

Stauskas, Ovidijus

Working Paper

On the Limit Theory of Mixed to Unity VARs: Panel Setting With Weakly Dependent Errors

Working Paper, No. 2019:2

Provided in Cooperation with:

Department of Economics, School of Economics and Management, Lund University

Suggested Citation: Stauskas, Ovidijus (2019) : On the Limit Theory of Mixed to Unity VARs: Panel Setting With Weakly Dependent Errors, Working Paper, No. 2019:2, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/260272>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Working Paper 2019:2

Department of Economics
School of Economics and Management

On the Limit Theory of Mixed to Unity VARs: Panel Setting With Weakly Dependent Errors

Ovidijus Stauskas

January 2019



LUND
UNIVERSITY

On the Limit Theory of Mixed to Unity VARs: Panel Setting With Weakly Dependent Errors

Ovidijus Stauskas *

January 2019

Abstract

In this paper we re-visit a recent theoretical idea introduced by [Phillips and Lee \(2015\)](#). They examine an empirically relevant situation when multiple time series exhibit different degrees of non-stationarity. By bridging the asymptotic theory of the local to unity and mildly explosive processes, they construct a Wald test for the commonality of the long-run behavior of two series. Therefore, a vector autoregressive (VAR) setup is natural. However, inference is complicated by the fact that the statistic is degenerate under the null and divergent under the alternative. This is true if the parameters of the data generating process are known and a re-normalizing function can be constructed. If the parameters are unknown, as is usually the case in practice, the test statistic may be divergent even under the null. We solve this problem by converting the original setting of vector time series into a panel setting with N individual vector series. We consider asymptotics with fixed N as $T \rightarrow \infty$ and extend the results to sequential asymptotics when T passes to infinity before N . We show that the Wald test statistic converges to Chi-squared distribution which is free of nuisance parameters under the null hypothesis of common local to unity behavior.

Keywords Local to unity, mildly explosive, panel, weak dependence, Wald test

JEL classification codes: C12, C32, C33

*Lund University, Department of Economics, Lund, Sweden. E-mail address: ovidijus.stauskas@nek.lu.se, Office:275, EC1. I am grateful to Joakim Westerlund (Lund University), Peter Jochumzen (Lund University), Simon Reese (USC Dornsife INET), Stephan Smeekes (Maastricht University) and Henrik Bengtsson (Lund University).

1 Introduction

While local to unity stochastic processes have been extensively used as alternatives in unit root testing (see e.g. [Phillips \(1987\)](#), [Chan and Wei \(1987\)](#) or [Elliott and Jansson \(2003\)](#)), explosive processes have been useful in modelling bubbles in financial markets (see e.g. [Phillips, Wu, and Yu \(2011\)](#) or [Phillips, Shi, and Yu \(2015\)](#)). The type of explosive processes considered here is called mildly explosive and it was popularized by the works of [Phillips and Magdalinos \(2007\)](#) and [Magdalinos and Phillips \(2009a\)](#). Both local to unity and mildly explosive processes are in the vicinity of $O(T)$ and $O(k_T)$, respectively, where k_T is a function of the sample size. However, they are on the different sides of unit root.

[Phillips and Lee \(2015\)](#) considered the very challenging route of bridging the asymptotics of local to unity and mildly explosive cases when they occur in the same estimation procedure. It is an interesting technical question and a realistic scenario when a researcher is faced with potentially different degrees of non-stationarity. This naturally leads to a question as to whether the persistence of the series is of the same nature. Contrary to previous studies (e.g. [Phillips et al. \(2015\)](#)) where exact unit root behavior is usually taken as the null hypothesis, they allow local to unity under the null. This gives more flexibility as local to unity processes are very general and include pure unit roots as a special case.

Since it is very unlikely that both series are unrelated, it is natural to analyze them in a vector autoregressive (VAR) setting. In this setup, [Phillips and Lee \(2015\)](#) consider a Wald test for the equivalence of the largest autoregressive root of the two series. They show that if both series are local to unity, then the test statistic degenerates as the sample size T increases, whereas it diverges if one of the series is mildly explosive. The asymptotic size of such test is therefore 0, which means that any set of critical values can be used. Hence, while the test has discriminatory power between two persistence regimes, the procedure is not really practical. The problem lies in the fact that while Wald statistic is $O_p(1)$ under the null, it depends on the non-estimable localizing parameter c . The statistic can be made to go to zero by re-scaling it by L_T , which is a slowly diverging function of k_T which determines the vicinity. The problem is that L_T is only determined up to a certain rate of expansion with T since k_T is unknown. This means that L_T can basically take any value.

This degeneracy problem in a time series context was addressed in [Phillips and Lee \(2016\)](#) using the IVX estimator in a more general k -variate regression case. [Phillips and Magdalinos \(2009b\)](#) showed that IVX Wald test statistics have a pivotal Chi-squared distribution that is free of nuisance parameters, while [Phillips and Lee \(2016\)](#) extended this result to certain prototypical mixed VARs, covering the problem in [Phillips and Lee \(2015\)](#). However, they point out that the IVX instrument selection procedure remains sub-optimal. The suggested a way out is to impose more structure by assuming which

processes are mildly explosive or near integrated, which is equivalent to imposing a degree of persistence under the null as in the original paper (Phillips & Lee, 2015).

In this paper we consider the same testing problem as in Phillips and Lee (2015). The novelty is that we put it into a panel data context. The main idea is that the increased number of observations will simplify the testing problem. Our results support this. The analysis is based on sequential limit theory (see Phillips and Moon (1999b) and Moon and Phillips (2000)). Instances of it can be found in panel unit root testing (e.g. Im, Pesaran, and Shin (2003)) or testing for no cointegration (e.g. Pedroni (2004)). An additional problem arises, because unlike Phillips and Lee (2015), we do not want to assume independence but instead allow the errors to be weakly dependent over time, as in the presence of e.g. measurement errors (Alvarez & Arellano, 2004). As a solution to this problem, we use the Fully Modified Least Squares (FMLS) estimator. It is shown to be normal under the sequential asymptotics. Also, we allow cross-section heteroskedasticity under some regularity conditions. By using the panel setting, we make an exploratory step to examine explosiveness in panels and contribute to somewhat scarce theoretical literature on non-stationary panel VARs (see Binder et al. (2005) and references therein).

The remainder of this paper is organized as follows. Section 2 describes the model and provides assumptions we impose on the errors. Section 3 explores the asymptotic distribution of the estimator of the VAR coefficient matrix \mathbf{R}_T for fixed N and large N . Section 4 constructs and examines the Wald statistic for testing the common degree of persistence in the panel VAR setting. The Technical Appendix contains the proofs of all the necessary lemmas and theorems.

We use the following notation: \xrightarrow{P} , \xrightarrow{D} and $\xrightarrow{\mathcal{L}^p}$ represent convergence in probability, distribution and \mathcal{L}^p norm, respectively. Weak convergence in measure is represented by \Rightarrow and $(T, N)_{seq} \rightarrow$ stands for sequential limits, while distributional equality is given by $\stackrel{D}{=}$. The smallest sigma-algebra generated by a random variable under consideration is $\sigma(\cdot)$.

2 The Model

We consider the bivariate VAR(1) model as in Phillips and Lee (2015), but we add the cross-section dimension consisting of N individuals. We also impose the common coefficients among the N systems under consideration¹.

$$X_{it} = \mathbf{R}_T X_{it-1} + u_{it}, \quad t = 1, \dots, T, \quad i = 1, \dots, N \quad (2.1)$$

¹Hjalmarsson (2005) considered panels with $\rho_{iT} = 1 + \frac{c+\eta_i}{T}$, where $\eta_i \sim \mathcal{N}(0, \sigma_i^2)$ and suggested the median-based estimator of c . However, this approach is aimed at summarizing the persistence patterns in a data set rather than being employed in regression analysis. See Theorem 3.2 and Theorem 4.1 below.

$$\mathbf{R}_T = \begin{bmatrix} \rho_T & 0 \\ 0 & \theta_T \end{bmatrix} \quad (2.2)$$

Here, $X_{it} = [X_{1it} \ X_{2it}]^\top$ and $u_{it} = [u_{1it} \ u_{2it}]^\top$. We keep the coefficient matrix \mathbf{R}_T diagonal for clarity of the asymptotic theory, according to the common practice (e.g. [Phillips and Lee \(2016\)](#), [Phillips and Magdalinos \(2009b\)](#)). The effects of further lags on the large sample properties will be examined in future work. The coefficients represent local to unity and mildly explosive cases in AR(1) process, respectively:

$$\rho_T = 1 + \frac{c}{T}, \quad \theta_T = 1 + \frac{b}{k_T} \quad (2.3)$$

where $c < 0$ and $b > 0$. The strict inequalities are used to keep the processes in the lower and the upper vicinity of the unit root, respectively (see Proposition 1 (a, b) in the Technical Appendix). The mildly explosive coefficient θ_T depends on k_T such that $k_T = o(T)$. Hence, $\frac{b}{k_T}$ goes to zero at a slower rate than $\frac{c}{T}$, which is a necessary condition to discriminate between the locality to unity induced by ρ_T and the mild explosiveness induced by θ_T . In fact, k_T can be any function satisfying $k_T = o(T)$. One possibility is to set $k_T = T^\alpha$ with $\alpha \in (0, 1)$. However, this is by no means a restriction.

We focus on the square matrix \mathbf{R}_T and omit the intercept vector following the literature that explores estimators with non-standard asymptotic behavior either on the lower or upper vicinity of the unit root. For example, see [Phillips et al. \(2010\)](#), [Giraitis and Phillips \(2012\)](#) or [Arvanitis and Magdalinos \(2018\)](#). For the recent development on the mildly explosive processes with an intercept, see [Fei \(2018\)](#) and [Quo et al. \(2018\)](#).

We impose $X_{ji0} = o_p(\sqrt{k_T})$ as $T \rightarrow \infty$ and independent from $\sigma(u_{i1}, u_{i2}, \dots)$ for $j = 1, 2$. Assumptions 1 and 2 below state the properties of an individual u_{it} .

Assumption 1 (Cross-section independence): u_{it} and u_{ks} are independent for all $i \neq k$ and t, s which also implies that $\mathbb{E}[u_{it}u_{ks}^\top] = \mathbf{O}$ (zero matrix).

Assumption 2 (Individual error moments): The error vectors $\{u_{it}\}$ are a covariance-stationary sequence with $\mathbb{E}[u_{it}] = 0$ and $\mathbb{E}[|u_{it}|^{2+\beta}] < \infty$ for $\beta > 0$. They have an individual specific positive definite covariance matrix:

$$\mathbb{E}[u_{it}u_{it}^\top] = \mathbf{\Omega}_i = \begin{bmatrix} \Omega_{i,11} & \Omega_{i,21} \\ \Omega_{i,21} & \Omega_{i,22} \end{bmatrix} \quad (2.4)$$

It is a weakly dependent sequence with α -mixing dependence structure and the following mixing coefficients:

$$\alpha(m) = \sup_t \sup_{A \in \mathcal{F}_t^\infty, B \in \mathcal{F}_{-m}^{t-m}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \quad (2.5)$$

where $\mathcal{F}_{t_1}^{t_2}$ represents the smallest sigma algebra containing information within the specified time periods. For a generic matrix \mathbf{M} , $\|\mathbf{M}\| = \sqrt{\rho(\mathbf{M}^\top \mathbf{M})}$ represents the spectral norm² and $\rho(\cdot)$ is the largest eigenvalue operator.

