\rho_{2n}^2 - 1)\rho_{2n}^{-n} \frac{1}{1 - \rho_{1n}\rho_{2n}} \Big(- X_{1,n}X_{2,n} - \rho_{1n} \sum_{t=1}^{n} X_{1,t-1}\varepsilon_t + \rho_{2n} \sum_{t=1}^{n} X_{2,t-1}\varepsilon_t - \sum_{t=1}^{n} \varepsilon_t^2 \Big).$$

We know from previously that $X_{1,n} = O_p(n^{1/2})$ and $X_{2,n} = O_p((\rho_{2n}^2 - 1)^{-1/2}\rho_{2n}^n)$, so $X_{1,n}X_{2,n} = O_p(n^{1/2}(\rho_{2n}^2 - 1)^{-1/2}\rho_{2n}^n)$. Theorem 2.5 implies the second term $O_p(n)$, and Theorem 2.6 implies the third term is $O_p((\rho_{2n}^2 - 1)^{-1}\rho_{2n}^n)$. The last term is $O_p(n)$ by the WLLN. Note that

$1-\rho_{1n}\rho_{2n} \sim 1-\rho_{2n}^2$ since ρ_{2n} converges slower than ρ_{1n} by assumption. Then,

$$\begin{aligned}\frac{1}{n}(\rho_{2n}^2-1)\rho_{2n}^{-n}\sum_{t=1}^{n} X_{1,t-1}X_{2,t-1} &= \frac{1}{n}(\rho_{2n}^2-1)\rho_{2n}^{-n}\frac{1}{1-\rho_{1n}\rho_{2n}}O_p\Big(n^{1/2}(\rho_{2n}^2-1)^{-1/2}\rho_{2n}^{n}\Big)\\ &= O_p\Big(n^{-1/2}(\rho_{2n}^2-1)^{1/2}(1-\rho_{1n}\rho_{2n})^{-1}\Big)\\ &= O_p\Big(n^{-1/2}(\rho_{2n}^2-1)^{-1/2}\Big)\\ &= o_p(1).\end{aligned}$$

Thus, we have:

$$D_{n,(ii,iii)}^{-1}\sum_{t=1}^{n} X_{t-1}X_{t-1}'D_{n,(ii,iii)}^{-1} \to_d \begin{bmatrix} n^{-2}\sum_{t=1}^{n} X_{1,t-1}^2 & 0\\ 0 & Z_{2n}^2\end{bmatrix} + o_p(I_2).$$

Next, we have:

$$D_{n,(ii,iii)}^{-1}\sum_{t=1}^{n} X_{t-1}\varepsilon_t = \begin{bmatrix} n^{-1}\sum_{t=1}^{n} X_{1,t-1}\varepsilon_t\\ (\rho_{2n}^2-1)\rho_{2n}^{-n}\sum_{t=1}^{n} X_{2,t-1}\varepsilon_t\end{bmatrix} = \begin{bmatrix} n^{-1}\sum_{t=1}^{n} X_{1,t-1}\varepsilon_t\\ Y_{2n}Z_{2n}\end{bmatrix}.$$

For joint convergence of all terms in (7), we use the vector martingale FCLT. Let

$$\xi_{n,j} = \begin{bmatrix} \sigma^{-1}n^{-1/2}\\ (\rho_{2n}^2-1)^{1/2}\rho_{2n}^{-j}\\ (\rho_{2n}^2-1)^{1/2}\rho_{2n}^{-(n-j+1)}\end{bmatrix}\varepsilon_j. \tag{25}$$

Then, we have:

$$\sum_{j=1}^{\lfloor nt\rfloor}\mathbb{E}(\xi_{n,j}\xi_{n,j}'|\mathcal{F}_{n,j-1}) = \sigma^2$$

$$\cdot\begin{bmatrix} \sigma^{-2}n^{-1}\lfloor nt\rfloor & * & *\\ \sigma^{-1}n^{-1/2}(\rho_{2n}^2-1)^{1/2}\sum_{j=1}^{\lfloor nt\rfloor}\rho_{2n}^{-j} & (\rho_{2n}^2-1)\sum_{j=1}^{\lfloor nt\rfloor}\rho_{2n}^{-2j} & *\\ \sigma^{-1}n^{-1/2}(\rho_{2n}^2-1)^{1/2}\sum_{j=1}^{\lfloor nt\rfloor}\rho_{2n}^{-(n-j+1)} & (\rho_{2n}^2-1)\sum_{j=1}^{\lfloor nt\rfloor}\rho_{2n}^{-n-1} & (\rho_{2n}^2-1)\sum_{j=1}^{\lfloor nt\rfloor}\rho_{2n}^{-2(n-j+1)}\end{bmatrix}.$$

The diagonal terms converge to $\sigma^{-2}t$, 1 and 1 respectively, where we use geometric series and Lemma C.3 to prove convergence for the (2,2) and (3,3) element as in the proof of Theorem 3.3. For the (2,1) element, we have:

$$\sigma^{-1}n^{-1/2}(\rho_{2n}^2-1)^{1/2}\sum_{j=1}^{\lfloor nt\rfloor}\rho_{2n}^{-j} = O_p((\rho_{2n}^2-1)^{-1/2}n^{-1/2}) = o_p(1).$$

The (3,1) element converges to 0 similarly. The (3,2) element goes to 0 by Lemma C.3. Therefore, the conditional variance converges to $\sigma^2 diag(\sigma^{-2}t,1,1)$. The Lindeberg condition holds by Lemma A.4. Note that the second and third term have no t-dependence, so we can view the second and third components of $\xi_{n,j}$ as constant functions in the space $D[0,1]$. Therefore, they converge to what the vector martingale CLT would imply, while the first component of $\xi_{n,j}$ converges to the vector martingale FCLT-implied Brownian motion. Therefore, we have that

$$\sum_{j=1}^{\lfloor nt \rfloor} \xi_{n,j} \to_d \Big(W(t), Z_2, Y_2 \Big)$$

on $D[0,1]$. Thus, joint convergence of all terms in (7) holds since both $n^{-2} \sum_{t=1}^{n} X_{1,t-1}^2$ and $n^{-1} \sum_{t=1}^{n} X_{1,t-1}\varepsilon_t$ derive from $\sigma^{-1} n^{-1/2} \sum_{j=1}^{\lfloor nt \rfloor} \varepsilon_j$, so we can use the FCLT and the continuous mapping theorem. Thus, we conclude that

$$\begin{aligned} D_{n,(ii,iii)} T_n'(\hat{\phi}_n - \phi) &\to_d \left(\begin{bmatrix} \sigma^2 \int_0^1 J_{c_1}(t)^2 dt & 0 \\ 0 & Z_2^2 \end{bmatrix} \right)^{-1} \begin{bmatrix} \sigma^2 \int_0^1 J_{c_1}(t) dW(t) \\ Y_2 Z_2 \end{bmatrix} \\ &= \begin{bmatrix} (\int_0^1 J_{c_1}(t)^2 dt)^{-1} \int_0^1 J_{c_1}(t) dW(t) \\ Z_2^{-1} Y_2 \end{bmatrix}. \end{aligned}$$

$\square$

3.2.6 $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(iii)$, $\rho_{2n} \in C(iii)$

Let q_n and q be defined as in (9), and change the definition of $Q_{n,i}$ and Q slightly:

$$Q_{n,iii} := \frac{(\rho_{1n}^2 - 1)^{1/2}(\rho_{1n}^2)^{1/2} - 1}{1 - \rho_{1n}\rho_{2n}}, \quad Q_{iii} := \lim_{n\to\infty} Q_{n,iii} \in [0, \infty). \tag{26}$$

The properties between q and Q_{iii} outlined before Theorem 3.1 still hold.

Theorem 3.6. *Suppose $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(iii)$, $\rho_{2n} \in C(iii)$. Let $D_{n,(iii,iii)}$ be the following diagonal matrix:*

$$D_{n,(iii,iii)} = \begin{bmatrix} (\rho_{1n}^2 - 1)^{-1}\rho_{1n}^n & 0 \\ 0 & (\rho_{2n}^2 - 1)^{-1}\rho_{2n}^n \end{bmatrix}.$$

Then, when $q \neq 1$,

$$\left[D_{n,(iii,iii)}^{-1} \sum_{t=1}^n X_{t-1}X_{t-1}' D_{n,(iii,iii)}^{-1}, D_{n,(iii,iii)}^{-1} \sum_{t=1}^n X_{t-1}\varepsilon_t\right]' \to_d \left[\begin{bmatrix} Z_1^2 & Q_{iii}Z_1Z_2 \\ Q_{iii}Z_1Z_2 & Z_2^2 \end{bmatrix}, \begin{bmatrix} Y_1Z_1 \\ Y_2Z_2 \end{bmatrix}\right]', \tag{27}$$

$$D_{n,(iii,iii)}T_n'(\hat{\phi}_n - \phi_n) \to_d \begin{bmatrix} Z_1^2 & Q_{iii}Z_1Z_2 \\ Q_{iii}Z_1Z_2 & Z_2^2 \end{bmatrix}^{-1} \begin{bmatrix} Y_1Z_1 \\ Y_2Z_2 \end{bmatrix}. \tag{28}$$

where Z_1 and Y_1 follows the definition in Theorem 2.6 with innovation sequence $(\varepsilon_t)_{t\in\mathbb{N}}$, and Z_2 and Y_2 have innovation sequence $(-\varepsilon_t)_{t\in\mathbb{N}}$.

Diagonalization yields two mildly explosive AR(1) processes. The diagonal entries of $D_{n,(iii,iii)}$ follow the rates of Theorem 2.6 as expected. Also, when ρ_{1n} and ρ_{2n} converge at different rates, $Q_{iii} = 0$, hence the components of the limiting distribution are just the limit distributions of a mildly explosive AR(1) process.

Proof. We have:

$$D_{n,(iii,iii)}^{-1} \sum_{t=1}^n X_{t-1}X_{t-1}' D_{n,(iii,iii)}^{-1} = \begin{bmatrix} (\rho_{1n}^2 - 1)^2\rho_{1n}^{-2n}\sum_{t=1}^n X_{1,t-1}^2 & * \\ (\rho_{1n}^2 - 1)\rho_{1n}^{-n}(\rho_{2n}^2 - 1)\rho_{2n}^{-n}\sum_{t=1}^n X_{1,t-1}X_{2,t-1} & (\rho_{2n}^2 - 1)^2\rho_{2n}^{-2n}\sum_{t=1}^n X_{2,t-1}^2 \end{bmatrix}.$$

The diagonal elements in the matrix are covered by Theorem 2.6. We focus on the off-diagonal elements. From before, we know that $X_{1,n}X_{2,n} = O_p((\rho_{1n}^2 - 1)^{-1/2}\rho_{1n}^n(\rho_{2n}^2 - 1)^{-1/2}\rho_{2n}^n)$, the second term is $O_p((\rho_{1n}^2 - 1)^{-1}\rho_{1n}^n)$, the third term is $O_p((\rho_{2n}^2 - 1)^{-1}\rho_{2n}^n)$, and the last term is $O_p(n)$. Thus, from (8), we have:

$$\sum_{t=1}^n X_{1,t-1}X_{2,t-1} = -\frac{1}{1 - \rho_{1n}\rho_{2n}}X_{1,n}X_{2,n}(1 + o_p(1)).$$

Then, we have from Phillips and Magdalinos (2007),

$$(\rho_{1n}^2 - 1)\rho_{1n}^{-n}(\rho_{2n}^2 - 1)\rho_{2n}^{-n}\sum_{t=1}^n X_{1,t-1}X_{2,t-1} = Q_{n,iii}Z_{1n}Z_{2n} + o_p(1).$$

Again by Phillips and Magdalinos (2007), we have:

$$D^{-1}_{n,(iii,iii)}\sum_{t=1}^{n}X_{t-1}X'_{t-1}D^{-1}_{n,(iii,iii)} = \begin{bmatrix} Z^2_{1n} & Q_{n,iii}Z_{1n}Z_{2n} \\ Q_{n,iii}Z_{1n}Z_{2n} & Z^2_{2n}\end{bmatrix} + o_p(I_2)$$

Next, we have:

$$D^{-1}_{n,(iii,iii)}\sum_{t=1}^{n}X_{t-1}\varepsilon_t = \begin{bmatrix}(\rho^2_{1n}-1)\rho^{-n}_{1n}\sum_{t=1}^{n}X_{1,t-1}\varepsilon_t \\ (\rho^2_{2n}-1)\rho^{-n}_{2n}\sum_{t=1}^{n}X_{2,t-1}\varepsilon_t\end{bmatrix} = \begin{bmatrix}Y_{1n}Z_{2n}\\ Y_{2n}Z_{2n}\end{bmatrix}.$$

We only need to show joint convergence of Z_{1n}, Y_{1n}, Z_{2n}, and Y_{2n}. Let

$$\xi_{n,t} = \begin{bmatrix} -(\rho^2_{1n}-1)^{1/2}\rho^{-t}_{1n} \\ -(\rho^2_{1n}-1)^{1/2}\rho^{-(n-t+1)}_{1n} \\ (\rho^2_{2n}-1)^{1/2}\rho^{-t}_{2n} \\ (\rho^2_{2n}-1)^{1/2}\rho^{-(n-t+1)}_{2n}\end{bmatrix}\varepsilon_t, \quad \sum_{t=1}^{n}\xi_{n,t} = \begin{bmatrix}Z_{1n}\\ Y_{1n}\\ Z_{2n}\\ Y_{2n}\end{bmatrix}. \tag{29}$$