This implies that the following Wold decomposition which is free of deterministic components exists (see [Phillips and Solo \(1992\)](#)):

$$u_{it} = \mathbf{C}_i(L)\epsilon_{it} = \sum_{j=0}^{\infty} \mathbf{C}_{ij}\epsilon_{it-j} \quad (2.6)$$

where ϵ_{it} is IID with $\mathbb{E}[\epsilon_{it}] = 0$, $\mathbb{E}\|\epsilon_{it}\|^{2+\beta} < \infty$ for $\beta > 0$ and

$$\mathbb{E}[\epsilon_{it}\epsilon_{it}^\top] = \boldsymbol{\Sigma}_{i\epsilon} = \begin{bmatrix} \sigma_{i,11}^2 & \sigma_{i,21} \\ \sigma_{i,21} & \sigma_{i,22}^2 \end{bmatrix} \quad (2.7)$$

which is positive definite. Here, $\mathbf{C}_i(L)$ is the infinite-order lag polynomial such that $\sum_{j=1}^{\infty} j^{\frac{1}{2}} \|\mathbf{C}_{ij}\| < \infty$.

For the further analysis, we employ the Beveridge-Nelson (BN) decomposition on the lag operator $\mathbf{C}_i(L)$ to separate components of u_{it} :

$$u_{it} = \mathbf{C}_i(L)\epsilon_{it} = \mathbf{C}_i(1)\epsilon_{it} - \Delta\tilde{\epsilon}_{it} \quad (2.8)$$

Here, $\mathbf{C}_i(1) = \sum_{j=0}^{\infty} \mathbf{C}_{ij}$, $\Delta = 1 - L$ and $\tilde{\epsilon}_{it} = \sum_{j=0}^{\infty} \tilde{\mathbf{C}}_{ij}\epsilon_{it-j}$ is a covariance-stationary linear process with $\tilde{\mathbf{C}}_{ij} = \sum_{m=j+1}^{\infty} \mathbf{C}_{im}$. With the BN decomposition and Assumption 2, we can define *one-sided* long-run covariance matrix of u_{it} :

$$\begin{aligned} \boldsymbol{\Lambda}_i &= \sum_{h=1}^{\infty} \mathbb{E}[u_{it}u_{it-h}^\top] \\ &= \begin{bmatrix} \Lambda_{i,11} & \Lambda_{i,21} \\ \Lambda_{i,21} & \Lambda_{i,22} \end{bmatrix} \end{aligned} \quad (2.9)$$

²Note that if \mathbf{M} is a vector, the spectral norm becomes a simple Euclidean norm.

As well as the long-run variance matrix of u_{it} :

$$\begin{aligned}
\mathbf{\Gamma}_i &= \mathbf{\Omega}_i + \mathbf{\Lambda}_i + \mathbf{\Lambda}_i^\top \\
&= \mathbf{C}_i(1)\mathbf{\Sigma}_{i\epsilon}\mathbf{C}_i(1)^\top \\
&= \begin{bmatrix} \Gamma_{i,11} & \Gamma_{i,21} \\ \Gamma_{i,21} & \Gamma_{i,22} \end{bmatrix}
\end{aligned} \tag{2.10}$$

The one-sided covariance matrix $\mathbf{\Lambda}_i$ will be important in describing the non-random bias component of the asymptotic analysis of VAR(1) least squares (LS) estimator in the local to unity part. The long-run variance $\mathbf{\Gamma}_i$ will be used in the Lindeberg condition in Lemma D in the Technical Appendix.

3 Asymptotics of the FMLS Estimator

3.1 Fixed N

We consider the simpler fixed N limits, first. This presents the panel setting but also exposes the problems that arise in the time series case. Using the matrix version of the least squares estimator, we can estimate the coefficient matrix \mathbf{R}_T with the pooled matrix estimator:

$$\hat{\mathbf{R}}_T = \left(\sum_{i=1}^N \sum_{t=2}^T X_{it} X_{it-1}^\top \right) \left(\sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right)^{-1} \tag{3.1}$$

Clearly, since both processes in 2.1 exhibit a different behavior around the unit root, different rates of convergence will apply. Therefore, as in [Phillips and Lee \(2015\)](#), we make use of two asymptotically equivalent normalization matrices \mathbf{D}_T and \mathbf{F}_T . In particular

$$\mathbf{D}_T = \begin{bmatrix} T & 0 \\ 0 & k_T \theta_T^\top \end{bmatrix}, \quad \mathbf{F}_T = \begin{bmatrix} T & 0 \\ 0 & \frac{\theta_T^\top}{\theta_T^2 - 1} \end{bmatrix} \tag{3.2}$$

Here, the asymptotic equivalence comes from the fact that $\theta_T^2 - 1 = 2\frac{b}{k_T} + \frac{b^2}{k_T^2} = 2\frac{b}{k_T} + o(1)$ as $T \rightarrow \infty$. Using \mathbf{F}_T gives the same normalization asymptotically but helps to cancel out the nuisance parameter b . Using \mathbf{D}_T is more handy in proofs. Given these normalization matrices, we formulate a generalized version of the Theorem 2.1 in [Phillips and Lee \(2015\)](#).

Theorem 3.1 Under Assumptions 1 and 2,

$$(\hat{\mathbf{R}}_T - \mathbf{R}_T)\mathbf{F}_T \Rightarrow \boldsymbol{\Phi} + \mathbf{Z} \quad (3.3)$$

as N stays fixed and $T \rightarrow \infty$. Here, $\boldsymbol{\Phi}, \mathbf{Z} \in \mathbb{R}^{2 \times 2}$ are random matrices with the following elements:

$$\boldsymbol{\Phi} = \begin{bmatrix} \frac{\sum_{i=1}^N C_{1i}^2(1) \int_0^1 J_{1ic}(r) dB_{1i}(r)}{\sum_{i=1}^N C_{1i}^2(1) \int_0^1 J_{1ic}^2(r) dr} & \frac{\sum_{i=1}^N C_{1i}(1) C_{2i}(1) X_{2i}(b) Y_{1i}(b)}{\sum_{i=1}^N C_{2i}^2(1) X_{2i}^2(b)} \\ \frac{\sum_{i=1}^N C_{1i}(1) C_{2i}(1) \int_0^1 J_{1ic}(r) dB_{2i}(r)}{\sum_{i=1}^N C_{1i}^2(1) \int_0^1 J_{1ic}^2(r) dr} & \frac{\sum_{i=1}^N C_{2i}^2(1) X_{2i}(b) Y_{2i}(b)}{\sum_{i=1}^N C_{2i}^2(1) X_{2i}^2(b)} \end{bmatrix} \quad (3.4)$$

$$\mathbf{Z} = \begin{bmatrix} \frac{\sum_{i=1}^N \Lambda_{i,11}}{\sum_{i=1}^N C_{1i}^2(1) \int_0^1 J_{1ic}^2(r) dr} & 0 \\ \frac{\sum_{i=1}^N \Lambda_{i,21}}{\sum_{i=1}^N C_{1i}^2(1) \int_0^1 J_{1ic}^2(r) dr} & 0 \end{bmatrix}$$

Here, $J_{jic}(r) \equiv \int_0^r e^{(r-s)c} dB_{ji}(s) = \sigma_{i,jj} \int_0^r e^{(r-s)c} dW_{ji}(s)$ is the zero-mean Ornstein-Uhlenbeck (O-U) process, $B_{ji}(s)$ and $W_{ji}(s)$ are Brownian Motion and standard Wiener process, respectively, for $j = 1, 2$, both defined on the interval $[0, 1]$. $X_{2i}(b) \stackrel{D}{=} Y_{2i}(b) \stackrel{D}{=} \mathcal{N}(0, \frac{\sigma_{i,22}^2}{2b})$ and $Y_{1i}(b) \stackrel{D}{=} \mathcal{N}(0, \frac{\sigma_{i,11}^2}{2b})$.

Remark 1. By the results of Lemma D and Lemma F in the Technical Appendix, $J_{1ic}(r)$ is independent from $X_{2i}(b), Y_{2i}(b)$ and $Y_{1i}(b)$. Plus, $X_{2i}(b), Y_{2i}(b)$ and $Y_{1i}(b)$ are mutually independent, as well as independent for $i = 1, \dots, N$ due to Assumption 1.

The joint convergence is a result of Lemma F and Lemma D. \mathbf{Z} is a bias matrix that is absent in Phillips and Lee (2015) and it occurs due to the weakly dependent errors. Additionally, \mathbf{Z} has a zero column because the mildly explosive part is unbiased (see Lemma E in the Technical Appendix). The proof of Theorem 3.1 can be found in the Technical Appendix. Clearly, (3.4) includes nuisance parameters and c cannot be consistently estimated from a single time series³. This drives degeneracy if (3.4) is used in test statistics.

Note that the stochastic bias matrix \mathbf{Z} disappears for if we use the Fully Modified Least Squares (FMLS) estimator:

$$\hat{\mathbf{R}}_T^{FM} = \left(\sum_{i=1}^N \left[\sum_{t=2}^T X_{it} X_{it-1}^\top - T \hat{\Lambda}_i \right] \right) \left(\sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right)^{-1} \quad (3.5)$$

³The elements in Γ_i and Λ_i can be estimated non-parametrically using a single series. The parameter b can also be found from a single series, however, it disappears in (3.4)

where

$$\hat{\Lambda}_i = \begin{bmatrix} \hat{\Lambda}_{i,11} & \hat{\Lambda}_{i,12} \\ \hat{\Lambda}_{i,21} & \hat{\Lambda}_{i,22} \end{bmatrix} = \sum_{s=1}^{J-1} K\left(\frac{s}{J}\right) \frac{1}{T} \sum_{t=s+1}^T \hat{u}_{it-s} \hat{u}_{it}^\top \quad (3.6)$$

is a consistent estimator of Λ_i as $T \rightarrow \infty$. Here, $K(x) = (1 - |x|)\mathbf{1}(|x| \leq 1)$ is the Bartlett kernel and $J > 0$ is an associated bandwidth parameter. If we use the FMLS estimator, we obtain:

$$(\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T)\mathbf{F}_T \Rightarrow \Phi \quad (3.7)$$

as $T \rightarrow \infty$. To see this, it is sufficient to examine the first term after we substitute (2.1) into (3.5):

$$\begin{aligned} & \sum_{i=1}^N \left[\sum_{t=2}^T u_{it} X_{it-1}^\top - T \hat{\Lambda}_i \right] \mathbf{D}_T^{-1} \\ &= \begin{bmatrix} \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T X_{1it-1} u_{1it} & \sum_{i=1}^N \frac{1}{k_T \theta_T^T} \sum_{t=2}^T X_{2it-1} u_{1it} \\ \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T X_{1it-1} u_{2it} & \sum_{i=1}^N \frac{1}{k_T \theta_T^T} \sum_{t=2}^T X_{2it-1} u_{2it} \end{bmatrix} \\ & \quad - \begin{bmatrix} \sum_{i=1}^N \hat{\Lambda}_{i,11} & \frac{T}{k_T \theta_T^T} \sum_{i=1}^N \hat{\Lambda}_{i,12} \\ \sum_{i=1}^N \hat{\Lambda}_{i,21} & \frac{T}{k_T \theta_T^T} \sum_{i=1}^N \hat{\Lambda}_{i,22} \end{bmatrix} \end{aligned} \quad (3.8)$$

The bias matrix will vanish as $T \rightarrow \infty$ due to the bias correction (the first column) and the fact that $\hat{\Lambda}_i$ is consistent and $\frac{T}{k_T \theta_T^T} = o(1)$ (the second column). This type of estimator will be exploited in the next section in order to get a correct centering of the asymptotic distributions at 0 when $N \rightarrow \infty$.

3.2 Large N

In this section we will invoke the cross-section dimension and sequential limits to simplify (3.4). To simplify the analysis we consider the properties of the following matrices and vectors in two cases as $T \rightarrow \infty$:

When $\mathbf{R}_T = \text{diag}(\rho_T, \theta_T)$:

$$\mathbf{F}_T^{-1} \sum_{t=2}^T X_{it-1} X_{it-1}^\top \mathbf{F}_T^{-1} \Rightarrow \mathbf{B}_i \in \mathbb{R}^{2 \times 2}, \quad \mathcal{B}_i = \mathbb{E}[\mathbf{B}_i] \quad (3.9)$$

$$\mathbf{F}_T^{-1} \text{vec} \left(\left[\sum_{t=2}^T u_{it} X_{it-1}^\top - T \hat{\Lambda}_i \right]^\top \right) \Rightarrow M_i \in \mathbb{R}^4, \quad \mathcal{M}_i = \mathbb{E}[M_i M_i^\top] \quad (3.10)$$

When $\mathbf{R}_T = \rho_T \mathbf{I}_2$:

$$\frac{1}{T^2} \sum_{t=2}^T X_{it-1} X_{it-1}^\top \Rightarrow \mathbf{C}_i \in \mathbb{R}^{2 \times 2}, \quad \Theta_i = \mathbb{E}[\mathbf{C}_i] \quad (3.11)$$

$$\frac{1}{T} \text{vec} \left[\sum_{t=2}^T u_{it} X_{it-1}^\top - T \hat{\Lambda}_i \right] \Rightarrow \Pi_i \in \mathbb{R}^4, \quad \Xi_i = \mathbb{E}[\Pi_i \Pi_i^\top] \quad (3.12)$$

Here, vec represents the vectorization operator. Clearly, for fixed N , (3.9) - (3.12) represent random vectors or matrices with only diffusion processes ($\mathbf{R}_T = \rho_T \mathbf{I}_2$) and diffusion processes together with Normals ($\mathbf{R}_T = \text{diag}(\rho_T, \theta_T)$) from (3.4). Assumption 3 provides requirements on the moment matrices in (3.9) - (3.12).

Assumption 3 (Error heterogeneity). $\mathbb{E} \|u_{it}\|^{4+\beta} < \infty$ and $\mathbb{E} \|\epsilon_{it}\|^{4+\beta} < \infty$ for $\beta > 0$ where u_{it}, ϵ_{it} are the same vectors in \mathbb{R}^2 as in Assumption 2. Also:

- i) $\sup_i \|\mathbf{B}_i\| < \infty, \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbf{B}_i = \mathbf{B}, \|\mathbf{B}\| < \infty,$
- ii) $\sup_i \|\Theta_i\| < \infty, \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Theta_i = \Theta, \|\Theta\| < \infty,$
- iii) $\sup_i \|\mathcal{M}_i\| < \infty, \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathcal{M}_i = \mathcal{M}, \|\mathcal{M}\| < \infty,$
- iv) $\sup_i \|\Xi_i\| < \infty, \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Xi_i = \Xi, \|\Xi\| < \infty,$
- v) $\lim_{N \rightarrow \infty} \left(\sum_{i=1}^N \mathcal{M}_i \right)^{-1} \mathcal{M}_i = \mathbf{O},$
- vi) $\lim_{N \rightarrow \infty} \left(\sum_{i=1}^N \Xi_i \right)^{-1} \Xi_i = \mathbf{O}.$

The conditions i) and ii) are required to invoke the WLLN for the matrices with heterogeneous means as $N \rightarrow \infty$ (see [Moon and Phillips \(2000\)](#)). The conditions iii) - vi) are sufficient to ensure that the multivariate Lindeberg condition for the CLT is satisfied as $N \rightarrow \infty$ and that no individual variance matrix dominates (see Theorem D.19A in [Greene \(2003\)](#) and Proposition 2.27 in [Van der Vaart \(2000\)](#)). They do not depend on the sample size and they will have a finite norm as long as certain products of individual moments of u_{it} coordinates are finite.

The structure of the matrices in (3.9) - (3.12) will be important for the Wald statistic. To present their elements, we follow [Hansen \(1995\)](#) and define the correlation coefficient

λ_i using the long-run variance matrix⁴ in (2.10):

$$\lambda_i = \frac{\Gamma_{i,21}}{\sqrt{\Gamma_{i,11}\Gamma_{i,22}}} \quad (3.13)$$

Here λ_i is the long-run (zero frequency) correlation between u_{1it} and u_{2it} . It is also the long-run correlation between W_{1it} and W_{2it} that is generated by u_{1it} and u_{2it} after the weak convergence in measure as $T \rightarrow \infty$.

The parameter λ_i is used to orthogonalize the Wiener processes generated by u_{1it} and u_{2it} :

$$W_{ji}(r) = \lambda_i W_{ki}(r) + \sqrt{1 - \lambda_i^2} W_{ki}^\perp(r) \quad (3.14)$$

where $j, k = 1, 2$ and $\mathbb{E}[W_{ki}(r)W_{ki}^\perp(r)] = 0$, hence independent because they are Gaussian. We use 3.14 in Proposition 3 in the Technical Appendix to facilitate calculation of covariances between diffusion processes and explicitly describe the matrices in (3.9) - (3.12). We use the following result by [Guillaume \(2017\)](#):

$$\begin{aligned} & \text{Cov} \left[\int_0^t f_1(r, W_{ki}(r)) dW_{ki}(r), \int_0^T f_2(r, W_{ji}(r)) dW_{ji}(r) \right] \\ &= \lambda_i \int_0^t \mathbb{E} \left[f_1(r, W_{ki}(r)) f_2(r, W_{ji}(r)) \right] dr \end{aligned} \quad (3.15)$$

with $t \leq T$, $j, k = 1, 2$. The functions f_1 and f_2 are the non-anticipatory and depend on W_{ki} , W_{ji} and time, respectively. Also, λ_i is the correlation coefficient in (3.13). For our purposes, we have $f_1(r, W_{ki}(r)) = J_{kic}(r)$ and $f_2(r, W_{ji}(r)) = J_{jic}(r)$, implying $f_1 = f_2$. We are able to generalize this approach to cases when k and j are switched in the integrator since we can define an enlarged sigma-algebra \mathcal{F}^{jk} to which $J_{kic}(r)$ and $J_{jic}(r)$ are adapted. Therefore, such integrals exist.

To derive the asymptotic behavior of $\hat{\mathbf{R}}_T^{FM}$, we define two new asymptotically equivalent block-normalization matrices to account for large N :

$$\mathbf{D}_{NT} = \mathbf{I}_2 \otimes \sqrt{N} \mathbf{D}_T, \quad \mathbf{F}_{NT} = \mathbf{I}_2 \otimes \sqrt{N} \mathbf{F}_T \quad (3.16)$$

where \mathbf{D}_T and \mathbf{F}_T are the same normalization matrices for fixed the N as in (3.2). \otimes represents the Kronecker product and \mathbf{I}_2 is a 2×2 identity matrix. For the estimation, we have to use (3.5) in order to remove the bias when invoking CLT. We re-scale the FMLS estimator to account for the second limit as $N \rightarrow \infty$:

$$\hat{\mathbf{R}}_T^{FM} = \left(\sum_{i=1}^N \left[\sum_{t=2}^T X_{it} X_{it-1}^\top - \sqrt{NT} \hat{\mathbf{\Lambda}}_i \right] \right) \left(\sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right)^{-1} \quad (3.17)$$

⁴[Hansen \(1995\)](#) uses two-sided long-run covariance matrix but it coincides with (2.10) by definition.

Here, $\hat{\Lambda}_i$ is the same as in (3.6). Given this, we can formulate the asymptotic result:

Theorem 3.2 *Under Assumptions 1 and 3,*

$$\mathbf{F}_{NT}(\hat{\boldsymbol{\phi}}_T^{FM} - \boldsymbol{\phi}_T) \xrightarrow{D} \mathcal{N}\left(0, (\mathbf{I}_2 \otimes \mathcal{B})^{-1} \mathcal{M} (\mathbf{I}_2 \otimes \mathcal{B})^{-1}\right) \quad (3.18)$$

as $(T, N)_{seq} \rightarrow \infty$, where $(\hat{\boldsymbol{\phi}}_T^{FM} - \boldsymbol{\phi}_T) = \text{vec}([\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T]^\top)$.

The proof of Theorem 3.2 can be found in the Technical Appendix. Following Proposition 3 and denoting $\Gamma_{i,jj} = C_{ji}^2(1)\sigma_{i,jj}^2$ for $j = 1$ or 2 , they are the limiting averages of the following matrices for each i :

$$\mathcal{B}_i = \begin{bmatrix} \Gamma_{i.11} \int_0^1 \int_0^r e^{2(r-s)c} & 0 \\ 0 & \Gamma_{i.22} \end{bmatrix} \quad (3.19)$$

$$\mathcal{M}_i = \begin{bmatrix} \Gamma_{i.11}^2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 & \Gamma_{i.21} \Gamma_{i.11} \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 \\ 0 & \frac{\Gamma_{i.11} \Gamma_{i.22}}{2b} & 0 & 0 \\ \Gamma_{i.21} \Gamma_{i.11} \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 & \Gamma_{i.11} \Gamma_{i.22} \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 \\ 0 & 0 & 0 & \frac{\Gamma_{i.22}^2}{2b} \end{bmatrix} \quad (3.20)$$

The zeros in \mathcal{B}_i occur due to Lemma F (c), while they occur in \mathcal{M}_i due to zero means of the Normals (Lemma D) or the asymptotic independence between the Normals and the functionals of Brownian Motion (Lemma F (a)).

Remark 2. *We still have $\frac{1}{2b}$ in \mathcal{M}_i , because pre-multiplication by \mathbf{F}_{NT} from the left gives $2b$ in the T limit, while the variance terms on the mildly explosive side are scaled by $\frac{1}{4b^2}$.*

We conjecture that the asymptotic behavior of the FMLS estimator can be described under the joint limits for T and N . However, as is well known (see e.g. [Moon and Phillips \(2000\)](#)), the rate of expansion needs to be controlled. Particularly, $\frac{N}{T} \rightarrow 0$ as $(T, N) \xrightarrow{j} \infty$. Under weakly dependent errors, this has to be imposed in order to avoid a rapid accumulation of bias terms. Additionally, we would need stronger assumptions on admissible kernels for the estimator in (3.6) together with assumptions on the bandwidth parameter expansion rate. See Assumption 3 and Assumption 4 in [Moon and Phillips \(2000\)](#). We do not pursue proofs of the joint limits for the clarity of exposition as in [Breitung and Pesaran \(2008\)](#) or [Harris et al. \(2010\)](#).

4 Wald Testing

4.1 Construction of the Wald Statistic

We invoke sequential limit theory to explore the large sample properties of the Wald test statistic. We test $H_0 : \mathbf{R}_T = \rho_T \mathbf{I}_2$ against the alternative $H_1 : \mathbf{R}_T = \text{diag}(\rho_T, \theta_T)$. We can additionally test if the off-diagonal terms are different from zero, however the *block statistic* does not exhibit a proper behavior under H_1 (see Remark 3 below). Similarly to [Phillips and Lee \(2015\)](#), we employ the vector $a_1^\top = [1 \ 0 \ 0 \ -1]$ to compactly write H_0 as $a_1^\top \text{vec}(\mathbf{R}_T) = 0$. We formulate the Wald statistic in the following way:

$$W_{NT} = \frac{\left[a_1^\top \text{vec}(\hat{\mathbf{R}}_T^{FM}) \right]^2}{a_1^\top \left[\mathbf{Q}^{-1} \otimes \mathbf{I}_2 \right] \hat{\Xi} \left[\mathbf{Q}^{-1} \otimes \mathbf{I}_2 \right] a_1} \quad (4.1)$$

In particular, $\mathbf{Q} = \sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \in \mathbb{R}^{2 \times 2}$. Also, $\hat{\mathbf{R}}_T^{FM}$ is the FMLS estimator of \mathbf{R}_T with an additionally scaled modification as in (3.17) using the same $\hat{\Lambda}_i$. The matrix $\hat{\Xi} \in \mathbb{R}^{4 \times 4}$ is a consistent estimator of the limiting average of Ξ_i .

If we construct the consistent estimator of $\hat{\Xi}$, the asymptotic behavior of W_{NT} under the null and Assumption 3 as $(T, N)_{seq} \rightarrow \infty$ is described in the following result:

Theorem 4.1 Under $H_0: \mathbf{R}_T = \rho_T \mathbf{I}_2$, as $(T, N)_{seq} \rightarrow \infty$

$$W_{NT} \xrightarrow{D} \chi_1^2 \quad (4.2)$$

where χ_1^2 represents Chi-squared distribution with 1 degree of freedom.