Then, we have:

$$V_n = \sum_{t=1}^{n}\mathbb{E}(\xi_{n,t}\xi'_{n,t}|\mathcal{F}_{n,t-1}) = \sigma^2$$

$$\cdot\begin{bmatrix}(\rho^2_{1n}-1)\sum_{t=1}^{n}\rho^{-2t}_{1n} & V_{n,12} & V_{n,13} & V_{n,14} \\ V_{n,21} & (\rho^2_{1n}-1)\sum_{t=1}^{n}\rho^{-2(n-t+1)}_{1n} & V_{n,23} & V_{n,24} \\ V_{n,31} & V_{n,32} & (\rho^2_{2n}-1)\sum_{t=1}^{n}\rho^{-2t}_{2n} & V_{n,34} \\ V_{n,41} & V_{n,42} & V_{n,43} & (\rho^2_{2n}-1)\sum_{t=1}^{n}\rho^{-2(n-t+1)}_{2n}\end{bmatrix}$$

where V_n is symmetric and

$$\begin{aligned}
V_{n,21} &= (\rho^2_{1n}-1)\sum_{t=1}^{n}\rho^{-n-1}_{1n}, \\
V_{n,31} &= -(\rho^2_{1n}-1)^{1/2}(\rho^2_{2n}-1)^{1/2}\sum_{t=1}^{n}(\rho_{1n}\rho_{2n})^{-t}, \\
V_{n,41} &= -(\rho^2_{1n}-1)^{1/2}(\rho^2_{2n}-1)^{1/2}\sum_{t=1}^{n}\rho^{-t}_{1n}\rho^{-(n-t+1)}_{2n}, \\
V_{n,23} &= -(\rho^2_{1n}-1)^{1/2}(\rho^2_{2n}-1)^{1/2}\sum_{t=1}^{n}\rho^{-(n-t+1)}_{1n}\rho^{-t}_{2n}, \\
V_{n,24} &= -(\rho^2_{1n}-1)^{1/2}(\rho^2_{2n}-1)^{1/2}\sum_{t=1}^{n}(\rho_{1n}\rho_{2n})^{-(n-t+1)}, \\
V_{n,34} &= (\rho^2_{2n}-1)\sum_{t=1}^{n}\rho^{-n-1}_{2n}.
\end{aligned}$$

The diagonal terms of V_n converge to 1 by using geometric series. $V_{n,21}$ and $V_{4,3}$ converge to 0 by Lemma C.3. $V_{n,31}$ and $V_{n,24}$ converge to Q by using geometric series. Finally, for $V_{n,41}$ (and similarly $V_{n,23}$), we have:

$$\begin{aligned}
V_{n,41} &= -(\rho_{1n}^2 - 1)^{1/2}(\rho_{2n}^2 - 1)^{1/2} \sum_{t=1}^{n} \rho_{1n}^{-t} \rho_{2n}^{-(n-t+1)} \\
&= -(\rho_{1n}^2 - 1)^{1/2}(\rho_{2n}^2 - 1)^{1/2} \rho_{2n}^{-n-1} \sum_{t=1}^{n} \rho_{1n}^{-t} \rho_{2n}^{t} \\
&= -(\rho_{1n}^2 - 1)^{1/2}(\rho_{2n}^2 - 1)^{1/2} \rho_{2n}^{-n-1} \frac{\frac{\rho_{2n}}{\rho_{1n}}(1 - (\frac{\rho_{2n}}{\rho_{1n}})^{n+1})}{1 - \frac{\rho_{2n}}{\rho_{1n}}} \\
&= -(\rho_{1n}^2 - 1)^{1/2}(\rho_{2n}^2 - 1)^{1/2} \rho_{2n}^{-n} \frac{1 - (\frac{\rho_{2n}}{\rho_{1n}})^{n+1}}{\rho_{1n} - \rho_{2n}} \\
&= -(\rho_{1n}^2 - 1)^{1/2}(\rho_{2n}^2 - 1)^{1/2} \frac{\rho_{2n}^{-n} - \rho_{1n}^{-n} \frac{\rho_{2n}}{\rho_{1n}}}{\rho_{1n} - \rho_{2n}} \\
&\to 0.
\end{aligned}$$

We conclude that

$$V_n \to_p \sigma^2 \begin{bmatrix} 1 & 0 & Q_{iii} & 0 \\ 0 & 1 & 0 & Q_{iii} \\ Q_{iii} & 0 & 1 & 0 \\ 0 & Q_{iii} & 0 & 1 \end{bmatrix}$$

Clearly this is positive definite since $Q_{iii} \neq 1$. Thus, with the Lindeberg condition from Lemma A.5, we conclude that

$$\begin{bmatrix} Z_{1n} \\ Y_{1n} \\ Z_{2n} \\ Y_{2n} \end{bmatrix} = \sum_{t=1}^{n} \xi_{n,t} \to_d \begin{bmatrix} Z_1 \\ Y_1 \\ Z_2 \\ Y_2 \end{bmatrix} := \mathcal{N}\Big(0, \sigma^2 \begin{bmatrix} 1 & 0 & Q_{iii} & 0 \\ 0 & 1 & 0 & Q_{iii} \\ Q_{iii} & 0 & 1 & 0 \\ 0 & Q_{iii} & 0 & 1 \end{bmatrix}\Big).$$

Then, by continuous mapping theorem, we can get the limiting distribution over our normalized OLS:

$$D_{n,(iii,iii)} T_n'(\hat{\phi}_n - \phi_n) \to_d \begin{bmatrix} Z_1^2 & Q_{iii} Z_1 Z_2 \\ Q_{iii} Z_1 Z_2 & Z_2^2 \end{bmatrix}^{-1} \begin{bmatrix} Y_1 Z_1 \\ Y_2 Z_2 \end{bmatrix}.$$

□

3.3 Repeated Eigenvalues

In the repeated eigenvalues case, we can no longer diagonalize R_n, but we can employ the Jordan form. Let

$$R_n = \tilde{T}_n J_n \tilde{T}_n^{-1},$$

with

$$\begin{aligned} \tilde{T}_n &= \frac{1}{\rho_n^2} \begin{bmatrix} 0 & \rho_n^2 \\ -1 & \rho_n \end{bmatrix}, \\ \tilde{T}_n^{-1} &= \begin{bmatrix} \rho_n & -\rho_n^2 \\ 1 & 0 \end{bmatrix}, \\ J_n &= \begin{bmatrix} \rho_n & 0 \\ 1 & \rho_n \end{bmatrix}, \end{aligned}$$

where we use (3). The Jordan matrix has the following power properties:

$$\begin{aligned} J_n^k &= \rho_n^{k-1} \begin{bmatrix} \rho_n & 0 \\ k & \rho_n \end{bmatrix}, \\ J_n^{-k} &= \rho_n^{-k-1} \begin{bmatrix} \rho_n & 0 \\ -k & \rho_n \end{bmatrix}. \end{aligned}$$

Transforming the companion form in (2) with $X_t = \tilde{T}_n^{-1} x_t$ and $U_t = \tilde{T}_n^{-1} u_t$, we get

$$X_t = J_n X_{t-1} + U_t, \tag{30}$$

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \rho_n & 0 \\ 1 & \rho_n \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \varepsilon_t \begin{bmatrix} \rho_n \\ 1 \end{bmatrix}.$$

Since we cannot diagonalize, we cannot separate X_t into two AR(1) processes. This generally makes the asymptotics harder to prove. Similar to the distinct eigenvalues case, we must find a suitable normalization diagonal matrix D_n, so our OLS will look like (7):

$$D_n \tilde{T}_n' (\hat{\phi}_n - \phi_n) = \big(D_n^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}' D_n^{-1}\big)^{-1} \big(D_n^{-1} \sum_{t=1}^{n} X_{t-1} \varepsilon_t\big). \tag{31}$$

To help us with the off-diagonal terms of the matrix, the following decomposition is helpful:

$$\begin{aligned} &(I_4 - J_n \otimes J_n) \sum_{t=1}^{n} vec(X_{t-1} X_{t-1}') \\ &= vec\Big(- \sum_{t=1}^{n} U_t U_t' + J_n \sum_{t=1}^{n} X_{t-1} U_t' + \sum_{t=1}^{n} U_t X_{t-1}' J_n' - X_n X_n' \Big). \end{aligned} \tag{32}$$

Furthermore, we must use the *vec* operator and the Kronecker product $\otimes$. We list some key facts in Lemma C.4.

We now solve for asymptotics. We will have three cases depending on the regime of ρ_n.

3.3.1 $\rho_n \in C(i)$

Theorem 3.7. *Suppose $\rho_n \in C(i)$. Let $D_{n,(i)}$ be the following diagonal matrix:*

$$D_{n,(i)} = \begin{bmatrix} (1-\rho_n^2)^{-1/2} & 0 \\ 0 & (1-\rho_n^2)^{-3/2} \end{bmatrix}.$$

Define

$$V_{(i)} := \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

Then,

$$\sqrt{n} D_{n,(i)} \tilde{T}_n'(\hat{\phi}_n - \phi_n) \to_d \mathcal{N}(0, V_{(i)}^{-1}). \tag{33}$$

We can see the effects of having to use a Jordan form rather than a diagonalization in this theorem. We have a strange $(1-\rho_n^2)^{-3/2}$-normalization for the second component because the rotation couldn't separate $X_{1,t}$ from $X_{2,t}$: the second component is an AR(1) process with an AR(1) innovation. Furthermore, the components of the limit distribution no longer resemble each other, as was in Theorem 3.1.

Proof. From (30), we see that the X_{2t} has an innovation that includes an AR(1) term in the $C(i)$ case, so we expect to need to normalize according to $X_n X_n'$. We can write $X_n = \sum_{j=1}^n J_n^{n-j} U_j = \sum_{j=0}^{n-1} J_n^j U_{n-j}$, so:

$$\begin{aligned}
\mathbb{E}(X_n X_n') &= \sum_{j=0}^{n-1} J_n^j \mathbb{E}(U_{n-j} U_{n-j}')(J_n^j)' \\
&= \sigma^2 \sum_{j=0}^{n-1} \rho_n^{2(j-1)} \begin{bmatrix} \rho_n & 0 \\ j & \rho_n \end{bmatrix} \begin{bmatrix} \rho_n \\ 1 \end{bmatrix} \begin{bmatrix} \rho_n & 1 \end{bmatrix} \begin{bmatrix} \rho_n & j \\ 0 & \rho_n \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} \rho_n^2 \sum_{j=0}^{n-1} \rho_n^{2j} & \rho_n \sum_{j=0}^{n-1} \rho_n^{2j}(j+1) \\ \rho_n \sum_{j=0}^{n-1} \rho_n^{2j}(j+1) & \sum_{j=0}^{n-1} \rho_n^{2j}(j+1)^2 \end{bmatrix} \\
&\sim \sigma^2 \begin{bmatrix} (1-\rho_n)^{-1} & \frac{1}{4}(1-\rho_n)^{-2} \\ \frac{1}{4}(1-\rho_n)^{-2} & \frac{1}{4}(1-\rho_n)^{-3} \end{bmatrix}
\end{aligned}$$

by Lemma B.1. Then, if we let

$$D_{n,(i)} = \begin{bmatrix} (1-\rho_n^2)^{-1/2} & 0 \\ 0 & (1-\rho_n^2)^{-3/2} \end{bmatrix},$$

we have

$$\begin{aligned}
D_{n,(i)}^{-1} \mathbb{E}(X_n X_n') D_{n,(i)}^{-1} &\sim \sigma^2 \begin{bmatrix} 1 & \frac{1}{4}(1-\rho_n^2)^2(1-\rho_n)^{-2} \\ \frac{1}{4}(1-\rho_n^2)^2(1-\rho_n)^{-2} & \frac{1}{4}(1-\rho_n^2)^3(1-\rho_n)^{-3} \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} 1 & \frac{1}{4}(1+\rho_n)^2 \\ \frac{1}{4}(1+\rho_n)^2 & \frac{1}{4}(1+\rho_n)^3 \end{bmatrix} \\
&\to \sigma^2 \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.
\end{aligned}$$

Similarly, the above calculations imply that

$$\max_{1\leq t\leq n}\mathbb{E}\|D_{n,(i)}^{-1}X_t\|^2<\infty \tag{34}$$

.

Now, we expect $n^{-1/2}D_{n,(i)}^{-1}$ normalization to be enough to make equation (32) stable. However, we actually require another normalization to ensure convergence. By (32), we know that

$$(I_4-J_n\otimes J_n)\frac{1}{n}\sum_{t=1}^{n}vec(X_{t-1}X_{t-1}')=vec(G_n)$$

$$:=vec\Big(-\frac{1}{n}\sum_{t=1}^{n}U_tU_t'+J_n\frac{1}{n}\sum_{t=1}^{n}X_{t-1}U_t'+\frac{1}{n}\sum_{t=1}^{n}U_tX_{t-1}'J_n'-\frac{1}{n}X_nX_n'\Big).$$

Define

$$\Delta_n=\begin{bmatrix}1 & 0\\ 1 & 1-\rho_n^2\end{bmatrix}. \tag{35}$$

We will normalize by $\Delta_nG_n\Delta_n'$. We have the following facts by Lemma B.2:

$$\|n^{-1/2}\Delta_nX_n\|_{L_2}=o_p(1), \tag{36}$$

$$\Delta_n\frac{1}{n}\sum_{t=1}^{n}U_tU_t'\Delta_n'\to_p\sigma^2\rho_2\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}, \tag{37}$$

$$\left\|vec\Big(\Delta_nJ_n\frac{1}{n}\sum_{t=1}^{n}X_{t-1}U_t'\Delta_n'\Big)\right\|_{L_2}^2=o(1),. \tag{38}$$

Then, we have

$$\begin{aligned}(I_4-J_n\otimes J_n)\frac{1}{n}\sum_{t=1}^{n}vec(X_{t-1}X_{t-1}')&=vec(G_n)\\ &=(\Delta_n^{-1}\otimes\Delta_n^{-1})vec(\Delta_nG_n\Delta_n')\\ &=(\Delta_n\otimes\Delta_n)^{-1}\sigma^2vec\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}+o_p(1),\end{aligned}$$

and since

$$\frac{1}{n}\sum_{t=1}^{n}vec(X_{t-1}X_{t-1}')=(D_{n,(i)}\otimes D_{n,(i)})\sum_{t=1}^{n}vec(D_{n,(i)}^{-1}X_{t-1}X_{t-1}'D_{n,(i)}^{-1}),$$

we have

$$M_nvec\Big(\frac{1}{n}D_{n,(i)}^{-1}\sum_{t=1}^{n}X_{t-1}X_{t-1}'D_{n,(i)}^{-1}\Big)=\sigma^2vec\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}+o_p(1),$$

where

$$
\begin{aligned}
M_n &= (\Delta_n \otimes \Delta_n)(I_4 - J_n \otimes J_n)(D_{n,(i)} \otimes D_{n,(i)}) \\
&= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1-\rho_n^2 & 0 & 0 \\ 1 & 0 & 1-\rho_n^2 & 0 \\ 1 & 1-\rho_n^2 & 1-\rho_n^2 & (1-\rho_n^2)^2 \end{bmatrix} \\
&\cdot \begin{bmatrix} 1-\rho_n^2 & 0 & 0 & 0 \\ -\rho_n & 1-\rho_n^2 & 0 & 0 \\ -\rho_n & 0 & 1-\rho_n^2 & 0 \\ -1 & -\rho_n & -\rho_n & 1-\rho_n^2 \end{bmatrix} \\
&\cdot \begin{bmatrix} (1-\rho_n^2)^{-1} & 0 & 0 & 0 \\ 0 & (1-\rho_n^2)^{-2} & 0 & 0 \\ 0 & 0 & (1-\rho_n^2)^{-2} & 0 \\ 0 & 0 & 0 & (1-\rho_n^2)^{-3} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1-\rho_n & 1 & 0 & 0 \\ 1-\rho_n & 0 & 1 & 0 \\ (\rho_n-2)\rho_n & 1-\rho_n & 1-\rho_n & 1 \end{bmatrix}.
\end{aligned}
$$

Then,

$$
M_n^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \rho_n - 1 & 1 & 0 & 0 \\ \rho_n - 1 & 0 & 1 & 0 \\ \rho_n^2 - 2\rho_n + 2 & \rho_n - 1 & \rho_n - 1 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix},
$$

hence

$$
\begin{aligned}
vec\Big(\frac{1}{n} D_{n,(i)}^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}' D_{n,(i)}^{-1}\Big) &= \sigma^2 \rho^2 M_n^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + o_p(1) \\
&= \sigma^2 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \end{bmatrix} + o_p(1) \\
&= \sigma^2 vec \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} + o_p(1),
\end{aligned}
$$

so

$$
\frac{1}{n} D_{n,(i)}^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}' D_{n,(i)}^{-1} \to_p \sigma^2 V_{(i)}.
$$

Clearly $V_{(i)}$ is positive definite.