The proof of Theorem 4.1 can be found in the Technical Appendix. Here, the denominator converges to the variance of a normal variable under the square in (4.1), where the matrix in the quadratic form is $\left[\Theta^{-1} \otimes \mathbf{I}_2 \right] \Xi \left[\Theta^{-1} \otimes \mathbf{I}_2 \right]$ which has the following components:

$$\begin{aligned} \Theta &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Theta_i, \text{ with} \\ \Theta_i &= \begin{bmatrix} \Gamma_{i.11} & \lambda_i \sqrt{\Gamma_{i.11} \Gamma_{i.22}} \\ \lambda_i \sqrt{\Gamma_{i.11} \Gamma_{i.22}} & \Gamma_{i.22} \end{bmatrix} \int_0^1 \int_0^r e^{2(r-s)c} ds dr \\ &= \begin{bmatrix} \Gamma_{i.11} & \Gamma_{i.21} \\ \Gamma_{i.21} & \Gamma_{i.22} \end{bmatrix} \int_0^1 \int_0^r e^{2(r-s)c} ds dr = \Gamma_i \int_0^1 \int_0^r e^{2(r-s)c} ds dr \end{aligned} \quad (4.3)$$

where we find the long-run variance matrix in (2.10) if we use $\lambda_i = \frac{\Gamma_{i,21}}{\sqrt{\Gamma_{i,11}\Gamma_{i,22}}}$. Also:

$$\begin{aligned} \mathfrak{E} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathfrak{E}_i, \text{ with} \\ \mathfrak{E}_i &= \begin{bmatrix} \Gamma_{i,11}^2 & \Gamma_{i,21}\Gamma_{i,11} & \Gamma_{i,21}\Gamma_{i,11} & \Gamma_{i,21}^2 \\ \Gamma_{i,21}\Gamma_{i,11} & \Gamma_{i,11}\Gamma_{i,22} & \Gamma_{i,21}^2 & \Gamma_{i,21}\Gamma_{i,22} \\ \Gamma_{i,21}\Gamma_{i,11} & \Gamma_{i,21}^2 & \Gamma_{i,11}\Gamma_{i,22} & \Gamma_{i,21}\Gamma_{i,22} \\ \Gamma_{i,21}^2 & \Gamma_{i,21}\Gamma_{i,22} & \Gamma_{i,21}\Gamma_{i,22} & \Gamma_{i,22}^2 \end{bmatrix} \int_0^1 \int_0^r e^{2(r-s)c} ds dr \end{aligned} \quad (4.4)$$

The structure of (4.3) and (4.4) is derived in Proposition 3 (g) in the Technical Appendix. By using a panel dimension, we do not need the additional re-scaling of the statistic by a diverging function L_T (such that $\frac{k_T^2 L_T}{T^2} \rightarrow 0$ as $T \rightarrow \infty$) which is the case in [Phillips and Lee \(2015\)](#). This means that we do not need to have any knowledge about the functional form of k_T and we allow the Wald statistic to have a distribution, instead of asymptotically letting it converge to 0 and resorting to any arbitrary set of critical values.

A consistent estimation of \mathfrak{E} also allows the statistic to diverge under H_1 . Using the same normalization as in Theorem 4.1, we can show that under $H_1 : \mathbf{R}_T = \text{diag}(\rho_T, \theta_T)$ the Wald test statistic diverges for any $b > 0$ and $c < 0$.

Theorem 4.2 *Under $H_1 : \mathbf{R}_T = \text{diag}(\rho_T, \theta_T)$ as $(T, N)_{seq} \rightarrow \infty$*

$$W_{NT} \rightarrow \infty \quad (4.5)$$

because the numerator has a dominant term $\left(\frac{\sqrt{NT}b}{k_T}\right)^2$ and the denominator converges to

$$\left(\frac{\mathcal{S}_2}{\mathcal{S}_1} \int_0^1 \int_0^r e^{2(r-s)c} ds dr\right)^{-1}$$

Here, $\mathcal{S}_1 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i,11}^2$ and $\mathcal{S}_2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i,11}$.

Clearly, under the Assumption 3, the denominator is finite, since the integral is always finite for $c < 0$. Therefore, under the alternative, W_{NT} diverges as $(T, N)_{seq} \rightarrow \infty$.

The proof of Theorem 4.2 can be found in the Technical Appendix. Under the alternative hypothesis, the denominator matrix in the quadratic form of (4.1) is asymptotically singular. However, we investigate a special case where we avoid a potentially asymptotically undefined statistic.

4.2 Estimation of the Parameters in the Wald Statistic

As suggested by the Theorem 3.2 and also following [Hjalmarsson \(2006\)](#)⁵, a natural estimator of Ξ would have the form of the sample moment matrix:

$$\hat{\Xi} = \frac{1}{N} \sum_{i=1}^N \left[\left(\frac{1}{T} \sum_{t=2}^T \text{vec}[\hat{u}_{it} X_{it-1}^\top - T \hat{\Lambda}_i] \right) \left(\frac{1}{T} \sum_{s=2}^T \text{vec}[\hat{u}_{is} X_{is-1}^\top - T \hat{\Lambda}_i] \right)^\top \right] \quad (4.6)$$

where \hat{u}_{it} is the residual vector for the individual i . Using $\hat{u}_{it} = u_{it} - (\hat{\mathbf{R}}_T - \mathbf{R}_T) X_{it-1}$ we receive:

$$\begin{aligned} \text{vec} \left[\frac{1}{T} \sum_{t=2}^T \hat{u}_{it} X_{it-1}^\top - \hat{\Lambda}_i \right] &= \text{vec} \left[\frac{1}{T} \sum_{t=2}^T u_{it} X_{it-1}^\top - \hat{\Lambda}_i - T(\hat{\mathbf{R}}_T - \mathbf{R}_T) \frac{1}{T^2} \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right] \\ &= \text{vec} \left[\frac{1}{T} \sum_{t=2}^T u_{it} X_{it-1}^\top - \hat{\Lambda}_i \right] + O_p(1) \end{aligned} \quad (4.7)$$

as $T \rightarrow \infty$ under $H_0 : \mathbf{R}_T = \rho_T \mathbf{I}_2$. The bias term in the T limit can be simplified using the N limit, however, (4.7) becomes explosive under H_1 :

$$\text{vec} \left[\frac{1}{T} \sum_{t=2}^T u_{it} X_{it-1}^\top \right] = \left[\dots, \frac{k_T \theta_T^\top}{T} \frac{1}{k_T \theta_T^\top} \sum_{t=2}^T X_{2it-1} u_{1it}, \frac{k_T \theta_T^\top}{T} \frac{1}{k_T \theta_T^\top} \sum_{t=2}^T X_{2it-1} u_{2it} \right]^\top \quad (4.8)$$

The last two terms in the vector are $O_p\left(\frac{k_T \theta_T^\top}{T}\right)$ as $T \rightarrow \infty$.

Therefore, we follow an alternative approach. Since we have an analytic expression of Ξ_i in (4.4) using the generalized Itô Isometry in [Guillaume \(2017\)](#), we define the direct estimator which is fixed under both H_0 and H_1 in the following expression:

$$\hat{\Xi} = \sum_{i=1}^N T^2 \hat{\Xi}_i = \sum_{i=1}^N \left(T^2 \hat{\Xi}_i^{NP} \right) \int_0^1 \int_0^r e^{2(r-s)\hat{\epsilon}} ds dr \quad (4.9)$$

where $\hat{\Xi}_i^{NP}$ is the non-parametric estimator part of the matrix in (4.4). We additionally scale the estimator for the individual i by T^2 because the statistic in (4.1) will not then

⁵Such estimator was employed in panel predictive regression setup with the local to unity predictor.

need extra scaling when $\hat{\Xi}$ is inserted and $\frac{1}{N}$ will occur in front of (4.9) due to self-normalization. For each individual, $\hat{\Xi}_i^{NP}$ can be fully constructed from $\hat{\Gamma}_i$. Following [Hansen \(1995\)](#) and the definition in (2.10):

$$\begin{aligned} \hat{\Gamma}_i &= \begin{bmatrix} \hat{\Gamma}_{i.11} & \hat{\Gamma}_{i.21} \\ \hat{\Gamma}_{i.21} & \hat{\Gamma}_{i.22} \end{bmatrix} = \frac{1}{T} \sum_{t=1}^T \hat{u}_{it} \hat{u}_{it}^\top + \sum_{s=0}^{J-1} K\left(\frac{s}{J}\right) \frac{1}{T} \sum_{t=1}^T \hat{u}_{it-s} \hat{u}_{it}^\top \\ &+ \left(\sum_{s=0}^{J-1} K\left(\frac{s}{J}\right) \frac{1}{T} \sum_{t=1}^T \hat{u}_{it-s} \hat{u}_{it}^\top \right)^\top \end{aligned} \quad (4.10)$$

where K is an admissible kernel (e.g. Bartlett as in (3.6) or Parzen) which produces positive definite matrices and J is, again, a bandwidth parameter with an expansion rate slower than the expansion of the sample size, i.e. $\frac{J}{T} \rightarrow 0$ as $J, T \rightarrow \infty$.

For the parametric component, the natural estimator of c is $T(\hat{\rho}_T - 1)$, because $c = T(\rho_T - 1)$ as in [Moon and Phillips \(2000\)](#). Clearly, due to weakly dependent errors, we have to use the FMLS estimator in the following equation:

$$X_{1it} = \rho_T X_{1it-1} + u_{1it} \quad (4.11)$$

from which we find:

$$T(\hat{\rho}_T^{FM} - \rho_T) = \frac{\frac{1}{N} \sum_{i=1}^N \left[\frac{1}{T} \sum_{t=2}^T X_{1it-1} u_{1it} - NT \hat{\Lambda}_{i.11} \right]}{\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T X_{1it-1}^2} \quad (4.12)$$

Using $\rho_T = 1 + \frac{c}{T}$, we get $\hat{c} - c = T(\hat{\rho}_T^{FM} - 1) - c = T(\hat{\rho}_T^{FM} - \rho_T)$. Clearly from (4.12), $\hat{c} \xrightarrow{P} c$ sequentially because $\frac{1}{N} \sum_{i=1}^N C_{1i}^2(1) \int_0^1 J_{1ic}(r) dB_{1i}(r) \xrightarrow{P} 0$ as $N \rightarrow \infty$ which we found by using WLLN. Then, by the CMT:

$$\int_0^1 \int_0^r e^{2(r-s)\hat{c}} ds dr \xrightarrow{P} \int_0^1 \int_0^r e^{2(r-s)c} ds dr \quad (4.13)$$

Remark 3. Similarly to [Phillips and Lee \(2015\)](#) we can generalize the Wald test and define the block test with $H_0 : \mathbf{A}^\top rvec(\mathbf{R}_T) = 0$. Here

$$\mathbf{A}^\top = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} a_1^\top \\ a_2^\top \\ a_3^\top \end{bmatrix} \quad (4.14)$$

is the matrix which additionally takes into account the off-diagonal terms in \mathbf{R}_T and tests if the coefficients are zero. Also, $rvec$ is the row vectorization operator⁶. The block (B) Wald test statistic has the following form and the behavior under the null:

$$W_{NT}^B = \left[\mathbf{A}^\top rvec(\hat{\mathbf{R}}_T^{FM}) \right]^\top \left(\mathbf{A}^\top \left[\mathbf{I}_2 \otimes \mathbf{Q}^{-1} \right] \hat{\mathbf{\Xi}} \left[\mathbf{I}_2 \otimes \mathbf{Q}^{-1} \right] \mathbf{A} \right)^{-1} \left[\mathbf{A}^\top rvec(\hat{\mathbf{R}}_T^{FM}) \right] \quad (4.15)$$

$$\xrightarrow{D} \chi_3^2$$

as $(T, N)_{seq} \rightarrow \infty$ which is Chi-squared distribution with 3 degrees of freedom. However, under H_1 the statistic is asymptotically undefined because the matrix in the quadratic form is singular. Contrary to the case in (4.2), we cannot escape this problem because the matrix needs to be actually inverted in the block statistic. See the discussion in Proposition 4 in the Technical Appendix.