Finally, we return to the original OLS in (31). We have:

$$\sqrt{n}D_{n,(i)}\tilde{T}_n'(\hat{\phi}_n - \phi_n) = \Big(\frac{1}{n}D_{n,(i)}^{-1}\sum_{t=1}^{n} X_{t-1}X_{t-1}'D_{n,(i)}^{-1}\Big)^{-1}\Big(n^{-1/2}D_{n,(i)}^{-1}\sum_{t=1}^{n} X_{t-1}\varepsilon_t\Big) \tag{39}$$

The second term on the RHS satisfies a martingale CLT. Define

$$\xi_{n,t} = n^{-1/2}D_{n,(i)}^{-1}X_{t-1}\varepsilon_t \implies \sum_{t=1}^{n}\xi_{n,t} = n^{-1/2}D_{n,(i)}^{-1}\sum_{t=1}^{n} X_{t-1}\varepsilon_t \tag{40}$$

We have

$$\sum_{t=1}^{n}\mathbb{E}(\xi_{n,t}\xi_{n,t}'|\mathcal{F}_{n,t-1}) = \sigma^2\frac{1}{n}D_{n,(i)}^{-1}\sum_{t=1}^{n} X_{t-1}X_{t-1}'D_{n,(i)}^{-1} \to_p \sigma^4 V_{(i)}$$

from our work above. The Lindeberg condition holds by Lemma B.3, so we conclude that

$$n^{-1/2}D_{n,(i)}^{-1}\sum_{t=1}^{n} X_{t-1}\varepsilon_t = \sum_{t=1}^{n}\xi_{n,t} \to_d \mathcal{N}(0, \sigma^4 V_{(i)}),$$

and continuous mapping theorem together with (39) implies

$$\sqrt{n}D_{n,(i)}\tilde{T}_n'(\hat{\phi}_n - \phi_n) \to_d \mathcal{N}(0, V_{(i)}^{-1}).$$

□

3.3.2 $\rho_n \in C(ii)$

Theorem 3.8. *Suppose $\rho_n \in C(ii)$. Let $D_{n,(ii)}$ be the following diagonal matrix:*

$$D_{n,(ii)} = \begin{bmatrix} n & 0 \\ 0 & n^2 \end{bmatrix}.$$

Define

$$H_c(r) := \int_0^r e^{c(r-u)} J_c(u) du.$$

Then,

$$D_{n,(ii)} \tilde{T}_n' (\hat{\phi}_n - \phi_n) \to_d \begin{bmatrix} \int_0^1 J_c(r)^2 dr & \int_0^1 J_c(r) H_c(r) dr \\ \int_0^1 J_c(r) H_c(r) dr & \int_0^1 H_c(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 J_c(r) dW(r) \\ \int_0^1 H_c(r) dW(r) \end{bmatrix}. \qquad (41)$$

Again, the second component requires an n^2-normalization rather than an n-normalization. Furthermore, the limit distribution involves $H_c(r)$ rather than only $J_c(r)$.

Proof. In this case, it is easier to analyze each term of the OLS expansion separately rather than try to work with vectors. From (30), we have:

$$X_{1,t} = \rho_n X_{1,t-1} + \rho_n \varepsilon_t,$$
$$X_{2,t} = \rho_n X_{2,t-1} + X_{1,t-1} + \varepsilon_t.$$

Focus on the matrix term in (31). For $r \in [0,1]$, we have by recursion:

$$\begin{aligned} X_{2,\lfloor nr \rfloor} &= \sum_{j=1}^{\lfloor nr \rfloor} \rho_n^{\lfloor nr \rfloor - j} (X_{1,j-1} + \varepsilon_j) \\ &= \sum_{j=1}^{\lfloor nr \rfloor} \rho_n^{\lfloor nr \rfloor - j} X_{1,j-1} + O_p(n^{1/2}) \\ &= \rho_n^{-1} \int_0^{\lfloor nr \rfloor} \rho_n^{\lfloor nr \rfloor - \lfloor u \rfloor} X_{1,\lfloor u \rfloor} du + O_p(n^{1/2}) \\ &= n \rho_n^{-1} \int_0^{\lfloor nr \rfloor / n} \rho_n^{\lfloor nr \rfloor - \lfloor ns \rfloor} X_{1,\lfloor ns \rfloor} ds + O_p(n^{1/2}). \end{aligned}$$

Then, by Lemma B.4, we have

$$\sigma^{-1} n^{-3/2} X_{2,\lfloor nr \rfloor} = \rho_n^{-1} \int_0^{\lfloor nr \rfloor / n} \rho_n^{\lfloor nr \rfloor - \lfloor ns \rfloor} \frac{X_{1,\lfloor ns \rfloor}}{\sigma \sqrt{n}} ds + O_p(n^{-1}) \to_d H_c(r).$$

where the convergence is on $D[0,1]$. Then, we know

$$\sigma^{-2} n^{-2} \sum_{t=1}^{n} X_{1,t-1}^2 \to_d \int_0^1 J_c(r)^2 dr$$

from Phillips (1987b), and we use the same strategy in the following argument:

$$\begin{aligned}\sigma^{-2}n^{-4}\sum_{t=1}^{n}X_{2,t-1}^2 &= \sigma^{-2}n^{-4}\int_0^n X_{2,\lfloor r\rfloor}^2 dr \\ &= \int_0^1 \left(\frac{X_{2,\lfloor nr\rfloor}}{\sigma n^{3/2}}\right)^2 dr \\ &\to_d \int_0^1 H_c(r)^2 dr.\end{aligned}$$

Importantly, we know joint convergence

$$\left[\frac{X_{1,\lfloor nr\rfloor}}{\sigma\sqrt{n}}, \frac{X_{2,\lfloor nr\rfloor}}{\sigma n^{3/2}}\right]' \to_d \left[J_c(r), H_c(r)\right]'$$

on $D[0,1]\times D[0,1]$ since both limits derive continuously from the basic FCLT result. Thus, we should pick our normalizing diagonal matrix to be $D_{n,(ii)} = diag(n, n^2)$. We have

$$\sigma^{-2}n^{-3}\sum_{t=1}^{n}X_{1,t-1}X_{2,t-1} = \int_0^1 \frac{X_{1,\lfloor nr\rfloor}}{\sigma\sqrt{n}}\frac{X_{2,\lfloor nr\rfloor}}{\sigma n^{3/2}}dr \to_d \int_0^1 J_c(r)H_c(r)dr.$$

So, we have via continuous mapping:

$$D_{n,(ii)}^{-1}\sum_{t=1}^{n}X_{t-1}X_{t-1}'D_{n,(ii)}^{-1} \to_d \sigma^2\begin{bmatrix}\int_0^1 J_c(r)^2 dr & \int_0^1 J_c(r)H_c(r)dr \\ \int_0^1 J_c(r)H_c(r)dr & \int_0^1 H_c(r)^2 dr\end{bmatrix}.$$

This is positive definite by Cauchy-Schwarz (clearly equality does not hold).

Next, we analyze the vector term in (31) and its components. $\sum_{t=1}^n X_{1,t-1}\varepsilon_t$ converges as in Theorem 2.5 since we have innovations $\rho_n\varepsilon_t = (1+O(1/n))\varepsilon_t$. For $\sum_{t=1}^n X_{2,t-1}\varepsilon_t$, let $S_t = \sum_{j=1}^t \varepsilon_t$, and let $\tilde{S}_t = W(T_t)$ be the Skorokhod embedding of S_t into the probability space that the Brownian motion lives in (see Billingsley (1995)). Using recursion, we have:

$$\begin{aligned}X_{1,t} &= \sum_{j=1}^{t}\rho_n^{t-j}\varepsilon_j \\ &= \sum_{j=1}^{t}\rho_n^{t-j}\Delta S_j \\ &= \rho_n^{-1}S_t - \rho_n^t\sum_{j=1}^{t}S_j(\rho_n^{-j-1}-\rho_n^{-j}) \\ &= \rho_n^{-1}\Big(S_t - (1-\rho_n)\sum_{j=1}^{t}S_j\rho_n^{t-j}\Big) \\ &=_d \rho_n^{-1}\Big(\tilde{S}_t - (1-\rho_n)\sum_{j=1}^{t}\tilde{S}_j\rho_n^{t-j}\Big) \\ &=: \tilde{X}_{1,t}.\end{aligned}$$

Then, we have that

$$\begin{aligned}\sigma^{-2}n^{-2}\sum_{t=1}^{n}X_{2,t-1}\varepsilon_t &= \sigma^{-2}n^{-2}\sum_{t=1}^{n}\Big(\sum_{j=1}^{t-1}\rho_n^{t-1-j}X_{1,j-1}\Big)\varepsilon_t + o_p(1)\\ &=_d \sigma^{-2}n^{-2}\sum_{t=1}^{n}\Big(\sum_{j=1}^{t-1}\rho_n^{t-1-j}\tilde{X}_{1,j-1}\Big)\Delta\tilde{S}_t + o_p(1).\end{aligned}$$

We investigate the term in the parentheses:

$$\begin{aligned}\sigma^{-1}n^{-3/2}\sum_{j=1}^{t-1}\rho_n^{t-1-j}\tilde{X}_{1,j-1} &= \rho_n^{-1}\sigma^{-1}n^{-3/2}\sum_{j=1}^{t-1}\rho_n^{t-1-j}\Big(\tilde{S}_{j-1}-(1-\rho_n)\sum_{i=1}^{j-1}\tilde{S}_i\rho_n^{j-1-i}\Big)\\ &= \rho_n^{-1}\sigma^{-1}n^{-3/2}\sum_{j=1}^{t-1}\rho_n^{t-1-j}\Big(\tilde{S}_{j-1}-(1-\rho_n)\int_0^j \tilde{S}_{\lfloor r\rfloor}\rho_n^{j-1-\lfloor r\rfloor}dr\Big)\\ &= \rho_n^{-2}\sigma^{-1}n^{-3/2}\int_0^{t-1}\rho_n^{t-1-\lfloor u\rfloor}\Big(\tilde{S}_{\lfloor u\rfloor}-(1-\rho_n)\int_0^{\lfloor u\rfloor+1}\tilde{S}_{\lfloor r\rfloor}\rho_n^{\lfloor u\rfloor-\lfloor r\rfloor}dr\Big)du\\ &= \rho_n^{-2}\int_0^{\frac{t-1}{n}}\rho_n^{n(\frac{t-1}{n})-\lfloor nu\rfloor}\Big(\frac{\tilde{S}_{\lfloor nu\rfloor}}{\sigma\sqrt{n}}-n(1-\rho_n)\int_0^{\frac{\lfloor nu\rfloor}{n}+\frac{1}{n}}\frac{\tilde{S}_{\lfloor nr\rfloor}}{\sigma\sqrt{n}}\rho_n^{\lfloor nu\rfloor-\lfloor nr\rfloor}dr\Big)du.\end{aligned}$$

Define

$$g_n(x)=\int_0^{x}\rho_n^{n(\frac{t-1}{n})-\lfloor nu\rfloor}\Big(\frac{\tilde{S}_{\lfloor nu\rfloor}}{\sigma\sqrt{n}}-n(1-\rho_n)\int_0^{\frac{\lfloor nu\rfloor}{n}+\frac{1}{n}}\frac{\tilde{S}_{\lfloor nr\rfloor}}{\sigma\sqrt{n}}\rho_n^{\lfloor nu\rfloor-\lfloor nr\rfloor}dr\Big)du,$$

and

$$\begin{aligned}g(x) &= \int_0^x e^{c(x-u)}\Big(W(u)+c\int_0^u W(r)e^{c(u-r)}dr\Big)du\\ &= \int_0^x e^{c(x-u)}J_c(u)du\\ &= H_c(x),\end{aligned}$$

where the second equality holds by integration by parts. FCLT, Lemma B.4, and continuity imply

$$\sup_{1\le t\le n}\Big|g_n(t/n)-g(t/n)\Big|\to_p 0.$$

Then, by Lemma B.5, we have

$$\begin{aligned}\sigma^{-2}n^{-2}\sum_{t=1}^{n}\Big(\sum_{j=1}^{t-1}\rho_n^{t-1-j}\tilde{X}_{1,j-1}\Big)\Delta\tilde{S}_t &= \rho_n^{-2}\sigma^{-1}n^{-1/2}\sum_{t=1}^{n}g_n\Big(\frac{t-1}{n}\Big)\Delta W(T_t)\\ &= \int_0^1 H_c(t)dW(t)+o_p(1).\end{aligned}$$

We conclude that

$$\sigma^{-2} n^{-2} \sum_{t=1}^{n} X_{2,t-1} \varepsilon_t \to_d \int_0^1 H_c(t) dW(t).$$

Using the FLCT and continuity again, we conclude via (31) that

$$\begin{aligned} D_{n,(ii)} \tilde{T}_n' (\hat{\phi}_n - \phi_n) &= \left(D_{n,(ii)}^{-1} \sum_{t=1}^{n} X_{t-1} X_{t-1}' D_{n,(ii)}^{-1} \right)^{-1} \left(D_{n,(ii)}^{-1} \sum_{t=1}^{n} X_{t-1} \varepsilon_t \right) \\ &\to_d \begin{bmatrix} \int_0^1 J_c(r)^2 dr & \int_0^1 J_c(r) H_c(r) dr \\ \int_0^1 J_c(r) H_c(r) dr & \int_0^1 H_c(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 J_c(r) dW(r) \\ \int_0^1 H_c(r) dW(r) \end{bmatrix}. \end{aligned}$$

□

3.3.3 $\rho_n \in C(iii)$

Theorem 3.9. *Suppose $\rho_n \in C(iii)$. Let $D_{n,(iii)}$ be the following diagonal matrix:*

$$D_{n,(iii)} = \begin{bmatrix} (\rho_n^2 - 1)^{-1/2} & 0 \\ 0 & (\rho_n^2 - 1)^{-3/2} \end{bmatrix}.$$

Define

$$\xi := \mathcal{N}(0, V_1),\ V_1 := \sigma^2 \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

and

$$V_\xi := \begin{bmatrix} \xi_1^2 & -\xi_1(\xi_1 - \xi_2) \\ -\xi_1(\xi_1 - \xi_2) & \xi_1^2 + (\xi_1 - \xi_2)^2 \end{bmatrix}.$$

Furthermore, let

$$V_{(iii)} := \sigma^2 \begin{bmatrix} 1 & -1 & 0 & 1 \\ -1 & 2 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}.$$

Then,

$$(\rho_n^2 - 1)^{-1/2} J_n^n D_{n,(iii)} \tilde{T}_n'(\hat{\phi}_n - \phi_n) \to_d \Big(\xi' \otimes (V_\xi)^{-1}\Big) \mathcal{N}(0, V_{(iii)}) \tag{42}$$

.