5 Concluding Remarks

In this paper, we re-visited the theory for bridging the asymptotic behavior of the local to unity and the mildly explosive processes when they occur in a single estimation procedure. Converting the time series setting to the panel setting and making use of the additional identification power that stems from the cross-section, we resolve the problem of the asymptotic degeneracy of the Wald test statistic under the null. Our approach hinges on using sequential limit analysis and the direct estimator of the covariance matrix which enters the Wald statistic. Similarly to Phillips and Lee (2016) in the time series case, we find a Chi-squared distribution by employing the FMLS estimator. This is simpler than using IVX that turned out to be sub-optimal for such mixed to unity cases in time series.

⁶ $rvec(\mathbf{M}) = vec(\mathbf{M}^\top)$ for a generic matrix \mathbf{M} , therefore the order of the Kronecker product in the (4.15) changes. However, $\hat{\mathbf{\Xi}}$ stays the same

References

- Alvarez, J., & Arellano, M. (2004). *Robust likelihood estimation of dynamic panel data models*. CEMFI, Centro de Estudios Monetarios y Financieros.
- Arvanitis, S., & Magdalinos, T. (2018). Mildly explosive autoregression under stationary conditional heteroskedasticity. *Working Paper*.
- Binder, M., Hsiao, C., & Pesaran, H. (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory*, 21(4), 795–837.
- Breitung, J., & Pesaran, H. (2008). *Unit roots and cointegration in panels*. Springer.
- Brockwell, P., & Davis, R. (2013). *Time series: theory and methods*. Springer Science & Business Media.
- Chan, N. H., & Wei, C.-Z. (1987). Asymptotic inference for nearly nonstationary ar (1) processes. *The Annals of Statistics*, 1050–1063.
- Elliott, G., & Jansson, M. (2003). Testing for unit roots with stationary covariates. *Journal of econometrics*, 115(1), 75–89.
- Fei, Y. (2018). Limit theory for mildly integrated process with intercept. *Economics Letters*, 163, 98–101.
- Giraitis, L., & Phillips, P. C. (2012). Mean and autocovariance function estimation near the boundary of stationarity. *Journal of Econometrics*, 169(2), 166–178.
- Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.
- Guillaume, T. (2017). A useful result on the covariance between ito integrals. *Journal of Advances in Mathematics and Computer Science*, 25(6), 1–12.
- Hansen, B. E. (1995). Rethinking the univariate approach to unit root testing: Using covariates to increase power. *Econometric Theory*, 11(5), 1148–1171.
- Harris, D., Harvey, D. I., Leybourne, S. J., & Sakkas, N. D. (2010). Local asymptotic power of the im-pesaran-shin panel unit root test and the impact of initial observations. *Econometric Theory*, 26(1), 311–324.
- Hjalmarsson, E. (2005). Estimation of average local-to-unity roots in heterogenous panels. *FRB International Finance Discussion Paper No. 852*.
- Hjalmarsson, E. (2006). Predictive regressions with panel data. *International Finance Discussion Papers*.
- Im, K. S., Pesaran, H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of econometrics*, 115(1), 53–74.
- Magdalinos, T., & Phillips, P. C. (2009a). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory*, 25(2), 482–526.
- Moon, H., & Phillips, P. (2000). Estimation of autoregressive roots near unity using panel data. *Econometric Theory*, 16(6), 927–997.
- Øksendal, B. (2003). *Stochastic differential equations*. Springer.
- Pedroni, P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the ppp hypothesis. *Econometric theory*, 20(3), 597–625.

- Phillips, P. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3), 535–547.
- Phillips, P., & Lee, J. H. (2015). Limit theory for vars with mixed roots near unity. *Econometric Reviews*, 34(6-10), 1035–1056.
- Phillips, P., & Lee, J. H. (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. *Journal of Econometrics*, 192(2), 433–450.
- Phillips, P., & Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1), 115–130.
- Phillips, P., & Magdalinos, T. (2009b). Econometric inference in the vicinity of unity. *Singapore Management University, CoFie Working Paper*, 7.
- Phillips, P., Magdalinos, T., & Giraitis, L. (2010). Smoothing local-to-moderate unit root theory. *Journal of Econometrics*, 158(2), 274–279.
- Phillips, P., & Moon, H. (1999b). Linear regression limit theory for nonstationary panel data [Cowles Foundation Discussion Papers]. (1222). Retrieved from <https://EconPapers.repec.org/RePEc:cwl:cwldpp:1222>
- Phillips, P., Shi, S., & Yu, J. (2015). Testing for multiple bubbles: Limit theory of real-time detectors. *International Economic Review*, 56(4), 1079–1134.
- Phillips, P., & Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, 971–1001.
- Phillips, P., Wu, Y., & Yu, J. (2011). Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International economic review*, 52(1), 201–226.
- Quo, G., Sun, Y., & Wang, S. (2018). Testing for moderate explosiveness. *The Econometrics Journal*.
- Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.
- Westerlund, J., & Smeekes, S. (2018). Robust block bootstrap panel predictability tests. *Forthcoming in Econometric Reviews*.

Technical Appendix

Proofs of Lemmas B, C, D, E and F are based on techniques from Phillips and Lee (2015), Phillips and Magdalinos (2007) and Phillips and Magdalinos (2009b)

Auxiliary Propositions and Lemmas for Convergence

Proposition 1

- a) For any $c \in \mathbb{R}$ such that $1 + \frac{c}{T} > 0$, $1 + \frac{c}{T} = \exp(c/T\{1 + o(1)\})$ as $T \rightarrow \infty$.
- b) For each $b > 0$, $\theta_T^{-T} = o(k_T/T)$ as $T \rightarrow \infty$.
- c) $k_T(\theta_T^2 - 1) \rightarrow 2b$ as $T \rightarrow \infty$.

Proof of Proposition 1

a) Using the fact that $\log(1 + x) = x + O(x^2)$ as $x \rightarrow 0$, we receive the following as $T \rightarrow \infty$:

$$\begin{aligned}\log\left(1 + \frac{c}{T}\right) &= \frac{c}{T} + O(1/T^2) \\ &= \frac{c}{T}\{1 + O(1/T)\} \\ &= \frac{c}{T}\{1 + o(1)\} \implies 1 + \frac{c}{T} = \exp\left(\frac{c}{T}\{1 + o(1)\}\right)\end{aligned}$$

as $T \rightarrow \infty$. □

b) Similarly to the part a) and using properties of the natural logarithm we obtain:

$$\begin{aligned}\log\left(\frac{T}{k_T}\theta_T^{-T}\right) &= -T\log(\theta) + \log\left(\frac{T}{k_T}\right) \\ &= -T\log(1 + b/k_T) + \log\left(\frac{T}{k_T}\right) \\ &= -T[b/k_T + O(1/k_T^2)] + \log\left(\frac{T}{k_T}\right) \\ &= -T\frac{b}{k_T}\left[1 + \frac{1}{b}\frac{\log(T/k_T)}{T/k_T} + O(1/k_T)\right] \\ &= -T\frac{b}{k_T}[1 + o(1)] \implies \frac{T}{k_T}\theta_T^{-T} = \exp\left(-T\frac{b}{k_T}[1 + o(1)]\right) = o(1)\end{aligned}$$

as $T \rightarrow \infty$ since $b > 0$. □

c)

$$k_T(\theta_T^2 - 1) = k_T\left(\frac{2b}{k_T} + \frac{b^2}{k_T^2}\right) = 2b + \frac{b^2}{k_T} \rightarrow 2b$$

as $T \rightarrow \infty$

□

Proposition 2

a) $\frac{1}{\sqrt{T}}X_{1it} \Rightarrow C_{1i}(1) \int_0^r e^{(r-s)c} dB_{1i}(s) \equiv C_{1i}(1)J_{1ic}(r) \equiv C_{1i}(1)\sigma_{i.11}\check{J}_{1ic}(r)$ for $i = 1, \dots, N$ ($\check{J}_{1ic}(r)$ is the notation that will be relevant for the calculation of variances in Proposition 3).

b) $\frac{1}{T^2} \sum_{t=2}^T X_{1it-1}^2 \Rightarrow C_{1i}^2(1) \int_0^1 J_{1ic}^2(r) dr$ for $i = 1, \dots, N$

Proof of Proposition 2

a)

$$\begin{aligned} \frac{1}{\sqrt{T}}X_{1it} &= \frac{1}{\sqrt{T}} \sum_{j=1}^t \rho_T^{t-j} u_{1ij} + o_p(1) = \frac{1}{\sqrt{T}} \sum_{j=1}^t \rho_T^{t-j} C_{1i}(1) \epsilon_{1ij} + o_p(1) - \frac{1}{\sqrt{T}} \sum_{j=1}^t \rho_T^{t-j} \Delta \tilde{\epsilon}_{1ij} \\ &= \frac{1}{\sqrt{T}} \sum_{j=1}^t \rho_T^{t-j} C_{1i}(1) \epsilon_{1ij} + o_p(1) - \frac{\tilde{\epsilon}_{1it}}{\sqrt{T}} + \sum_{j=1}^t \Delta \rho_T^{t-j+1} \tilde{\epsilon}_{1ij-1} \end{aligned}$$

The last two terms converge to 0 in $\mathcal{L}1$ norm for any $t = 1, \dots, T$. Given that $\Delta \rho_T^{t-j+1} = \frac{c}{T} \rho_T^{t-j}$, we obtain:

$$\begin{aligned} \mathbb{E} \left[\left| -\frac{\tilde{\epsilon}_{1it}}{\sqrt{T}} + \frac{1}{\sqrt{T}} \sum_{j=1}^t \Delta \rho_T^{t-j+1} \tilde{\epsilon}_{1ij-1} \right| \right] &\leq \frac{\mathbb{E}[|\tilde{\epsilon}_{1i1}|]}{\sqrt{T}} + \frac{c}{T^{\frac{3}{2}}} \sum_{j=1}^t \exp \left[(t-j) \frac{c}{T} \{1 + o(1)\} \right] \mathbb{E}[|\tilde{\epsilon}_{1i1}|] \\ &\leq \frac{\mathbb{E}[|\tilde{\epsilon}_{1i1}|]}{\sqrt{T}} + \mathbb{E}[|\tilde{\epsilon}_{1i1}|] \frac{tc}{T^{\frac{3}{2}}} = o(1) \end{aligned}$$

as $T \rightarrow \infty$ because $\mathbb{E}[|\tilde{\epsilon}_{1i1}|] < \infty$. The last inequality comes from the fact that $0 < \exp \left[(t-j) \frac{c}{T} \{1 + o(1)\} \right] \leq 1$ for all $j = 1, \dots, t$ since $c < 0$. The first term converges according to the functional theory in [Phillips \(1987\)](#). □

b) The result follows from part a), CMT with the function $f(x) = x^2$ and the fact that $\frac{t}{T} - \frac{t-1}{T} = \frac{1}{T}$. □

Proposition 3

- a) $Var \left[C_{1i}^2(1) \int_0^1 J_{1ic}(r) dB_{1i}(r) \right] = C_{1i}^4(1) \sigma_{i.11}^4 \int_0^1 \int_0^r e^{2(r-s)c} ds dr$
- b) $\mathbb{E} \left[C_{1i}^2(1) \int_0^1 J_{1ic}^2(r) dr \right] = C_{1i}^2(1) \sigma_{i.11}^2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr$
- c) $\mathbb{E} \left[C_{1i} C_{2i} \int_0^1 J_{1ic}(r) J_{2ic}(r) dr \right] = C_{1i}(1) C_{2i}(1) \sigma_{i.11} \sigma_{i.22} \lambda_i \int_0^1 \int_0^r e^{2(r-s)c} ds dr$
- d) $Var[C_{1i}(1) C_{2i}(1) X_{2i}(b) Y_{1i}(b)] = C_{1i}^2(1) C_{2i}^2(1) \frac{\sigma_{i.11}^2 \sigma_{i.22}^2}{4b^2}$
- e) $Var[C_{2i}^2(1) X_{2i}(b) Y_{2i}(b)] = C_{2i}^4(1) \frac{\sigma_{i.22}^4}{4b^2}$
- f) $\mathbb{E}[C_{2i}^3(1) C_{1i}(1) X_{2i}^2(b) Y_{2i}(b) Y_{1i}(b)] = 0$

g) Cross-terms. Here, we are interested in the occurrence of the correlation coefficient λ_i . Therefore, we omit long-run variance terms for brevity and adopt the \sim notation immediately. Also, $j, k = 1, 2$ and $j \neq k$.

- i) $\mathbb{E} \left[\left(\int_0^1 \tilde{J}_{kic}(r) dW_{ji}(r) \right)^2 \right] = \int_0^1 \int_0^r e^{2(r-s)c} ds dr$
- ii) $\mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ji} \int_0^1 \tilde{J}_{jic}(r) dW_{ki}(r) \right] = \lambda_i^2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr$
- iii) $\mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \int_0^1 \tilde{J}_{jic}(r) dW_{ki}(r) \right] = \lambda_i \int_0^1 \int_0^r e^{2(r-s)c} ds dr$
- iv) $\mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \int_0^1 \tilde{J}_{kic}(r) dW_{ji}(r) \right] = \lambda_i \int_0^1 \int_0^r e^{2(r-s)c} ds dr$
- v) $\mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \int_0^1 \tilde{J}_{jic}(r) dW_{ji}(r) \right] = \lambda_i^2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr$

*Here, we write the scaling constants explicitly. In Theorems 3.2, 4.1 and 4.2 we compactly use $C_{i,jj}^2(1) \sigma_{i,jj}^2 = \Gamma_{i,jj}$ with $j = 1$ or 2 .

Proof of Proposition 3

a)

$$\begin{aligned} \mathbb{E} \left[\left(C_{1i}^2(1) \int_0^1 J_{1ic}(r) dB_{1i}(r) \right)^2 \right] &= C_{1i}^4(1) \sigma_{i.11}^4 \mathbb{E} \left[\left(\int_0^1 \tilde{J}_{1ic}(r) dW_{1i}(r) \right)^2 \right] \\ &= C_{1i}^4(1) \sigma_{i.11}^4 \int_0^1 \mathbb{E}[\tilde{J}_{1ic}^2(r)] dr \\ &= C_{1i}^4(1) \sigma_{i.11}^4 \int_0^1 \int_0^r e^{2(r-s)c} ds dr \end{aligned}$$

where the second and the third equality are obtained by using the $\hat{\text{I}}$ to Isometry twice. The proof for the process 2 is identical. \square

b)

$$\begin{aligned}\mathbb{E}\left[C_{1i}^2(1)\int_0^1 J_{1ic}^2(r)dr\right] &= C_{1i}^2(1)\sigma_{i.11}^2\int_0^1\int_{\Omega}\tilde{J}_{1ic}^2(\omega,r)d\mathbb{P}(\omega)dr \\ &= C_{1i}^2(1)\sigma_{i.11}^2\int_0^1\mathbb{E}[\tilde{J}_{1ic}^2(r)]dr \\ &= C_{1i}^2(1)\sigma_{i.11}^2\int_0^1\int_0^r e^{2(r-s)c}dsdr\end{aligned}$$

where we exchanged the expectation and the inner integral by using Fubini's Theorem on $\Omega \times \mathbb{R}$, where Ω is our sample space. The proof for the process 2 is identical. \square

c)

$$\begin{aligned}\mathbb{E}\left[C_{1i}(1)C_{2i}(1)\int_0^1 J_{1ic}(r)J_{2ic}(r)dr\right] &= C_{1i}(1)C_{2i}(1)\sigma_{i.11}\sigma_{i.22}\int_0^1\int_{\Omega}\tilde{J}_{1ic}(\omega,r)\tilde{J}_{2ic}(\omega,r)dP(\omega)dr \\ &= C_{1i}(1)C_{2i}(1)\sigma_{i.11}\sigma_{i.22}\int_0^1\mathbb{E}[\tilde{J}_{1ic}(r)\tilde{J}_{2ic}(r)]dr \\ &= C_{1i}(1)C_{2i}(1)\sigma_{i.11}\sigma_{i.22}\lambda_i\int_0^1\int_0^r e^{2(r-s)c}dsdr\end{aligned}$$

where we moved from the second to the third equality using the generalized $\hat{\text{I}}$ to Isometry by [Guillaume \(2017\)](#) as in (3.15) in order to find the covariance between zero-mean $\hat{\text{I}}$ to Integrals that are functionals of two different dependent Wiener processes. \square

d)

$$\begin{aligned}\mathbb{E}[(C_{1i}(1)C_{2i}(1)X_{2i}(b)Y_{1i}(b))^2] &= C_{1i}^2(1)C_{2i}^2(1)\mathbb{E}[X_{2i}^2(b)]\mathbb{E}[Y_{1i}^2(b)] \\ &= C_{1i}^2(1)C_{2i}^2(1)\frac{\sigma_{i.11}^2\sigma_{i.22}^2}{4b^2}\end{aligned}$$

because of independence shown in Lemma D. \square

e)

$$\begin{aligned}\mathbb{E}[(C_{2i}^2(1)X_{2i}(b)Y_{2i}(b))^2] &= C_{2i}^4(1)\mathbb{E}[X_{2i}^2(b)]\mathbb{E}[Y_{2i}^2(b)] \\ &= C_{2i}^4(1)\frac{\sigma_{i.22}^4}{4b^2}\end{aligned}$$

because of independence shown in Lemma D. \square

f)

$$\begin{aligned} & \mathbb{E}[C_{2i}^3(1)C_{1i}(1)X_{2i}^2(b)Y_{2i}(b)Y_{1i}(b)] \\ &= C_{2i}^3(1)C_{1i}(1)\mathbb{E}[X_{2i}^2(b)]\mathbb{E}[Y_{2i}(b)]\mathbb{E}[Y_{1i}(b)] = 0 \end{aligned}$$

because of independence shown in Lemma D. □

g) The proof follows the similar lines provided in [Guillaume \(2017\)](#), where $\hat{\text{Ito}}$ Integrals are constructed as the limits of the following incremental sums: $\sum_{s=0}^{n-1} f_s(W(r_{s+1}) - W(r_s))$. Given that \mathcal{F} is the limiting sigma-algebra that nests all $\mathcal{F}_r \subset \mathcal{F}_{r+1}$ and \mathcal{B} is Borel sigma-algebra on \mathbb{R} , then $f(r)$ is a random function ($f(r, \omega)$) such that:

- 1) $f : [0, \infty] \times \Omega \rightarrow \mathbb{R}$;
- 2) $\mathcal{F} \times \mathcal{B}$ measurable;
- 3) \mathcal{F}_r - adapted;
- 4) $\mathbb{E} \left[\int_S^T f^2(r) dr \right] < \infty$ for finite $S < T$

It can be approximated by a simple process $\sum_{j=0}^{n-1} f_s \mathbf{1}\{r_s, r_{s+1}\}(r)$. Here, f_s 's are square-integrable random variables. The approximation converges to $f(r)$ in the mean square sense and its existence is guaranteed if 1) - 4) are satisfied (see [Øksendal \(2003\)](#), p. 27-28). This approximation allows us to exchange the limit and the expectation operator. Additionally, f_s 's are \mathcal{F}_{r_s} measurable.

The key point to notice, is that we can define \mathcal{F}^{jk} which covers both processes for $j \neq k$ because we have a vector of dependent Wiener processes. This is necessary for the existence of $\hat{\text{Ito}}$ Integrals of our required form:

$$I_{kj} = \int_0^1 \tilde{J}_{kic}(r) dW_{ji}(r)$$

Once their existence is guaranteed, we can proceed with calculating variances and covariances by conditioning on \mathcal{F}^j or \mathcal{F}^k individually as in [Guillaume \(2017\)](#). This is important in order to avoid complications when conditioning products of the different Wiener processes on an expanded sigma-algebra that is admissible to all of them. Such product is not a martingale with respect to an expanded sigma-algebra.

We firstly invoke the discussed approximation by [Guillaume \(2017\)](#):

$$\tilde{J}_{kic}(r) = \lim_{n \rightarrow \infty} \sum_{s=0}^{n-1} f_s(W_{ki}) \mathbf{1}\{r_s, r_{s+1}\}(r)$$

because O-U process satisfies 1) - 4). We put the Wiener process in the argument to stress the fact that this approximated noise component of O-U process is a function of

W_{ki} . Because O-U processes for $k = 1, 2$ are Itô Integrals themselves which differ only by the Wiener process inside but retain the same exponential function as the integrand, we stress the heterogeneity by inserting the different Wiener process in the argument.

Given that $s < m$ for two natural numbers s and m , we have $\sup_s((s+1) - m) = 0$. Also, if $s > m$, then $\sup_m((m+1) - s) = 0$. Using this in connection to the decomposition $W_{ji}(r) = \lambda_i W_{ki}(r) + \sqrt{1 - \lambda_i^2} W_{ki}^\perp(r)$ where $W_{ki}^\perp(r)$ is a Wiener process uncorrelated with W_{ki} , we first show that:

$$\begin{aligned} & \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s))(W_{ji}(r_{m+1}) - W_{ji}(r_m)) \right] \\ &= \mathbb{E} \left[\left(W_{ki}(r_{s+1}) - W_{ki}(r_s) \right) \left(\lambda_i W_{ki}(r_{m+1}) + \sqrt{1 - \lambda_i^2} W_{ki}^\perp(r_{m+1}) \right. \right. \\ & \quad \left. \left. - \lambda_i W_{ki}(r_m) - \sqrt{1 - \lambda_i^2} W_{ki}^\perp(r_m) \right) \right] \\ &= \lambda_i \mathbb{E} \left[(W_{ki}(r_{s+1}) W_{ki}(r_{m+1})) \right] - \lambda_i \mathbb{E} \left[W_{ki}(r_{s+1}) W_{ki}(r_m) \right] \\ & \quad - \lambda_i \mathbb{E} \left[W_{ki}(r_s) W_{ki}(r_{m+1}) \right] + \lambda_i \mathbb{E} \left[W_{ki}(r_s) W_{ki}(r_m) \right] \end{aligned}$$

Therefore, if $s < m$ and $s > m$, we have, respectively:

$$\begin{aligned} \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s))(W_{ji}(r_{m+1}) - W_{ji}(r_m)) \right] &= \lambda_i [r_{s+1} - r_{s+1} - r_s + r_s] = 0 \\ \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s))(W_{ji}(r_{m+1}) - W_{ji}(r_m)) \right] &= \lambda_i [r_{m+1} - r_m - r_{m+1} + r_m] = 0 \end{aligned}$$

This result is important for the cases ii), iii) and v) when re-writing the covariances as the products of incremental sums. However, we can use the decomposition of Wiener process directly in the integral for the case i) and iv) because we will conveniently obtain a product of W_{ki} and W_{ki}^\perp which is 0 in expectation.

i) Using the aforementioned decomposition, we write the integral as:

$$\int_0^1 \tilde{J}_{kic}(r) dW_{ji}(r) = \lambda_i \int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) + \sqrt{1 - \lambda_i^2} \int_0^1 \tilde{J}_{kic}(r) dW_{ki}^\perp(r)$$

Therefore, the variance can be calculated as:

$$\begin{aligned} & \mathbb{E} \left[\left(\int_0^1 \tilde{J}_{kic}(r) dW_{ji}(r) \right)^2 \right] \\ &= \lambda_i^2 \mathbb{E} \left[\left(\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \right)^2 \right] + 2\lambda_i \sqrt{1 - \lambda_i^2} \mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \int_0^1 \tilde{J}_{kic}(r) dW_{ki}^\perp(r) \right] \\ & \quad + (1 - \lambda_i^2) \mathbb{E} \left[\left(\int_0^1 \tilde{J}_{kic}(r) dW_{ki}^\perp(r) \right)^2 \right] \end{aligned}$$