Proof. Similar to the $C(i)$ case, we expect to need to normalize according to $X_n X_n'$. We have

$$J_n^{-n} X_n = \sum_{j=1}^n J_n^{-j} U_j,$$

and

$$\begin{aligned}
\mathbb{E}\Big[(J_n^{-n}X_n)(J_n^{-n}X_n)'\Big] &= \sum_{j=1}^n J_n^{-j} \mathbb{E}(U_j U_j')(J_n^{-j})' \\
&= \sigma^2 \sum_{j=1}^n \rho_n^{2(-j-1)} \begin{bmatrix} \rho_n & 0 \\ -j & \rho_n \end{bmatrix} \begin{bmatrix} \rho_n \\ 1 \end{bmatrix} \begin{bmatrix} \rho_n & 1 \end{bmatrix} \begin{bmatrix} \rho_n & -j \\ 0 & \rho_n \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} \rho_n^2 \sum_{j=0}^{n-1} \rho_n^{-2j} & -\rho_n^{-1} \sum_{j=1}^{n-1} \rho_n^{-2j} j \\ -\rho_n^{-1} \sum_{j=1}^{n-1} \rho_n^{-2j} j & \rho_n^{-2} \sum_{j=1}^{n} \rho_n^{-2j} j^2 \end{bmatrix} \\
&\sim \sigma^2 \begin{bmatrix} (\rho_n^2 - 1)^{-1} & -\frac{1}{4}(\rho_n - 1)^{-2} \\ -\frac{1}{4}(\rho_n - 1)^{-2} & \frac{1}{4}(\rho_n - 1)^{-3} \end{bmatrix}
\end{aligned}$$

by Lemma B.6. Then, if we let

$$D_{n,(iii)} = \begin{bmatrix} (\rho_n^2 - 1)^{-1/2} & 0 \\ 0 & (\rho_n^2 - 1)^{-3/2} \end{bmatrix},$$

we have

$$D_{n,(iii)}^{-1} \mathbb{E}\Big[(J_n^{-n} X_n)(J_n^{-n} X_n)'\Big] D_{n,(iii)}^{-1} \to V_1 = \sigma^2 \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

Clearly V_1 is positive definite.

Let

$$\xi_n := D_{n,(iii)}^{-1} J_n^{-n} X_n.$$

A martingale CLT with

$$\zeta_{n,j} := D_{n,(iii)}^{-1} J_n^{-j} U_j, \quad \xi_n = \sum_{j=1}^{n} \zeta_{n,j}, \tag{43}$$

implies that

$$\xi_n \to_d \xi := \mathcal{N}(0, V_1)$$

since the conditional variance is

$$\sum_{j=1}^{n} \mathbb{E}\big(\zeta_{n,j}\zeta_{n,j}' | \mathcal{F}_{n,j-1}\big) = D_{n,(iii)}^{-1} \mathbb{E}\Big[(J_n^{-n} X_n)(J_n^{-n} X_n)'\Big] D_{n,(iii)}^{-1} \to V_1.$$

The Lindeberg condition holds by Lemma (B.7).

Next, recall the decomposition in (32). By taking the opposite signs, we have:

$$(J_n \otimes J_n - I_4) \sum_{t=1}^{n} vec(X_{t-1}X_{t-1}')$$

$$= vec\Big(\sum_{t=1}^{n} U_t U_t' - J_n \sum_{t=1}^{n} X_{t-1} U_t' - \sum_{t=1}^{n} U_t X_{t-1}' J_n' + X_n X_n' \Big).$$

Note that

$$\begin{aligned} X_n &= J_n^n D_{n,(iii)} \xi_n \\ \implies vec(X_n X_n') &= \Big(J_n^n D_{n,(iii)} \otimes J_n^n D_{n,(iii)} \Big) vec(\xi_n \xi_n'). \end{aligned}$$

Next, we have

$$\Big\| J_n^{-n} D_{n,(iii)}^{-1} J_n \sum_{t=1}^{n} X_{t-1} U_t' D_{n,(iii)}^{-1} (J_n^{-n})' \Big\|_\infty \le o_p(1), \tag{44}$$

$$\Big\| J_n^{-n} D_{n,(iii)}^{-1} \sum_{t=1}^{n} U_t U_t' D_{n,(iii)}^{-1} (J_n^{-n})' \Big\|_\infty \le o_p(1) \tag{45}$$

by Lemma B.8. Note that we just need the max norm (the largest element) since we will eventually vectorize. Then, the inverse property of the Kronecker product implies

$$\Big(J_n^n D_{n,(iii)} \otimes J_n^n D_{n,(iii)}\Big)^{-1} = D_{n,(iii)}^{-1} J_n^{-n} \otimes D_{n,(iii)}^{-1} J_n^{-n},$$

hence

$$A_n \sum_{t=1}^{n} vec(X_{t-1}X_{t-1}') = vec(\xi_n \xi_n') + o_p(1),$$

where, using the mixed product property of the Kronecker product the commutativity of certain matrices, we have:

$$\begin{aligned} A_n &= \Big(D_{n,(iii)}^{-1} J_n^{-n} \otimes D_{n,(iii)}^{-1} J_n^{-n}\Big)\Big(J_n \otimes J_n - I_4\Big) \\ &= \Big(D_{n,(iii)}^{-1} \otimes D_{n,(iii)}^{-1}\Big)\Big(J_n^{-n} \otimes J_n^{-n}\Big)\Big(J_n \otimes J_n - I_4\Big) \\ &= \Big(D_{n,(iii)}^{-1} \otimes D_{n,(iii)}^{-1}\Big)\Big(J_n \otimes J_n - I_4\Big)\Big(J_n^{-n} \otimes J_n^{-n}\Big). \end{aligned}$$

Then, we have:

$$\Big(D_{n,(iii)}^{-1} \otimes D_{n,(iii)}^{-1}\Big)\Big(J_n \otimes J_n - I_4\Big) vec\Big(J_n^{-n} \sum_{t=1}^{n} X_{t-1}X_{t-1}'(J_n^{-n})'\Big) = vec(\xi_n \xi_n') + o_p(1).$$

To do the proper normalization, we do:

$$B_n vec\Big(D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^{n} X_{t-1}X_{t-1}'(J_n^{-n})' D_{n,(iii)}^{-1}\Big) = vec(\xi_n \xi_n') + o_p(1),$$

where

$$\begin{aligned} B_n &= \Big(D_{n,(iii)}^{-1} \otimes D_{n,(iii)}^{-1}\Big)\Big(J_n \otimes J_n - I_4\Big)\Big(D_{n,(iii)} \otimes D_{n,(iii)}\Big) \\ &= \Big(D_{n,(iii)}^{-1} J_n D_{n,(iii)} \otimes D_{n,(iii)}^{-1} J_n D_{n,(iii)}\Big) - I_4 \\ &= \begin{bmatrix} \rho_n & 0 \\ \rho_n^2 - 1 & \rho_n \end{bmatrix} \otimes \begin{bmatrix} \rho_n & 0 \\ \rho_n^2 - 1 & \rho_n \end{bmatrix} - I_4 \\ &= (\rho_n^2 - 1) \begin{bmatrix} 1 & 0 & 0 & 0 \\ \rho_n & 1 & 0 & 0 \\ \rho_n & 0 & 1 & 0 \\ \rho_n^2 - 1 & \rho_n & \rho_n & 1 \end{bmatrix}. \end{aligned}$$

Let M_n be the matrix. Then

$$M_n \to M := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}.$$

This is invertible, so we have additionally:

$$M_n^{-1} \to M^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -1 & -1 & 1 \end{bmatrix}.$$

Thus, we have:

$$vec\Big((\rho_n^2 - 1)D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^{n} X_{t-1}X_{t-1}'(J_n^{-n})' D_{n,(iii)}^{-1}\Big) = M^{-1} vec(\xi_n \xi_n') + o_p(1).$$

We can expand the RHS and do asymptotics since we know the limiting distribution of ξ_n:

$$M^{-1}vec(\xi_n\xi_n') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \xi_{1,n}^2 \\ \xi_{1,n}\xi_{2,n} \\ \xi_{1,n}\xi_{2,n} \\ \xi_{2,n}^2 \end{bmatrix} = \begin{bmatrix} \xi_{1,n}^2 \\ -\xi_{1,n}(\xi_{1,n} - \xi_{2,n}) \\ -\xi_{1,n}(\xi_{1,n} - \xi_{2,n}) \\ \xi_{1,n}^2 + (\xi_{1,n} - \xi_{2,n})^2 \end{bmatrix}.$$

We have by the continuous mapping theorem on ξ_n:

$$V_\xi^{(n)} := \begin{bmatrix} \xi_{1,n}^2 & -\xi_{1,n}(\xi_{1,n} - \xi_{2,n}) \\ -\xi_{1,n}(\xi_{1,n} - \xi_{2,n}) & \xi_{1,n}^2 + (\xi_{1,n} - \xi_{2,n})^2 \end{bmatrix} \to V_\xi := \begin{bmatrix} \xi_1^2 & -\xi_1(\xi_1 - \xi_2) \\ -\xi_1(\xi_1 - \xi_2) & \xi_1^2 + (\xi_1 - \xi_2)^2 \end{bmatrix}.$$

Then, since $M^{-1}vec(\xi_n\xi_n') = vec(V_\xi) + o_p(1)$, we have

$$(\rho_n^2 - 1)D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^{n} X_{t-1}X_{t-1}'(J_n^{-n})' D_{n,(iii)}^{-1} \to_d V_\xi.$$

Next, we must deal with the vector term in the OLS. We have:

$$\begin{aligned} (\rho_n^2 - 1)^{1/2} D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^{n} X_{t-1}\varepsilon_t &= (\rho_n^2 - 1)^{1/2} D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^{n} \Big(\sum_{j=1}^{t-1} J_n^{t-1-j} U_j\Big)\varepsilon_t \\ &= (\rho_n^2 - 1)^{1/2} D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^{n} \Big(\sum_{j=1}^{n} J_n^{t-1-j} U_j\Big)\varepsilon_t \\ &\quad - (\rho_n^2 - 1)^{1/2} D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^{n} \Big(\sum_{j=t}^{n} J_n^{t-1-j} U_j\Big)\varepsilon_t. \end{aligned}$$

We want the second term to be $o_p(1)$. Let R_n be the second term:

$$R_n = (\rho_n^2 - 1)^{1/2} D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^{n} \Big(\sum_{j=t}^{n} J_n^{t-1-j} U_j\Big)\varepsilon_t. \tag{46}$$

In Lemma B.9, we show that $\|R_n\|_\infty = o_p(1)$, so

$$
\begin{aligned}
(\rho_n^2-1)^{1/2} D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^n X_{t-1}\varepsilon_t &= (\rho_n^2-1)^{1/2} D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^n \Big(\sum_{j=1}^n J_n^{t-1-j} U_j \Big) \varepsilon_t + o_p(1) \\
&= \Big((\rho_n^2-1)^{1/2} D_{n,(iii)}^{-1} \sum_{t=1}^n J_n^{-(n-t+1)} \varepsilon_t D_{n,(iii)} \Big) \Big(D_{n,(iii)}^{-1} \sum_{j=1}^n J_n^{-j} U_j \Big) + o_p(1) \\
&= N_n \xi_n + o_p(1),
\end{aligned} \tag{47}
$$

since

$$
\xi_n = D_{n,(iii)}^{-1} J_n^{-n} X_n = D_{n,(iii)}^{-1} \sum_{j=1}^n J_n^{-j} U_j
$$

and we define

$$
\begin{aligned}
N_n &:= (\rho_n^2-1)^{1/2} D_{n,(iii)}^{-1} \sum_{t=1}^n J_n^{-(n-t+1)} \varepsilon_t D_{n,(iii)} \\
&= (\rho_n^2-1)^{1/2} \sum_{t=1}^n \varepsilon_t \rho_n^{-(n-t+1)} \begin{bmatrix} \rho_n & 0 \\ -(n-t+1)(\rho_n^2-1) & \rho_n \end{bmatrix}.
\end{aligned}
$$

We truncate each term to get independence, just as in the proof of Theorem 2.6. Let k_n be a sequence so that

$$
k_n(\rho_n^2-1) \to \infty, \quad k_n/n \to 0.
$$

Such a sequence exists since ρ_n is in the $C(iii)$ regime. Now, truncate as follows:

$$
\begin{aligned}
\tilde{\xi}_n &= D_{n,(iii)}^{-1} \sum_{j=1}^{k_n} J_n^{-j} U_j \\
\tilde{N}_n &= (\rho_n^2-1)^{1/2} \sum_{t=k_n+1}^n \varepsilon_t \rho_n^{-(n-t+1)} \begin{bmatrix} \rho_n & 0 \\ -(n-t+1)(\rho_n^2-1) & \rho_n \end{bmatrix}.
\end{aligned} \tag{48}
$$

By Lemma B.10, we have $\xi_n = \tilde{\xi}_n + o_p(1)$ and $N_n = \tilde{N}_n + o_p(1)$. Furthermore, $\tilde{\xi}_n$ and $\tilde{N}_n$ are independent since they are sums of disjoint ε_t. Let $\tilde{V}_\xi^{(n)}$ be $V_\xi^{(n)}$ with corresponding truncated components $\tilde{\xi}_{1,n}$ and $\tilde{\xi}_{2,n}$. Thus, we have:

$$
(\rho_n^2-1)^{-1/2} J_n^n D_{n,(iii)} \tilde{T}_n'(\hat{\phi}_n - \phi_n) = (\tilde{V}_\xi^{(n)})^{-1} \tilde{N}_n \tilde{\xi}_n + o_p(1).
$$

It remains to show the asymptotics of the above. We achieve a limit distribution by vectorizing N_n and doing a martingale CLT. Let $v_{n,t} = (\rho_n, -(n-t+1)(\rho_n^2-1), 0, \rho_n)'$, and define

$$
\chi_{n,t} = (\rho_n^2-1)^{1/2} \rho_n^{-(n-t+1)} v_{n,t} \varepsilon_t, \quad vec(N_n) = \sum_{t=1}^n \chi_{n,t}. \tag{49}
$$

We apply the martingale CLT. For the conditional variance, we have

$$
\begin{aligned}
&\sum_{t=1}^{n} \mathbb{E}\big(\chi_{n,t}\chi_{n,t}'|\mathcal{F}_{n,t}\big) \\
&= \sum_{t=1}^{n} \mathbb{E}\Big[(\rho_n^2-1)\rho_n^{-2(n-t+1)} v_{n,t}v_{n,t}'\varepsilon_t^2|\mathcal{F}_{n,t}\Big] \\
&= \sigma^2(\rho_n^2-1)\sum_{t=1}^{n} \rho_n^{-2(n-t+1)} v_{n,t}v_{n,t}' \\
&= \sigma^2 \begin{bmatrix} 1 & -(\rho_n^2-1)^2\sum_{t=1}^n \rho_n^{-2t}t & 0 & 1 \\ -(\rho_n^2-1)^2\sum_{t=1}^n \rho_n^{-2t}t & (\rho_n^2-1)^3\sum_{t=1}^n \rho_n^{-2t}t^2 & 0 & -(\rho_n^2-1)^2\sum_{t=1}^n \rho_n^{-2t}t \\ 0 & 0 & 0 & 0 \\ 1 & -(\rho_n^2-1)^2\sum_{t=1}^n \rho_n^{-2t}t & 0 & 1 \end{bmatrix} \\
&\to V_{(iii)} \\
&:= \sigma^2 \begin{bmatrix} 1 & -1 & 0 & 1 \\ -1 & 2 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}
\end{aligned}
$$

using Lemma B.6. The Lindeberg condition holds by Lemma B.11. Then, we have that

$$
vec(N_n) \to_d \mathcal{N}(0, V_{(iii)}) \implies vec(\tilde{N}_n) \to_d \mathcal{N}(0, V_{(iii)}).
$$

Therefore, we have:

$$
\begin{aligned}
(\rho_n^2-1)^{-1/2} J_n^n D_{n,(iii)} \tilde{T}_n'(\hat{\phi}_n - \phi_n) &= vec\Big((\tilde{V}_\xi^{(n)})^{-1}\tilde{N}_n\tilde{\xi}_n\Big) + o_p(1) \\
&= \Big(\tilde{\xi}_n' \otimes (\tilde{V}_\xi^{(n)})^{-1}\Big) vec(\tilde{N}_n) + o_p(1).
\end{aligned}
$$

By independence of $\tilde{\xi}_n$ and $\tilde{N}_n$, we conclude that we converge in distribution to a mixed Gaussian distribution by (31):

$$
(\rho_n^2-1)^{-1/2} J_n^n D_{n,(iii)} \tilde{T}_n'(\hat{\phi}_n - \phi_n) \to_d \Big(\xi' \otimes (V_\xi)^{-1}\Big)\mathcal{N}(0, V_{(iii)}). \tag{50}
$$

The mixed normal has a positive definite variance by Lemma B.12. □

3.4 Close Distinct Eigenvalues

Recall the definition of q from (9). The problem with $q = 1$ is that we get a singular limit matrix in both the $C(i)$ and $C(i)$ case and the $C(iii)$ and $C(iii)$ case. We must use a different normalization. Luckily, we can modify the Jordan form method. Define

$$T_n(\theta_n) = \frac{1}{\theta_n^2}\begin{bmatrix} 0 & \theta_n^2 \\ -1 & \theta_n \end{bmatrix},$$
$$T_n^{-1}(\theta_n) = \begin{bmatrix} \theta_n & -\theta_n^2 \\ 1 & 0 \end{bmatrix}$$

for some free parameter θ_n. Using (3), we have

$$\begin{aligned} T_n^{-1}(\theta_n)R_nT_n(\theta_n) &= \begin{bmatrix} \theta_n & -\theta_n^2 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \rho_{1n}+\rho_{2n} & -\rho_{1n}\rho_{2n} \\ 1 & 0 \end{bmatrix}\frac{1}{\theta_n^2}\begin{bmatrix} 0 & \theta_n^2 \\ -1 & \theta_n \end{bmatrix} \\ &= \begin{bmatrix} \frac{\rho_{1n}\rho_{2n}}{\theta_n} & \theta_n(\rho_{1n}+\rho_{2n}) - \rho_{1n}\rho_{2n} - \theta_n^2 \\ \frac{\rho_{1n}\rho_{2n}}{\theta_n^2} & \rho_{1n}+\rho_{2n} - \frac{\rho_{1n}\rho_{2n}}{\theta_n} \end{bmatrix}. \end{aligned}$$

Notice that choosing $\theta_n = \rho_{1n}$ or $\theta_n = \rho_{2n}$ makes the above resemble a Jordan form:

$$\begin{aligned} \theta_n = \rho_{1n} \implies T_n^{-1}(\rho_{1n}) &= \frac{1}{\rho_{1n}^2}\begin{bmatrix} 0 & \rho_{1n}^2 \\ -1 & \rho_{1n} \end{bmatrix}, \quad T_n^{-1}(\rho_{1n})R_nT_n(\rho_{1n}) = \begin{bmatrix} \rho_{2n} & 0 \\ \frac{\rho_{2n}}{\rho_{1n}} & \rho_{1n} \end{bmatrix}, \\ \theta_n = \rho_{2n} \implies T_n^{-1}(\rho_{2n}) &= \frac{1}{\rho_{2n}^2}\begin{bmatrix} 0 & \rho_{2n}^2 \\ -1 & \rho_{2n} \end{bmatrix}, \quad T_n^{-1}(\rho_{2n})R_nT_n(\rho_{2n}) = \begin{bmatrix} \rho_{1n} & 0 \\ \frac{\rho_{1n}}{\rho_{2n}} & \rho_{2n} \end{bmatrix}. \end{aligned}$$

The idea is that the above should behave similarly to the Jordan form since ρ_{1n}/ρ_{2n} should be close to 1. We make this notion precise: we have

$$\delta_n := 1 - \frac{1+\rho_{2n}}{1+\rho_{1n}}\frac{1-\rho_{1n}^2}{1-\rho_{2n}^2} = 1 - \frac{1+\rho_{2n}}{1+\rho_{1n}}q \to 0$$

so then

$$\begin{aligned} \rho_{1n} &= \rho_{2n} + (1-\rho_{2n})\delta_n, \\ \implies \frac{\rho_{1n}}{\rho_{2n}} &= 1 + \frac{1}{\rho_{2n}}(1-\rho_{2n})\delta_n. \end{aligned}$$

Now, do the familiar transformation with $X_t = T_n^{-1}(\rho_{2n})x_t$, $K_n = T_n^{-1}(\rho_{2n})R_nT_n(\rho_{2n})$, and $U_t = T_n^{-1}(\rho_{2n})u_t$:

$$X_t = K_nX_{t-1} + U_t, \tag{51}$$

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \rho_{1n} & 0 \\ \frac{\rho_{1n}}{\rho_{2n}} & \rho_{2n} \end{bmatrix}\begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \rho_{2n} \\ 1 \end{bmatrix}\varepsilon_t.$$

3.4.1 $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(i)$, $q = 1$

In this case, since both eigenvalues are near-stationary, K_n behaves like the Jordan decomposition in the repeated $C(i)$ case. The OLS estimator is the same as in Theorem 3.7:

Theorem 3.10. *Suppose $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(i)$, and $q = 1$. Let $D_n(\rho_{2n})$ be the following diagonal matrix:*

$$D_n(\rho_{2n}) = \begin{bmatrix} (1-\rho_{2n}^2)^{-1/2} & 0 \\ 0 & (1-\rho_{2n}^2)^{-3/2} \end{bmatrix}.$$

Define

$$V_{(i)} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

Then,

$$\sqrt{n} D_n(\rho_{2n}) T_n(\rho_{2n})' (\hat{\phi}_n - \phi_n) \to_d \mathcal{N}(0, V_{(i)}^{-1}). \tag{52}$$

Proof. The proof follows similarly to the proof in Theorem 3.7, where we define $D_n(\rho_{2n})$ as above, and $\Delta_n(\rho_{2n})$ as Δ_n with ρ_n replaced by ρ_{2n} in (35). $\square$

3.4.2 $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(iii)$, $\rho_{2n} \in C(iii)$, $q = 1$

In this case, we need to be more careful since K_n^{-n} will not be a good approximation for J_n^{-n}. We have two cases depending on how close ρ_{1n} and ρ_{2n} are to each other. Define

$$\tau := \lim_{n\to\infty} n(\rho_{2n} - \rho_{1n}) \in [-\infty, \infty].$$

We care about τ because we have

$$\begin{aligned} \rho_{2n}^{-n} &= \left(\rho_{1n} + \frac{n(\rho_{2n} - \rho_{1n})}{n}\right)^{-n} \\ &= \rho_{1n}^{-n}\left(1 + \frac{1}{\rho_{1n}}\frac{n(\rho_{2n} - \rho_{1n})}{n}\right)^{-n} \\ &= \rho_{1n}^{-n}(1 + o(1))e^{-\rho_{1n}^{-1} n(\rho_{2n}-\rho_{1n})}, \end{aligned}$$

hence $\rho_{2n}^{-n} \sim \rho_{1n}^{-n}$ when $\tau = 0$.

We can see the two cases appear through the following calculation. Let $r_n := (\rho_{1n} - \rho_{2n})/\rho_{2n} \to 0$, so

$$\begin{aligned} K_n^n &= \left(\begin{bmatrix} \rho_{2n} & 0 \\ 0 & \rho_{2n} \end{bmatrix} + \begin{bmatrix} \rho_{1n} - \rho_{2n} & 0 \\ \frac{\rho_{1n}}{\rho_{2n}} & 0 \end{bmatrix}\right)^n \\ &= \rho_{2n}^n\left(I_2 + \begin{bmatrix} r_n & 0 \\ \frac{\rho_{1n}}{\rho_{2n}^2} & 0 \end{bmatrix}\right)^n \\ &= \rho_{2n}^n \sum_{j=0}^n \binom{n}{j} \begin{bmatrix} r_n & 0 \\ \frac{\rho_{1n}}{\rho_{2n}^2} & 0 \end{bmatrix}^j \\ &= \rho_{2n}^n \begin{bmatrix} 1 + \sum_{j=1}^n \binom{n}{j} r_n^j & 0 \\ \frac{\rho_{1n}}{\rho_{2n}^2} r_n^{-1} \sum_{j=1}^n \binom{n}{j} r_n^j & 1 \end{bmatrix} \\ &= \rho_{2n}^n \begin{bmatrix} (1 + r_n)^n & 0 \\ \frac{\rho_{1n}}{\rho_{2n}^2} r_n^{-1}\big((1 + r_n)^n - 1\big) & 1 \end{bmatrix} \\ &= \begin{bmatrix} \rho_{1n}^n & 0 \\ \frac{\rho_{1n}}{\rho_{2n}} \frac{\rho_{1n}^n - \rho_{2n}^n}{\rho_{1n} - \rho_{2n}} & \rho_{2n}^n \end{bmatrix} \\ \implies K_n^{-n} &= \begin{bmatrix} \rho_{2n}^{-n} & 0 \\ \frac{\rho_{1n}}{\rho_{2n}} \frac{\rho_{1n}^{-n} - \rho_{2n}^{-n}}{\rho_{1n} - \rho_{2n}} & \rho_{1n}^{-n} \end{bmatrix}. \end{aligned}$$

Then,

$$\begin{aligned} \frac{\rho_{1n}^{-n} - \rho_{2n}^{-n}}{\rho_{1n} - \rho_{2n}} &= -n\rho_{1n}^{-n} \frac{1 - \left(1 + \frac{1}{\rho_{1n}}\frac{n(\rho_{2n}-\rho_{1n})}{n}\right)^{-n}}{n(\rho_{2n} - \rho_{1n})} = -n\rho_{1n}^{-n}(1 + o(1))\frac{1 - e^{\tau/\rho_1}}{\tau}, \\ \rho_{2n}^{-n} &= \rho_{1n}^{-n}\left(1 + \frac{1}{\rho_{1n}}\frac{n(\rho_{2n} - \rho_{1n})}{n}\right)^{-n} = \rho_{1n}^{-n}(1 + o(1))e^{-\tau/\rho_1}. \end{aligned}$$

Thus, $\tau = 0$ implies

$$\frac{\rho_{1n}^{-n} - \rho_{2n}^{-n}}{\rho_{1n} - \rho_{2n}} = -n\rho_{1n}^{-n}(1 + o(1)),$$
$$\rho_{2n}^{-n} = \rho_{1n}^{-n}(1 + o(1)),$$

hence

$$K_n^{-n} = (I_2 + o(I_2))J_n^{-n},$$

where J_n is defined at the beginning of section 3.3 with ρ_n replaced by ρ_{2n}. Therefore, when $\tau = 0$, we have the same OLS result. When $\tau \neq 0$, we have normalization by K_n^n rather than by J_n^n.