The middle covariance term is zero. To see this, we re-write:

$$\begin{aligned}
& \mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \int_0^1 \tilde{J}_{kic}(r) dW_{ki}^\perp(r) \right] \\
&= \mathbb{E} \left[\lim_{n \rightarrow \infty} \sum_{s=0}^{n-1} f_s(W_{ki}(r_s)) (W_{ki}(r_{s+1}) - W_{ki}(r_s)) \lim_{n \rightarrow \infty} \sum_{m=0}^{n-1} f_m(W_{ki}(r_m)) (W_{ki}^\perp(r_{m+1}) - W_{ki}^\perp(r_m)) \right] \\
&= \lim_{n \rightarrow \infty} \sum_{s=0}^{n-1} \sum_{m=0}^{n-1} \mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ki}(r_m)) (W_{ki}(r_{s+1}) - W_{ki}(r_s)) \right] \mathbb{E} \left[(W_{ki}^\perp(r_{m+1}) - W_{ki}^\perp(r_m)) \right] = 0
\end{aligned}$$

where we split the expectation because of the independence which comes from the decomposition and the fact that the expectation of the Wiener increments is 0. Clearly, the first and the third terms satisfy the Itô Isometry, therefore:

$$\begin{aligned}
\mathbb{E} \left[\left(\int_0^1 \tilde{J}_{kic}(r) dW_{ji}(r) \right)^2 \right] &= \lambda_i^2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr + (1 - \lambda_i^2) \int_0^1 \int_0^r e^{2(r-s)c} ds dr \\
&= \int_0^1 \int_0^r e^{2(r-s)c} ds dr
\end{aligned}$$

□

ii) The covariance term can be computed in the following way:

$$\begin{aligned}
& \mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ji}(r) \int_0^1 \tilde{J}_{jic}(r) dW_{ki}(r) \right] \\
&= \lim_{n \rightarrow \infty} \sum_{s=0}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_s(W_{ji}(r_s)) (W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{s+1}) - W_{ji}(r_s)) \middle| \mathcal{F}_{r_s}^k \right] \right] \\
&+ \lim_{n \rightarrow \infty} \sum_{\substack{s,m=0 \\ s>m}}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_m)) (W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{m+1}) - W_{ji}(r_m)) \middle| \mathcal{F}_{r_m}^k \right] \right] \\
&+ \lim_{n \rightarrow \infty} \sum_{\substack{s,m=0, \\ m>s}}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_m)) (W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{m+1}) - W_{ji}(r_m)) \middle| \mathcal{F}_{r_s}^k \right] \right] \\
&= \lim_{n \rightarrow \infty} \sum_{s=0}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_s(W_{ji}(r_s)) \middle| \mathcal{F}_{r_s}^k \right] \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{s+1}) - W_{ji}(r_s)) \middle| \mathcal{F}_{r_s}^k \right] \right] \\
&+ \lim_{n \rightarrow \infty} \sum_{\substack{s,m=0 \\ s>m}}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_m)) \middle| \mathcal{F}_{r_m}^k \right] \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{m+1}) - W_{ji}(r_m)) \middle| \mathcal{F}_{r_m}^k \right] \right] \\
&+ \lim_{n \rightarrow \infty} \sum_{\substack{s,m=0, \\ m>s}}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_m)) \middle| \mathcal{F}_{r_s}^k \right] \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{m+1}) - W_{ji}(r_m)) \middle| \mathcal{F}_{r_s}^k \right] \right]
\end{aligned}$$

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} \sum_{s=0}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_s)) | \mathcal{F}_{r_s}^k \right] \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{s+1}) - W_{ji}(r_s)) \right] \right] \\
&+ \lim_{n \rightarrow \infty} \sum_{\substack{s,m=0 \\ s>m}}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_m)) | \mathcal{F}_{r_m}^k \right] \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{m+1}) - W_{ji}(r_m)) \right] \right] \\
&+ \lim_{n \rightarrow \infty} \sum_{\substack{s,m=0, \\ m>s}}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_m)) | \mathcal{F}_{r_s}^k \right] \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{m+1}) - W_{ji}(r_m)) \right] \right] \\
&= \lim_{n \rightarrow \infty} \sum_{s=0}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_s)) | \mathcal{F}_{r_s}^k \right] \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{s+1}) - W_{ji}(r_s)) \right] \right]
\end{aligned}$$

which follows from both increments of W_{ki} and W_{ji} being independent from the sigma-algebra that we condition on and due to their independence on the different sub-intervals of the partition (demonstrated above). Now:

$$\begin{aligned}
&\mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ji}(r_{s+1}) - W_{ji}(r_s)) \right] \\
&= \mathbb{E} \left[\left(W_{ki}(r_{s+1}) - W_{ki}(r_s) \right) \left(\lambda_i W_{ki}(r_{s+1}) + \sqrt{1 - \lambda_i^2} W_{ki}^\perp(r_{s+1}) \right. \right. \\
&\quad \left. \left. - \lambda_i W_{ki}(r_s) - \sqrt{1 - \lambda_i^2} W_{ki}^\perp(r_s) \right) \right] \\
&= \lambda_i \mathbb{E} \left[W_{ki}^2(r_{s+1}) \right] - \lambda_i \mathbb{E} \left[W_{ki}(r_{s+1}) W_{ki}(r_s) \right] \\
&\quad - \lambda_i \mathbb{E} \left[W_{ki}(r_s) W_{ki}(r_{s+1}) \right] + \lambda_i \mathbb{E} \left[W_{ki}^2(r_s) \right] \\
&= \lambda_i [r_{s+1} - r_s] - \lambda_i r_s + \lambda_i r_s = \lambda_i [r_{s+1} - r_s]
\end{aligned}$$

Therefore, the covariance term collapses to:

$$\begin{aligned}
&\mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ji}(r) \int_0^1 \tilde{J}_{jic}(r) dW_{ki}(r) \right] \\
&= \lim_{n \rightarrow \infty} \lambda_i \sum_{s=0}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_s)) | \mathcal{F}_{r_s}^k \right] \right] (r_{s+1} - r_s) \\
&= \lim_{n \rightarrow \infty} \lambda_i \sum_{s=0}^{n-1} \mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_s)) \right] (r_{s+1} - r_s) \\
&= \lambda_i \int_0^1 \mathbb{E} \left[\tilde{J}_{kic}(r) \tilde{J}_{jic}(r) \right] dr = \lambda_i^2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr
\end{aligned}$$

where the convergence happens in the mean square sense and we come to the last equality using the original result by [Guillaume \(2017\)](#) to find the covariance between the two O-U processes. \square

iii)

$$\begin{aligned}
& \mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \int_0^1 \tilde{J}_{jic}(r) dW_{ki}(r) \right] \\
&= \lim_{n \rightarrow \infty} \sum_{s=0}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_s(W_{ji}(r_s)) (W_{ki}(r_{s+1}) - W_{ki}(r_s))^2 \mid \mathcal{F}_{r_s}^k \right] \right] \\
&+ \lim_{n \rightarrow \infty} \sum_{\substack{s,m=0 \\ s>m}}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_m)) (W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ki}(r_{m+1}) - W_{ki}(r_m)) \mid \mathcal{F}_{r_m}^k \right] \right] \\
&+ \lim_{n \rightarrow \infty} \sum_{\substack{s,m=0 \\ m>s}}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_m)) (W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ki}(r_{m+1}) - W_{ki}(r_m)) \mid \mathcal{F}_{r_s}^k \right] \right] \\
&= \lim_{n \rightarrow \infty} \sum_{s=0}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_s(W_{ji}(r_s)) \mid \mathcal{F}_{r_s}^k \right] \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s))^2 \right] \right] \\
&+ \lim_{n \rightarrow \infty} \sum_{\substack{s,m=0 \\ s>m}}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_m)) \mid \mathcal{F}_{r_m}^k \right] \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ki}(r_{m+1}) - W_{ki}(r_m)) \right] \right] \\
&+ \lim_{n \rightarrow \infty} \sum_{\substack{s,m=0 \\ m>s}}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f_s(W_{ki}(r_s)) f_m(W_{ji}(r_m)) \mid \mathcal{F}_{r_s}^k \right] \mathbb{E} \left[(W_{ki}(r_{s+1}) - W_{ki}(r_s)) (W_{ki}(r_{m+1}) - W_{ki}(r_m)) \right] \right] \\
&= \lim_{n \rightarrow \infty} \sum_{s=0}^{n-1} \mathbb{E} \left[f_s(W_{ki}(r_s)) f_s(W_{ji}(r_s)) \right] (r_{s+1} - r_s) = \int_0^1 \mathbb{E} \left[\tilde{J}_{kic}(r) \tilde{J}_{jic}(r) \right] dr \\
&= \lambda_i \int_0^1 \int_0^r e^{2(r-s)c} ds dr
\end{aligned}$$

where the convergence happens in the mean square sense. We use the the independence of the Wiener increments (for the same process) and we come to the last equality using the original result by [Guillaume \(2017\)](#) to find the covariance between the two O-U processes. \square

iv) As in i), we can conveniently use the decomposition directly in the integral:

$$\begin{aligned}
& \mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \int_0^1 \tilde{J}_{kic}(r) dW_{ji}(r) \right] \\
&= \lambda_i \mathbb{E} \left[\left(\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \right)^2 \right] + \sqrt{1 - \lambda_i^2} \mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \int_0^1 \tilde{J}_{kic}(r) dW_{ki}^\perp(r) \right] \\
&= \lambda_i \int_0^1 \int_0^r e^{2(r-s)c} ds dr
\end{aligned}$$

because the second term is zero due to independence of W_{ki} and W_{ki}^\perp which can be shown in the same way as in the second term in i). The rest follows the usual Itô Isom-

etry. □

v) By applying the result of [Guillaume \(2017\)](#) iteratively:

$$\begin{aligned} \mathbb{E} \left[\int_0^1 \tilde{J}_{kic}(r) dW_{ki}(r) \int_0^1 \tilde{J}_{jic}(r) dW_{ji}(r) \right] &= \lambda_i \int_0^1 \mathbb{E} \left[\tilde{J}_{kic}(r) \tilde{J}_{jic}(r) \right] dr \\ &= \lambda_i^2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr \end{aligned}$$

Note that re-writing in terms of the incremental sum would result in the same expression as in ii). □