Theorem 3.11. *Suppose $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(iii)$, $\rho_{2n} \in C(iii)$, and $q = 1$. Define ξ, V_ξ, and $V_{(iii)}$ as in Theorem 3.9. Define*

$$D_n(\rho_{2n}) = \begin{bmatrix} (\rho_{2n}^2 - 1)^{-1/2} & 0 \\ 0 & (\rho_{2n}^2 - 1)^{-3/2} \end{bmatrix}.$$

If $\tau = 0$, then

$$(\rho_n^2 - 1)^{-1/2} J_n^n D_n(\rho_{2n}) T_n(\rho_{2n})'(\hat{\phi}_n - \phi_n) \to_d \Big(\xi' \otimes (V_\xi)^{-1}\Big)\mathcal{N}(0, V_{(iii)}). \tag{53}$$

If $\tau \neq 0$, then

$$(\rho_n^2 - 1)^{-1/2} K_n^n D_n(\rho_{2n}) T_n(\rho_{2n})'(\hat{\phi}_n - \phi_n) \to_d \Big(\xi' \otimes (V_\xi)^{-1}\Big)\mathcal{N}(0, V_{(iii)}). \tag{54}$$

Proof. The proof follows Theorem 3.9. □

4 Discussion

The results in Section 3 establish asymptotic limit theory for drifting AR(2) parameter sequences ϕ_{1n} and ϕ_{2n} with cases depending on the eigenvalues ρ_{1n} and ρ_{2n} of the companion matrix. Section 3.2 covers the case when $\rho_{1n} \neq \rho_{2n}$ and they are far apart, Section 3.3 covers the case when $\rho_{1n} = \rho_{2n}$, and Section 3.4 covers the case when the eigenvalues are distinct but close.

Continuity in the limit distributions of AR(2) does occur between some cases. We have continuity between cases where the components of the OLS limit distribution are independent AR(1) limit distributions:

- $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(i)$, $Q_i = 0$ versus $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(ii)$.
- $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(ii)$ versus $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(iii)$.
- $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(iii)$ versus $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(ii)$, $\rho_{2n} \in C(iii)$.
- $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(ii)$, $\rho_{2n} \in C(iii)$ versus $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(iii)$, $\rho_{2n} \in C(iii)$, $Q_{iii} = 0$.

For example, consider the second case. We see from Theorem 3.2 and Theorem 3.3 that the limit distributions are independent AR(1) limit distributions joined in a vector. The first component is the same for both. For the second component, recall from Section 2.2 that as $c_2 \to \infty$, ρ_{2n} tends from a local-to-unity sequence to a mildly explosive sequence, which has a Cauchy limit distribution. The local-to-unity distribution also approaches a Cauchy distribution as $c_2 \to \infty$, hence we have a continuous bridge between these cases.

We also have continuity between cases where the convergence rates are already similar:

- $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(i)$, $q = 1$ versus $\rho_n \in C(i)$.
- $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(iii)$, $\rho_{2n} \in C(iii)$, $q = 1$ versus $\rho_n \in C(iii)$.

As we saw in Section 3.4, we do a Jordan-esque decomposition in the cases where the eigenvalues are distinct but close together, so we end up with a limit distribution exactly the same as when the eigenvalue sequences are the same.

However, continuity fails across the other cases. Discontinuity in the limit distributions occurs between cases where ρ_{1n} and ρ_{2n} have different convergence speeds versus similar convergence speeds. There are discontinuities between the following cases:

- $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(i)$, $q \neq 1$ versus $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(i)$, $q = 1$.
- $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(i)$, $\rho_{2n} \in C(ii)$ versus $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(ii)$, $\rho_{2n} \in C(ii)$.
- $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(ii)$, $\rho_{2n} \in C(iii)$ versus $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(ii)$, $\rho_{2n} \in C(ii)$.
- $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(iii)$, $\rho_{2n} \in C(iii)$, $q \neq 1$ versus $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(iii)$, $\rho_{2n} \in C(iii)$, $q = 1$.

For example, consider the first case. The eigenvalues are distinct and mildly explosive, but we compare when they have different convergence speeds versus when they converge at a similar rate. The limit distribution when they have different convergence speeds is

$$\mathcal{N}(0, V_{(i,i)}^{-1}), \quad V_{(i,i)} = \begin{bmatrix} 1 & -Q_i \\ -Q_i & 1 \end{bmatrix}$$

by Theorem 3.1 and the limit distribution for the other is

$$\mathcal{N}(0, V_{(i)}^{-1}), \quad V_{(i)} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

by Theorem 3.10. When $q \to 1$, then $Q \to 1$, so $V_{(i,i)}$ becomes singular, whereas $V_{(i)}$ is invertible. Clearly, the limit distributions are discontinuous.

We are unable to comment on the continuity between $\rho_{1n} \neq \rho_{2n}$, $\rho_{1n} \in C(ii)$, $\rho_{2n} \in C(ii)$ versus $\rho_n \in C(ii)$ because of a lack of understanding of $H_c(r)$. Since both cases have the same convergence rate n, we suspect that the distributions vary continuously, but it is unclear how.

Failure of continuity means that we cannot use Andrews et al. (2020) to build a uniform confidence set since we violate Assumptions C1 and C2 in the paper. The procedure for AR(1) relied on creating a confidence set by inverting a test and using the continuously varying critical values from the limiting distributions generated from the drifting parameter sequences (see (2.17) of Andrews et al. (2020)). We tried to extend the successful results of AR(1) to AR(2) but found that it does not work.

Failure to construct a uniform confidence set for AR(2) with Andrews et al. (2020) does not mean that it is impossible. Other methods for constructing uniform confidence sets exist, such as Mikusheva (2007) and Magdalinos and Petrova (2022). The more recent paper uses endogenously constructed instruments to do uniform inference for AR(1). Perhaps the method could be extended to AR(2).

Appendix

A Proof Details of Section 3.2

Lemma A.1. *The martingale difference sequence $\xi_{n,t}$ defined in (14) satisfies the Lindeberg condition.*

Proof. Each component of $\xi_{n,t}$ satisfies a univariate Lindeberg condition by Giraitis and Phillips (2006), hence the Lindeberg condition is satisfied for the vector by Lemma C.2. □

Lemma A.2. *The martingale difference sequence $\xi_{n,j}$ defined in (17) satisfies the Lindeberg condition.*

Proof. The first component of $\xi_{n,j}$ satisfies a univariate Lindeberg condition by Giraitis and Phillips (2006). The second component is easily seen to satisfy it as well:

$$\begin{aligned}\sum_{j=1}^{n} \mathbb{E}\big(|n^{-1/2}\varepsilon_j|^2 \mathbf{1}\{|n^{-1/2}\varepsilon_j| > \delta\}\big) &= \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}\big(\varepsilon_j^2 \mathbf{1}\{\varepsilon_j^2 > n\delta\}\big) \\ &\leq \max_{1\leq j\leq n} \mathbb{E}\big(\varepsilon_j^2 \mathbf{1}\{\varepsilon_j^2 > n\delta\}\big) \\ &\to 0\end{aligned}$$

by uniform integrability of the innovation sequence (since it is iid). Thus, the Lindeberg condition is satisfied for the vector by Lemma C.2. □

Lemma A.3. *The martingale difference sequence $\xi_{n,t}$ defined in (20) satisfies the Lindeberg condition.*

Proof. The first component satisfies a univariate Lindeberg condition by Giraitis and Phillips (2006). The second and third components satisfy univariate Lindeberg conditions by Phillips and Magdalinos (2007). Hence the Lindeberg condition is satisfied for the vector by Lemma C.2. □

Lemma A.4. *The martingale difference sequence $\xi_{n,j}$ defined in (25) satisfies the Lindeberg condition.*

Proof. The first component satisfies a univariate Lindeberg condition by the same proof in Lemma A.2. The second and third components satisfy univariate Lindeberg conditions by Phillips and Magdalinos (2007). Hence the Lindeberg condition is satisfied for the vector by Lemma C.2. □

Lemma A.5. *The martingale difference sequence $\xi_{n,t}$ defined in (29) satisfies the Lindeberg condition.*

Proof. Each component of $\xi_{n,t}$ satisfies a univariate Lindeberg condition by Phillips and Magdalinos (2007), hence the Lindeberg condition is satisfied for the vector by Lemma C.2. □

B Proof Details of Section 3.3

Lemma B.1. *Suppose $\rho_n \in C(i)$. Then, we have*

$$\sum_{j=1}^{n} j^k \rho_n^{2j} \sim (1-\rho_n)^{-1-k} \frac{\Gamma(k+1)}{2^{k+1}}$$

for all $k > 0$.

Proof. Write the sum as an integral with the change of variables $u = (1-\rho_n)x$:

$$\begin{aligned}
\sum_{j=1}^{n} j^k \rho_n^{2j} &= \int_1^{n+1} \lfloor x \rfloor^k \rho_n^{2\lfloor x \rfloor} dx \\
&= (1-\rho_n)^{-1-k} \int_{1-\rho_n}^{(1-\rho_n)(n+1)} \left(\frac{\lfloor (1-\rho_n)^{-1} u \rfloor}{(1-\rho_n)^{-1}} \right)^k \rho_n^{2\lfloor (1-\rho_n)^{-1} u \rfloor} du.
\end{aligned}$$

From the assumptions, we know that $1-\rho_n \to 0$ and $\rho_n \in C(i)$. Noting that $\ln(1+z) = z + O(z^2)$ as $z \to 0$, we have:

$$\begin{aligned}
\rho_n^{2\lfloor (1-\rho_n)^{-1} u \rfloor} &= \big(1-(1-\rho_n)\big)^{2\lfloor (1-\rho_n)^{-1} u \rfloor} \\
&= \exp\Big\{ 2\lfloor (1-\rho_n)^{-1} u \rfloor \ln\big(1-(1-\rho_n)\big) \Big\} \\
&= \exp\Big\{ 2\lfloor (1-\rho_n)^{-1} u \rfloor \big(-(1-\rho_n) + O((1-\rho_n)^2) \big) \Big\} \\
&= \exp\Big\{ 2\lfloor (1-\rho_n)^{-1} u \rfloor \big(-(1-\rho_n) + O((1-\rho_n)^2) \big) \Big\} \\
&= \exp\Big\{ -2\lfloor (1-\rho_n)^{-1} u \rfloor / (1-\rho_n)^{-1} + O(1-\rho_n) \big) \Big\} \\
&\to e^{-2u}.
\end{aligned}$$

Therefore, we have

$$(1-\rho_n)^{1+k} \sum_{j=1}^{n} j^k \rho_n^{2j} \to \int_0^\infty u^k e^{-2u} du = 2^{-(1+k)} \int_0^\infty y^k e^{-y} dy = \frac{\Gamma(k+1)}{2^{1+k}}.$$

□

Lemma B.2. *The equations (36), (37), and (38) are true.*

Proof. For (36), we can use the operator norm and (34). Note that the operator norm equals the largest singular value (see any functional analysis text, such as Rudin (1991)):

$$\begin{aligned}
\|n^{-1/2} \Delta_n X_n\|_{L^2} &= \|n^{-1/2} \Delta_n D_{n,(i)} D_{n,(i)}^{-1} X_n\|_{L^2} \\
&\le \|n^{-1/2} \Delta_n D_{n,(i)}\|_{op} \|D_{n,(i)}^{-1} X_n\|_{L^2} \\
&= \left\| n^{-1/2} \begin{bmatrix} (1-\rho_n^2)^{-1/2} & 0 \\ (1-\rho_n^2)^{-1/2} & (1-\rho_n^2)^{-1/2} \end{bmatrix} \right\|_{op} \|D_{n,(i)}^{-1} X_n\|_{L^2} \\
&= O_p(n^{-1/2}(1-\rho_n^2)^{-1/2}) \\
&= o_p(1).
\end{aligned}$$

We prove (37) next with WLLN:

$$\Delta_n \frac{1}{n}\sum_{t=1}^{n} U_t U_t' \Delta_n' \to_p \sigma^2 \begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix} = \sigma^2 \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}.$$

For (38), note that the previous paragraph implies that each component of $\Delta_n U_t$ is $O(1)$. Then, using the operator norm again, we have:

$$\begin{aligned}
\left\| vec\Big(\Delta_n J_n \frac{1}{n}\sum_{t=1}^{n} X_{t-1}U_t'\Delta_n'\Big)\right\|_{L^2}^2 &= \left\| \frac{1}{n}\sum_{t=1}^{n} (\Delta_n U_t)\otimes(\Delta_n J_n X_{t-1})\right\|_{L^2}^2 \\
&= n^{-2}\sum_{t=1}^{n} \|\Delta_n U_t\|_{L^2}^2 \|\Delta_n J_n X_{t-1}\|_{L^2}^2 \\
&= O(1) n^{-2}\sum_{t=1}^{n} \|\Delta_n J_n X_{t-1}\|_{L^2}^2 \\
&\le O(1) n^{-1} \max_{1\le t\le n} \|\Delta_n J_n X_{t-1}\|_{L^2}^2 .
\end{aligned}$$

We have the bound

$$\begin{aligned}
\|\Delta_n J_n X_{t-1}\|^2 &= \left\| \begin{bmatrix}1 & 0\\ 1 & 1-\rho_n^2\end{bmatrix} \rho_n^{n-1} \begin{bmatrix}\rho_n & 0\\ n & \rho_n\end{bmatrix}\begin{bmatrix}X_{1,t-1}\\ X_{2,t-1}\end{bmatrix}\right\|^2 \\
&= \left\| \begin{bmatrix} X_{1,t-1}\rho_n^n \\ X_{1,t-1}(\rho_n^n + n\rho_n^{n-1}(1-\rho_n^2)) + X_{2,t-1}\rho_n^n(1-\rho_n^2)\end{bmatrix}\right\|^2 \\
&= X_{1,t-1}^2\rho_n^{2n} + X_{1,t-1}^2\big(\rho_n^n + n\rho_n^{n-1}(1-\rho_n^2)\big)^2 + X_{2,t-1}^2\rho_n^{2n}\big(1-\rho_n^2\big)^2 \\
&+ 2X_{1,t-1}X_{2,t-1}\big(\rho_n^n + n\rho_n^{n-1}(1-\rho_n^2)\big)\rho_n^n\big(1-\rho_n^2\big) \\
&\le X_{1,t-1}^2\rho_n^{2n} + X_{1,t-1}^2\big(\rho_n^n + n\rho_n^{n-1}(1-\rho_n^2)\big)^2 + X_{2,t-1}^2\rho_n^{2n}\big(1-\rho_n^2\big)^2 \\
&+ \big(X_{1,t-1}^2 + X_{2,t-1}^2\big)\big(\rho_n^n + n\rho_n^{n-1}(1-\rho_n^2)\big)\rho_n^n\big(1-\rho_n^2\big) \\
&\le X_{1,t-1}^2 + X_{1,t-1}^2 + X_{2,t-1}^2\big(1-\rho_n^2\big)^2 \\
&+ \big(X_{1,t-1}^2 + X_{2,t-1}^2\big)(1-\rho_n^2) + o_p(1) \\
&\le 3X_{1,t-1}^2 + 2\big(1-\rho_n^2\big)^2 X_{2,t-1}^2 + o_p(1)
\end{aligned}$$

by using $\rho_n^n \le 1$ and Lemma C.3. From (34), we know that

$$\max_{1\le t\le n} \mathbb{E}\big((1-\rho_n^2)^{-1}X_{1,t}^2 + (1-\rho_n^2)^{-3}X_{2,t}^2\big) = O(1),$$

so we can say that

$$\|\Delta_n J_n X_{t-1}\|_{\mathbb{R}^2}^2 \le O_p((1-\rho_n^2)^{-1}).$$

Thus,

$$\left\| vec\Big(\Delta_n J_n \frac{1}{n}\sum_{t=1}^{n} X_{t-1}U_t'\Delta_n'\Big)\right\|_{L^2}^2 \le O_p\Big(n^{-1}(1-\rho_n^2)^{-1}\Big) = o_p(1).$$

□

Lemma B.3. *The martingale difference sequence $\xi_{n,t}$ defined in (40) satisfies the Lindeberg condition.*

Proof. The key is (34): by Holder's inequality, we have uniform integrability of $\|D_{n,(i)}^{-1}X_t\|^2$. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of numbers to be chosen later. We have

$$
\begin{aligned}
L_n(\delta) &= \sum_{t=1}^{n} \mathbb{E}(\|\xi_{n,t}\|^2 \mathbf{1}\{\|\xi_{n,t}\| > \delta\}) \\
&= \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\big(\|D_{n,(i)}^{-1}X_{t-1}\|^2\varepsilon_t^2\mathbf{1}\{\|D_{n,(i)}^{-1}X_{t-1}\|^2\varepsilon_t^2 > n\delta^2\}\big) \\
&\le \max_{1\le t\le n} \mathbb{E}\big(\|D_{n,(i)}^{-1}X_{t-1}\|^2\varepsilon_t^2\mathbf{1}\{\|D_{n,(i)}^{-1}X_{t-1}\|^2\varepsilon_t^2 > n\delta^2\}\big) \\
&\le \max_{1\le t\le n} \mathbb{E}\big(\|D_{n,(i)}^{-1}X_{t-1}\|^2\varepsilon_t^2\mathbf{1}\{\|D_{n,(i)}^{-1}X_{t-1}\|^2\varepsilon_t^2 > n\delta^2\}\mathbf{1}\{\|D_{n,(i)}^{-1}X_{t-1}\|^2 \le a_n\}\big) \\
&+ \max_{1\le t\le n} \mathbb{E}\big(\|D_{n,(i)}^{-1}X_{t-1}\|^2\varepsilon_t^2\mathbf{1}\{\|D_{n,(i)}^{-1}X_{t-1}\|^2\varepsilon_t^2 > n\delta^2\}\mathbf{1}\{\|D_{n,(i)}^{-1}X_{t-1}\|^2 > a_n\}\big) \\
&\le a_n \max_{1\le t\le n} \mathbb{E}\big(\varepsilon_t^2\mathbf{1}\{\varepsilon_t^2 > n\delta^2/a_n\}\big) \\
&+ \sigma^2\mathbb{E}\big(\|D_{n,(i)}^{-1}X_{t-1}\|^2\mathbf{1}\{\|D_{n,(i)}^{-1}X_{t-1}\|^2 > a_n\}\big) \\
&\to 0
\end{aligned}
$$

by uniform integrability of ε_t^2 and $\|D_{n,(i)}^{-1}X_t\|^2$, and picking a_n such that $a_n \to \infty$, $n/a_n \to \infty$, and $a_n \max_{1\le t\le n} \mathbb{E}\big(\varepsilon_t^2\mathbf{1}\{\varepsilon_t^2 > n\delta^2/a_n\}\big) \to 0$. □

Lemma B.4. *Suppose $\rho_n \in C(ii)$. Then,*

$$\rho_n^{\lfloor nx\rfloor - \lfloor ny\rfloor} \to e^{c(x-y)}.$$

Proof.

$$\rho_n^{\lfloor nx\rfloor - \lfloor ny\rfloor} = \left[\left(1 + \frac{n(\rho_n - 1)}{n}\right)^n\right]^{\frac{\lfloor nx\rfloor}{n} - \frac{\lfloor ny\rfloor}{n}} \to e^{c(x-y)}.$$

□

Lemma B.5. *Let $(W(t), \mathcal{F}_t)$ be standard Brownian motion on $[0,1]$. Let the Skorokhod embedding of the random walk S_t into this space be given by $S_t =_d W(T_t)$ Suppose we have real functions (g_n) satisfying*

$$\sup_{1\le t\le n}\left|g_n\Big(\frac{t}{n}\Big) - g\Big(\frac{t}{n}\Big)\right| \to_p 0,$$

with $g : [0,1] \to \mathbb{R}$ is continuous. Assuming g_n and g are $\mathcal{F}_t$-measurable, then

$$\frac{1}{\sigma\sqrt{n}}\sum_{t=1}^{n} g_n\Big(\frac{t-1}{n}\Big)\Delta W(T_t) \to_p \int_0^1 g(t)dW(t).$$

Proof. This is a straightforward result from the definition of the stochastic integral. See a textbook such as Karatzas and Shreve (1998). □

Lemma B.6. *Suppose $\rho_n \in C(iii)$. Then, we have*

$$\sum_{j=1}^{n} j^k \rho_n^{-2j} \sim (\rho_n - 1)^{-1-k} \frac{\Gamma(k+1)}{2^{k+1}}$$

for all $k > 0$.

Proof. The proof is the same as for Lemma B.1. □

Lemma B.7. *The martingale difference sequence $\zeta_{n,j}$ defined in (43) satisfies the Lindeberg condition.*

Proof. First, note that

$$\zeta_{n,j} = D_{n,(iii)}^{-1} J_n^{-j} U_j = (\rho_n^2 - 1)^{1/2} \begin{bmatrix} \rho_n^{-j+1} \\ (-j+1)\rho_n^{-j}(\rho_n^2 - 1) \end{bmatrix} \varepsilon_j.$$

We prove Lindeberg conditions for each component. We take advantage of uniform integrability of the innovation sequence since they are iid. For the first component, we have:

$$\begin{aligned}
&\sum_{j=1}^{n} \mathbb{E}\big(|(\rho_n^2 - 1)^{1/2} \rho_n^{-j+1} \varepsilon_j|^2 \mathbf{1}\{|(\rho_n^2 - 1)^{1/2} \rho_n^{-j+1} \varepsilon_j| > \delta\}\big) \\
&= \sum_{j=1}^{n} (\rho_n^2 - 1) \rho_n^{2(-j+1)} \mathbb{E}\big(\varepsilon_j^2 \mathbf{1}\{\varepsilon_j^2 > \delta^2 (\rho_n^2 - 1)^{-1} \rho_n^{-2(-j+1)}\}\big) \\
&\leq \mathbb{E}\big(\varepsilon_j^2 \mathbf{1}\{\varepsilon_j^2 > \delta^2 (\rho_n^2 - 1)^{-1}\}\big)(\rho_n^2 - 1) \sum_{j=1}^{n} \rho_n^{2(-j+1)} \\
&= \mathbb{E}\big(\varepsilon_j^2 \mathbf{1}\{\varepsilon_j^2 > \delta^2 (\rho_n^2 - 1)^{-1}\}\big)(\rho_n^2 - 1) \frac{\rho_n^2 (1 - \rho_n^{2(-n)})}{\rho_n^2 - 1} \\
&\to 0.
\end{aligned}$$

We have the inequality since $\rho_n^{-2(-j+1)} \geq 1$. Next, we have:

$$\begin{aligned}
&\sum_{j=1}^{n} \mathbb{E}\big(|(\rho_n^2 - 1)^{3/2} (-j+1) \rho_n^{-j} \varepsilon_j|^2 \mathbf{1}\{(\rho_n^2 - 1)^{3/2} (-j+1) \rho_n^{-j} \varepsilon_j > \delta\}\big) \\
&= \sum_{j=1}^{n} (\rho_n^2 - 1)^3 (-j+1)^2 \rho_n^{-2j} \mathbb{E}\big(\varepsilon_j^2 \mathbf{1}\{\varepsilon_j^2 > \delta^2 (\rho_n^2 - 1)^{-3} (-j+1)^{-2} \rho_n^{2j}\}\big) \\
&\leq \mathbb{E}\big(\varepsilon_j^2 \mathbf{1}\{\varepsilon_j^2 > \delta^2 (\rho_n^2 - 1)^{-3}\}\big)(\rho_n^2 - 1)^3 \sum_{j=1}^{n} (j-1)^2 \rho_n^{-2j} \\
&= \mathbb{E}\big(\varepsilon_j^2 \mathbf{1}\{\varepsilon_j^2 > \delta^2 (\rho_n^2 - 1)^{-3}\}\big)(\rho_n^2 - 1)^3 O((\rho_n^2 - 1)^{-3}) \\
&\to 0.
\end{aligned}$$

We use Lemma B.6 for the sum.

Using Lemma C.2, we conclude that $\zeta_{n,j}$ satisfies the Lindeberg condition. □

Lemma B.8. *The equations (44) and (45) are true.*

Proof. The key for this lemma is that $\|J_n^{-n}\|_\infty = O(n\rho_n^{-n})$ and $\|D_{n,(iii)}^{-1}\|_\infty = O((\rho_n^2-1)^{3/2})$ by inspection. For (44), note that (47) implies that $\|(\rho_n^2-1)^{1/2}D_{n,(iii)}^{-1}J_n^{-n}\sum_{t=1}^n X_{t-1}\varepsilon_t\|_\infty = O_p(1)$, so we have:

$$\begin{aligned}\Big\|J_n^{-n}D_{n,(iii)}^{-1}J_n\sum_{t=1}^n X_{t-1}U_t'D_{n,(iii)}^{-1}(J_n^{-n})'\Big\|_\infty &\le \Big\|J_n^{-n}D_{n,(iii)}^{-1}J_n\sum_{t=1}^n X_{t-1}U_t'\Big\|_\infty \|D_{n,(iii)}^{-1}\|_\infty\|J_n^{-n}\|_\infty \\ &= O_p\big((\rho_n^2-1)^{-1/2})\big)\|D_{n,(iii)}^{-1}\|_\infty\|J_n^{-n}\|_\infty \\ &= O_p\big(n(\rho_n^2-1)\rho_n^{-n}\big) \\ &= o_p(1)\end{aligned}$$

by Lemma C.3. For (45), WLLN implies $\|\sum_{t=1}^n U_tU_t'\|_\infty = O_p(n)$, so

$$\Big\|J_n^{-n}D_{n,(iii)}^{-1}\sum_{t=1}^n U_tU_t'D_{n,(iii)}^{-1}(J_n^{-n})'\Big\|_\infty \le O_p(n)\|D_{n,(iii)}^{-1}\|_\infty^2\|J_n^{-n}\|_\infty^2 = o_p(1)$$

by Lemma C.3. □

Lemma B.9. *$\|R_n\|$ as defined in (46) is $o_p(1)$.*

Proof. First, we split R_n into R_{1n} and R_{2n}:

$$\begin{aligned}R_n &= (\rho_n^2-1)^{1/2}D_{n,(iii)}^{-1}J_n^{-n}\sum_{t=1}^n\Big(\sum_{j=t}^n J_n^{t-1-j}U_j\Big)\varepsilon_t \\ &= (\rho_n^2-1)^{1/2}D_{n,(iii)}^{-1}J_n^{-n}\sum_{t=1}^n J_n^{-1}U_t\varepsilon_t \\ &+ (\rho_n^2-1)^{1/2}D_{n,(iii)}^{-1}J_n^{-n}\sum_{t=1}^n\Big(\sum_{j=t+1}^n J_n^{t-1-j}U_j\Big)\varepsilon_t \\ &= R_{1n}+R_{2n},\end{aligned}$$

with

$$\begin{aligned}R_{1n} &= (\rho_n^2-1)^{1/2}D_{n,(iii)}^{-1}J_n^{-n}\sum_{t=1}^n J_n^{-1}U_t\varepsilon_t, \\ R_{2n} &= (\rho_n^2-1)^{1/2}D_{n,(iii)}^{-1}J_n^{-n}\sum_{t=1}^n\Big(\sum_{j=t+1}^n J_n^{t-1-j}U_j\Big)\varepsilon_t.\end{aligned}$$

We use that $\|J_n^{-n}\|_\infty = O(n\rho_n^{-n})$ and $\|D_{n,(iii)}^{-1}\|_\infty = O((\rho_n^2-1)^{3/2})$ as in Lemma B.8. For R_{1n}, using WLLN on $\sum_{t=1}^n U_t\varepsilon_t$, we have:

$$\begin{aligned}\|R_{1n}\|_\infty &\le O_p(n)\|(\rho_n^2-1)^{1/2}D_{n,(iii)}^{-1}J_n^{-n}\|_\infty \\ &= O_p(n)(\rho_n^2-1)^{1/2}\|D_{n,(iii)}^{-1}\|_\infty\|J_n^{-n}\|_\infty \\ &= O_p\big(n^2(\rho_n^2-1)^2\rho_n^{-n}\big) \\ &= o_p(1)\end{aligned}$$

by Lemma C.3. For R_{2n}, we can do a change of sums as follows:

$$\begin{aligned}
R_{2n} &= (\rho_n^2-1)^{1/2} D_{n,(iii)}^{-1} J_n^{-n} \sum_{t=1}^{n} \sum_{j=t+1}^{n} J_n^{t-1-j} \begin{bmatrix} \rho_n \\ 1 \end{bmatrix} \varepsilon_j \varepsilon_t \\
&= (\rho_n^2-1)^{1/2} D_{n,(iii)}^{-1} J_n^{-n} \sum_{j=2}^{n} \sum_{t=1}^{j-1} J_n^{t-1-j} \begin{bmatrix} \rho_n \\ 1 \end{bmatrix} \varepsilon_j \varepsilon_t \\
&= (\rho_n^2-1)^{1/2} \sum_{j=2}^{n} \Big(\sum_{t=1}^{j-1} D_{n,(iii)}^{-1} J_n^{-(n+j-t+1)} \begin{bmatrix} \rho_n \\ 1 \end{bmatrix} \varepsilon_t \Big) \varepsilon_j .
\end{aligned}$$

Then, we can see:

$$\begin{aligned}
\mathbb{E}\|R_{2n}\|^2 &= \sigma^2(\rho_n^2-1) \sum_{j=2}^{n} \left\| \sum_{t=1}^{j-1} D_{n,(iii)}^{-1} J_n^{-(n+j-t+1)} \begin{bmatrix} \rho_n \\ 1 \end{bmatrix} \varepsilon_t \right\|^2 \\
&= \sigma^4(\rho_n^2-1) \sum_{j=2}^{n} \sum_{t=1}^{j-1} \left\| D_{n,(iii)}^{-1} J_n^{-(n+j-t+1)} \begin{bmatrix} \rho_n \\ 1 \end{bmatrix} \right\|^2 \\
&= \sigma^4(\rho_n^2-1)^2 \sum_{j=2}^{n} \sum_{t=1}^{j-1} \rho_n^{-2(n+j-t+1)} \left\| \begin{bmatrix} \rho_n^2 \\ -(n+j-t+1)(\rho_n^2-1) \end{bmatrix} \right\|^2 \\
&= (1+O(1))(\rho_n^2-1)^4 \sum_{j=2}^{n} \sum_{t=1}^{j-1} \rho_n^{-2(n+j-t+1)} (n+j-t+1)^2 \\
&\le (1+O(1))(\rho_n^2-1)^4 (4n)^2 \rho_n^{-2n} \sum_{j=2}^{n} \rho_n^{-2(j-1)} \sum_{t=1}^{j-1} \rho_n^{2t} \\
&= (\rho_n^2-1)^4 n^2 \rho_n^{-2n} O(n(\rho_n^2-1)^{-1}) \\
&= O(\rho_n^{-2n} n^3 (\rho_n^2-1)^3) \\
&= o(1),
\end{aligned}$$

where we use Lemma C.3. We conclude $\|R_n\|_\infty = o_p(1)$. □

Lemma B.10. *Let $\tilde{\xi}_n$ and $\tilde{N}_n$ be defined as in (48). Then*

$$\begin{aligned}
\xi_n &= \tilde{\xi}_n + o_p(1), \\
N_n &= \tilde{N}_n + o_p(1).
\end{aligned}$$

Proof. The lemma is obvious once observing that $\tilde{\xi}_n$ only weights terms that carry non-trivial magnitude in the sum that describes ξ_n: $\|J_n^{-j}\|_\infty$ is greater for smaller j, and near 0 for larger j. The same observation holds for $\tilde{N}_n$. □

Lemma B.11. *The martingale difference sequence $\chi_{n,t}$ defined in (49) satisfies the Lindeberg condition.*

Proof. The proof of this lemma is very similar to Lemma B.7. We solve component-wise Lindeberg conditions. The third component is trivial. For the first and last components, by uniform integrability of the innovation sequence (it is iid), and since we have:

$$\begin{aligned}
&\sum_{t=1}^{n}\mathbb{E}\big(|(\rho_n^2-1)^{1/2}\rho_n^{-(n-t)}\varepsilon_t|^2\mathbf{1}\{|(\rho_n^2-1)^{1/2}\rho_n^{-(n-t)}\varepsilon_t|>\delta\}\big)\\
&=\sum_{t=1}^{n}(\rho_n^2-1)\rho_n^{-2(n-t)}\mathbb{E}\big(\varepsilon_t^2\mathbf{1}\{\varepsilon_t^2>\delta^2(\rho_n^2-1)^{-1}\rho_n^{2(n-t)}\}\big)\\
&\le \mathbb{E}(\varepsilon_1^2\mathbf{1}\{\varepsilon_1^2>\delta^2(\rho_n^2-1)^{-1}\})(\rho_n^2-1)\sum_{t=1}^{n}\rho_n^{-2(n-t)}\\
&=\mathbb{E}(\varepsilon_1^2\mathbf{1}\{\varepsilon_1^2>\delta^2(\rho_n^2-1)^{-1}\})(\rho_n^2-1)\frac{\rho_n^2(1-\rho_n^{-2(n-1)})}{\rho_n^2-1}\\
&\to 0.
\end{aligned}$$

The inequality comes from $\rho_n^{2(n-t)}\ge 1$, hence $\mathbf{1}\{\varepsilon_t^2>\delta^2(\rho_n^2-1)^{-1}\rho_n^{2(n-t)}\}\le\mathbf{1}\{\varepsilon_1>\delta(\rho_n^2-1)^{-1/2}\}$. For the second component, using uniform integrability again, we have:

$$\begin{aligned}
&\sum_{t=1}^{n}\mathbb{E}\big(|(\rho_n^2-1)^{3/2}\rho_n^{-(n-t+1)}(n-t+1)\varepsilon_t|^2\mathbf{1}\{|(\rho_n^2-1)^{3/2}\rho_n^{-(n-t+1)}(n-t+1)\varepsilon_t|>\delta\}\big)\\
&=\sum_{t=1}^{n}(\rho_n^2-1)^3\rho_n^{-2(n-t+1)}(n-t+1)^2\mathbb{E}\big(\varepsilon_t^2\mathbf{1}\{\varepsilon_t^2>\delta^2(\rho_n^2-1)^{-3}\rho_n^{2(n-t+1)}(n-t+1)^{-2}\}\big)\\
&\le \mathbb{E}\big(\varepsilon_1^2\mathbf{1}\{\varepsilon_1^2>\delta^2(\rho_n^2-1)^{-3}\}\big)(\rho_n^2-1)^3\sum_{t=1}^{n}t^2\rho_n^{-2t}\\
&=\mathbb{E}\big(\varepsilon_1^2\mathbf{1}\{\varepsilon_1^2>\delta^2(\rho_n^2-1)^{-3}\}\big)(\rho_n^2-1)^3O((\rho_n^2-1)^{-3})\\
&\to 0.
\end{aligned}$$

We use Lemma B.6 for the sum.

Using Lemma C.2, we conclude that $\chi_{n,t}$ satisfies the Lindeberg condition. ☐

Lemma B.12. *The variance matrix in the distribution of (50) is positive definite.*

Proof. First, we have

$$\begin{aligned}
\xi'\otimes(V_\xi)^{-1}&=\begin{bmatrix}\xi_1 & \xi_2\end{bmatrix}\otimes\xi_1^{-4}\begin{bmatrix}\xi_1^2+(\xi_1-\xi_2)^2 & \xi_1(\xi_1-\xi_2)\\ \xi_1(\xi_1-\xi_2) & \xi_1^2\end{bmatrix}\\
&=\begin{bmatrix}\xi_1^3+\xi_1(\xi_1-\xi_2)^2 & \xi_1^2(\xi_1-\xi_2) & \xi_1^2\xi_2+\xi_2(\xi_1-\xi_2)^2 & \xi_1\xi_2(\xi_1-\xi_2)\\ \xi_1^2(\xi_1-\xi_2) & \xi_1^3 & \xi_1\xi_2(\xi_1-\xi_2) & \xi_1^2\xi_2\end{bmatrix}.
\end{aligned}$$

Then, we have

$$\Big(\xi'\otimes(V_\xi)^{-1}\Big)V_{(iii)}\Big(\xi'\otimes(V_\xi)^{-1}\Big)'=\sigma^2\xi_1^{-4}\begin{bmatrix}\xi_1^2+(\xi_1-\xi_2)^2 & \xi_1(\xi_1-\xi_2)\\ \xi_1(\xi_1-\xi_2) & \xi_1^2\end{bmatrix}.$$

The determinant is then $\sigma^2>0$. ☐

C Additional Lemmas

We first present a technical lemma from Cytrynbaum (2024).

Lemma C.1. *For any positive constants* $(a_k)_{k=1}^d$ *and* $\delta > 0$*, we have:*

$$\sum_{k=1}^{d} a_k \mathbf{1}\Big\{\sum_{k=1}^{d} a_k > \delta\Big\} \leq d \sum_{k=1}^{d} a_k \mathbf{1}\Big\{a_k > \frac{\delta}{d}\Big\}$$

Proof. If $\sum_{k=1}^d a_k \leq \delta$, then the left-hand side is 0, so the inequality holds trivially. Else, $\exists k'$ such that $a_{k'} > \delta/d$ (if not, the sum would be less than δ). Let k' correspond to the maximum such $a_{k'}$. Then

$$\sum_{k=1}^{d} a_k \mathbf{1}\Big\{\sum_{k=1}^{d} a_k > \delta\Big\} \leq d a_{k'} \leq d \sum_{k=1}^{d} a_k \mathbf{1}\Big\{a_k > \frac{\delta}{d}\Big\}.$$

□

To help us with the Lindeberg conditions, we introduce a lemma that says if the components of a martingale difference sequence satisfy univariate Lindeberg conditions, then the vector satisfies the vector Lindeberg condition.

Lemma C.2. *Suppose* $\xi_{n,j} = [\xi_{n,j}^{(1)}, \ldots, \xi_{n,j}^{(d)}]$ *is a* $\mathcal{F}_{n,j}$*-martingale difference array with* $1 \leq d < \infty$*. For* $\delta > 0$*, define*

$$L_n^{(i)}(\delta) := \sum_{j=1}^{k_n} \mathbb{E}\big(|\xi_{n,j}^{(i)}|^2 \mathbf{1}\{|\xi_{n,j}^{(i)}| > \delta\}\big)$$

for $1 \leq i \leq d$ *and*

$$L_n(\delta) := \sum_{j=1}^{k_n} \mathbb{E}\big(\|\xi_{n,j}\|^2 \mathbf{1}\{\|\xi_{n,j}\| > \delta\}\big).$$

If $L_n^{(i)}(\delta) \to 0$ *for all* δ *and* i*, then* $L_n(\delta) \to 0$ *for all* δ*.*

Proof. We use Lemma C.1:

$$
\begin{aligned}
L_n(\delta) &= \sum_{j=1}^{k_n} \mathbb{E}\big(\|\xi_{n,j}\|^2 \mathbf{1}\{\|\xi_{n,j}\| > \delta\}\big) \\
&= \sum_{j=1}^{k_n} \mathbb{E}\big(\|\xi_{n,j}\|^2 \mathbf{1}\{\|\xi_{n,j}\|^2 > \delta^2\}\big) \\
&= \sum_{j=1}^{k_n} \mathbb{E}\Big(\sum_{i=1}^{d} |\xi_{n,j}^{(i)}|^2 \mathbf{1}\Big\{\sum_{i=1}^{d} |\xi_{n,j}^{(i)}|^2 > \delta^2\Big\}\Big) \\
&\leq \sum_{j=1}^{k_n} \mathbb{E}\Big(d \sum_{i=1}^{d} |\xi_{n,j}^{(i)}|^2 \mathbf{1}\Big\{|\xi_{n,j}^{(i)}|^2 > \frac{\delta^2}{d}\Big\}\Big) \\
&= d \sum_{i=1}^{d} \sum_{j=1}^{k_n} \mathbb{E}\Big(|\xi_{n,j}^{(i)}|^2 \mathbf{1}\Big\{|\xi_{n,j}^{(i)}| > \frac{\delta}{\sqrt{d}}\Big\}\Big) \\
&= d \sum_{j=1}^{k_n} L_n^{(i)}(d^{-1/2}\delta) \\
&\to 0.
\end{aligned}
$$

□

The following is a lemma from Phillips and Magdalinos (2007).

Lemma C.3. *Parameterize ρ_n as follows: $\rho_n = 1 + c/k_n$ for some sequence k_n.*

(a) When ρ_n is in the $C(i)$ regime, $\rho_n^n = o(k_n/n)$.

(b) When ρ_n is in the $C(iii)$ regime, $\rho_n^{-n} = o(k_n/n)$.

Note that $k_n^{-1} \sim (1-\rho_n) \sim (1-\rho_n^2)$ in the $C(i)$ case, and $k_n^{-1} \sim (\rho_n - 1) \sim (\rho_n^2 - 1)$ in the $C(iii)$ case. Therefore, $\big(n(1-\rho_n^2)\big)^k \rho_n^n \to 0$ in the $C(i)$ case and $\big(n(\rho_n^2-1)\big)^k \rho_n^{-n} \to 0$ in the $C(iii)$ case for any $k \geq 0$.

The following are facts about *vec* and $\otimes$.

Lemma C.4.

- *If $x, y \in \mathbb{R}_n$, then $vec(xy') = y \otimes x$.*
- *If A, B, and C are matrices such that ABC is well-defined, then $vec(ABC) = (C' \otimes A)vec(B)$.*
- *If A, A, B, C, and D are matrices such that AC and BD are well-defined, then $(A \otimes B)(C \otimes D) = AC \otimes BD$.*

References

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *The Annals of Mathematical Statistics*, 30(3):676–687.

Andrews, D. W. K. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica*, 61(1):139–165.

Andrews, D. W. K., Cheng, X., and Guggenberger, P. (2020). Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics*, 218(2):496–531.

Billingsley, P. (1995). *Probability and Measure*. John Wiley, New York, 3rd edition.

Billingsley, P. (1999). *Convergence of Probability Measures*. John Wiley, New York, 2nd edition.

Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary ar(1) processes. *The Annals of Statistics*, 15(3):1050–1063.

Cytrynbaum, M. (2024). Optimal stratification of survey experiments. *Working Paper*.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431.

Donsker, M. D. (1951). An invariance principle for certain probability limit theorems. *Memoris of the American Mathematical Society*, 6:1–12.

Giraitis, L. and Phillips, P. C. B. (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis*, 27(1):51–60.

Haldrup, N. (2002). An econometric analysis of i(2) variables. *Journal of Economic Surveys*, 12(5):595–650.

Haldrup, N. and Lildholdt, P. (2002). On robustness of unit root tests in the presence of double unit roots. *Journal of Time Series Analysis*, 23(2):127–250.

Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, 12(1):141–153.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin, 2nd edition.

Karatzas, I. and Shreve, S. E. (1998). *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York.

Magdalinos, T. and Petrova, K. (2022). Uniform and distribution-free inference with general autoregressive processes. *Working Paper*.

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, 75(5):1411–1452.

Phillips, P. C. B. (1977). Approximations to some finite sample distributions associated with a first-order stochastic difference equation. *Econometrica*, 45(2):463–485.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, 55(2):277–301.

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

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

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag, New York.

Rudin, W. (1991). *Functional Analysis*. McGraw-Hill, New York, 2nd edition.

Stock, J. H. (1991). Confidence intervals for the largest autoregressive root in u.s. macroeconomic time series. *Journal of Monetary Economics*, 28(3):435–459.