Hence, if we return the scaling by the long-run variances and use $\lambda_i = \frac{\Gamma_{i,21}}{\sqrt{\Gamma_{i,11}\Gamma_{i,22}}}$, we can describe the matrices in (3.9) - (3.12) as:

$$\mathcal{M}_i = \mathbb{E}[M_i M_i^\top]$$

$$\begin{aligned} &= \begin{bmatrix} \Gamma_{i,11}^2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 & \lambda_i \Gamma_{i,11}^{\frac{3}{2}} \sqrt{\Gamma_{i,22}} \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 \\ 0 & \frac{\Gamma_{i,11}\Gamma_{i,22}}{2b} & 0 & 0 \\ \lambda_i \Gamma_{i,11}^{\frac{3}{2}} \sqrt{\Gamma_{i,22}} \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 & \Gamma_{i,11}\Gamma_{i,22} \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 \\ 0 & 0 & 0 & \frac{\Gamma_{i,22}^2}{2b} \end{bmatrix} \\ &= \begin{bmatrix} \Gamma_{i,11}^2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 & \Gamma_{i,21}\Gamma_{i,11} \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 \\ 0 & \frac{\Gamma_{i,11}\Gamma_{i,22}}{2b} & 0 & 0 \\ \Gamma_{i,21}\Gamma_{i,11} \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 & \Gamma_{i,11}\Gamma_{i,22} \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 \\ 0 & 0 & 0 & \frac{\Gamma_{i,22}^2}{2b} \end{bmatrix} \end{aligned}$$

$$\Xi_i = \mathbb{E}[\Pi_i \Pi_i^\top]$$

$$= \begin{bmatrix} \Gamma_{i,11}^2 & \lambda_i \Gamma_{i,11}^{\frac{3}{2}} \sqrt{\Gamma_{i,22}} & \lambda_i \Gamma_{i,11}^{\frac{3}{2}} \sqrt{\Gamma_{i,22}} & \lambda_i^2 \Gamma_{i,11} \Gamma_{i,22} \\ \lambda_i \Gamma_{i,11}^{\frac{3}{2}} \sqrt{\Gamma_{i,22}} & \Gamma_{i,11} \Gamma_{i,22} & \lambda_i^2 \Gamma_{i,11} \Gamma_{i,22} & \lambda_i \sqrt{\Gamma_{i,11}} \Gamma_{i,22}^{\frac{3}{2}} \\ \lambda_i \Gamma_{i,11}^{\frac{3}{2}} \sqrt{\Gamma_{i,22}} & \lambda_i^2 \Gamma_{i,11} \Gamma_{i,22} & \Gamma_{i,11} \Gamma_{i,22} & \lambda_i \sqrt{\Gamma_{i,11}} \Gamma_{i,22}^{\frac{3}{2}} \\ \lambda_i^2 \Gamma_{i,11} \Gamma_{i,22} & \lambda_i \sqrt{\Gamma_{i,11}} \Gamma_{i,22}^{\frac{3}{2}} & \lambda_i \sqrt{\Gamma_{i,11}} \Gamma_{i,22}^{\frac{3}{2}} & \Gamma_{i,22}^2 \end{bmatrix} \int_0^1 \int_0^r e^{2(r-s)c} ds dr$$

$$= \begin{bmatrix} \Gamma_{i,11}^2 & \Gamma_{i,21} \Gamma_{i,11} & \Gamma_{i,21} \Gamma_{i,11} & \Gamma_{i,21}^2 \\ \Gamma_{i,21} \Gamma_{i,11} & \Gamma_{i,11} \Gamma_{i,22} & \Gamma_{i,21}^2 & \Gamma_{i,21} \Gamma_{i,22} \\ \Gamma_{i,21} \Gamma_{i,11} & \Gamma_{i,21}^2 & \Gamma_{i,11} \Gamma_{i,22} & \Gamma_{i,21} \Gamma_{i,22} \\ \Gamma_{i,21}^2 & \Gamma_{i,21} \Gamma_{i,22} & \Gamma_{i,21} \Gamma_{i,22} & \Gamma_{i,22}^2 \end{bmatrix} \int_0^1 \int_0^r e^{2(r-s)c} ds dr$$

$$\mathcal{B}_i = \mathbb{E}[\mathbf{B}_i] = \begin{bmatrix} \Gamma_{i,11} \int_0^1 \int_0^r e^{2(r-s)c} & 0 \\ 0 & \Gamma_{i,22} \end{bmatrix}, \quad \Theta_i = \mathbb{E}[\mathbf{C}_i] = \begin{bmatrix} \Gamma_{i,11} & \Gamma_{i,21} \\ \Gamma_{i,21} & \Gamma_{i,22} \end{bmatrix} \int_0^1 \int_0^r e^{2(r-s)c} ds dr$$

Lemma A

- a) $\frac{1}{T} \sum_{t=2}^T X_{1it-1} u_{1it} \Rightarrow C_{1i}^2(1) \int_0^1 J_{1ic}(r) dB_{1i}(r) + \Lambda_{i,11}$ for $i = 1, \dots, N$
- b) $\frac{1}{T} \sum_{t=2}^n X_{1it-1} u_{2it} \Rightarrow C_{1i}(1) C_{2i}(1) \int_0^1 J_{1ic}(r) dB_{2i}(r) + \Lambda_{i,21}$ for $i = 1, \dots, N$
- c) $\frac{1}{T} \sum_{t=2}^n X_{1it-1} u_{2it} \Rightarrow C_{1i}(1) C_{2i}(1) \int_0^1 J_{2ic}(r) dB_{1i}(r) + \Lambda_{i,12}$ for $i = 1, \dots, N$
- d) $\frac{1}{T} \sum_{t=2}^n X_{1it-1} u_{2it} \Rightarrow C_{2i}^2(1) \int_0^1 J_{2ic}(r) dB_{2i}(r) + \Lambda_{i,22}$ for $i = 1, \dots, N$

where $\Lambda_{i,jj}$ for $j = 1, 2$ come from Λ_i .

Proof of Lemma A

First, note that (2.9) can be expressed as:

$$\begin{aligned}
\Lambda_i &= \sum_{h=1}^{\infty} \mathbb{E}[u_{it} u_{it-h}^{\top}] \\
&= \sum_{h=1}^{\infty} \mathbb{E} \left[\sum_{j=0}^{\infty} \mathbf{C}_{ij} \epsilon_{it-j} \sum_{j=0}^{\infty} \epsilon_{it-h-j}^{\top} \mathbf{C}_{ij}^{\top} \right] \\
&= \sum_{j=0}^{\infty} \left(\sum_{h=1}^{\infty} \mathbf{C}_{ij+h} \right) \Sigma_{i\epsilon} \mathbf{C}_{ij}^{\top} \\
&= \sum_{j=0}^{\infty} \tilde{\mathbf{C}}_{ij} \Sigma_{i\epsilon} \mathbf{C}_{ij}^{\top} = \mathbb{E}[\tilde{\epsilon}_{it} u_{it}^{\top}] = \begin{bmatrix} \Lambda_{i,11} & \Lambda_{i,21} \\ \Lambda_{i,21} & \Lambda_{i,22} \end{bmatrix}
\end{aligned}$$

For both parts, we will invoke an approximation argument similar to the one in [Westerlund and Smeekes \(2018\)](#).

a)

$$\frac{1}{T} \sum_{t=2}^T X_{1it-1} u_{1it} = \frac{1}{T} \sum_{t=2}^T X_{1it-1} \mathbf{C}_{1i}(1) \epsilon_{1it} - \frac{1}{T} \sum_{t=2}^T X_{1it-1} \Delta \tilde{\epsilon}_{1it}$$

Now, for convenience define $\tilde{X}_{1it} = \sum_{j=1}^t \rho_T^{t-j} \epsilon_{1it} + o_p(\sqrt{T})$. The first term can be written in the following way:

i)

$$\frac{1}{T} \sum_{t=2}^T X_{1it-1} \mathbf{C}_{1i}(1) \epsilon_{1it} = \mathbf{C}_{1i}^2(1) \frac{1}{T} \sum_{t=2}^T \tilde{X}_{1it-1} \epsilon_{1it} + \frac{1}{T} \sum_{t=2}^T (X_{1it-1} - \mathbf{C}_{1i}(1) \tilde{X}_{1it-1}) \mathbf{C}_{1i}(1) \epsilon_{1it}$$

Note that using the BN decomposition in connection to $\Delta \rho_T^{t-j+1} = \frac{c}{T} \rho_T^{t-j}$, we can express $X_{1it} = \mathbf{C}_{1i}(1) \tilde{X}_{1it} - \sum_{j=1}^t \rho_T^{t-j} \Delta \tilde{\epsilon}_{1ij} = \mathbf{C}_{1i}(1) \tilde{X}_{1it} - \tilde{\epsilon}_{1it} + \frac{c}{T} \sum_{j=1}^t \rho_T^{t-j} \tilde{\epsilon}_{1ij-1}$. Taking the first lag and inserting this into the expression above we get:

$$\begin{aligned}
\frac{1}{T} \sum_{t=2}^T X_{1it-1} \mathbf{C}_{1i}(1) \epsilon_{1it} &= \mathbf{C}_{1i}^2(1) \frac{1}{T} \sum_{t=2}^T \tilde{X}_{1it-1} \epsilon_{1it} - \mathbf{C}_{1i}(1) \frac{1}{T} \sum_{t=2}^T \tilde{\epsilon}_{1it-1} \epsilon_{1it} \\
&+ \mathbf{C}_{1i}(1) \frac{c}{T^2} \sum_{t=2}^T \sum_{j=1}^{t-1} \rho_T^{t-j} \tilde{\epsilon}_{1ij-1} \epsilon_{1it}
\end{aligned}$$

Now, define \mathcal{F}_{it-1} as the smallest sigma-algebra containing all the information on the vector ϵ_{it} up until the period $t-1$ (similarly to Proposition 3 (g)). Also, $\mathbb{E}_{it-1}[\cdot]$ is the expectation conditional on such sigma-algebra. Clearly, $\mathbb{E}[\epsilon_{1it-1}\epsilon_{1it}] = \mathbb{E}[\tilde{\epsilon}_{1it-1}\mathbb{E}_{it-1}[\tilde{\epsilon}_{1it}]] = 0$, because IID errors imply Martingale Difference Sequence (MDS). Further:

$$\begin{aligned} \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=2}^T\tilde{\epsilon}_{1it-1}\epsilon_{1it}\right)^2\right] &= \frac{1}{T^2}\sum_{t=2}^T\mathbb{E}[\tilde{\epsilon}_{1it-1}^2\mathbb{E}_{it-1}[\epsilon_{1it}^2]] + \frac{1}{T^2}\sum_{\substack{s,t=2, \\ t>s}}^T\mathbb{E}[\tilde{\epsilon}_{1it-1}\tilde{\epsilon}_{1is-1}\epsilon_{1is}\mathbb{E}_{it-1}[\epsilon_{1it}]] \\ &+ \frac{1}{T^2}\sum_{\substack{s,t=2, \\ s>t}}^T\mathbb{E}[\tilde{\epsilon}_{1it-1}\tilde{\epsilon}_{1is-1}\epsilon_{1it}\mathbb{E}_{is-1}[\epsilon_{1is}]] = \frac{1}{T^2}\sum_{t=2}^T\sigma_{i.11}^2\mathbb{E}[\tilde{\epsilon}_{1it-1}^2] = O\left(\frac{1}{T}\right) \end{aligned}$$

as $T \rightarrow \infty$, because $\mathbb{E}[\tilde{\epsilon}_{1it-1}^2]$ is finite by Assumption 2. This implies that $\frac{1}{T}\sum_{t=2}^T\tilde{\epsilon}_{1it-1}\epsilon_{1it} \xrightarrow{P} 0$. Also:

$$\frac{1}{T^2}\sum_{t=2}^T\sum_{j=1}^{t-1}\rho_T^{t-j}\tilde{\epsilon}_{1ij-1}\epsilon_{1it} = O_p(1/T)$$

This is because $\sum_{j=1}^{t-1}\rho_T^{t-j}\tilde{\epsilon}_{1ij-1}$ asymptotically behaves just like X_{1it-1} and converges to O-U process by Proposition 2 (a) if normalized by \sqrt{T} , because $\tilde{\epsilon}_{1ij-1}$ is a linear process. Also, it is independent from ϵ_{1it} for all $j = 1, \dots, t-1$. Lastly, we are left with the term which converges according to the theory in Phillips (1987):

$$C_{1i}^2(1)\frac{1}{T}\sum_{t=2}^T\tilde{X}_{1it-1}\epsilon_{1it} \Rightarrow C_{1i}^2(1)\int_0^1J_{1ic}(r)dB_{1i}(r)$$

for $i = 1, \dots, N$.

ii) Looking at the second term:

$$\begin{aligned} \frac{1}{T}\sum_{t=2}^TX_{1it-1}\Delta\tilde{\epsilon}_{1it} &= -\frac{1}{T}\sum_{t=2}^T\Delta X_{1it}\tilde{\epsilon}_{1it} + \frac{1}{T}\sum_{t=2}^T(X_{1it}\tilde{\epsilon}_{1it} - X_{1it-1}\tilde{\epsilon}_{1it-1}) \\ &= -\frac{1}{T}\sum_{t=2}^T\Delta X_{1it}\tilde{\epsilon}_{1it} + \frac{1}{T}X_{1iT}\tilde{\epsilon}_{1iT} = -\frac{1}{T}\sum_{t=2}^T\Delta X_{1it}\tilde{\epsilon}_{1it} + O_p\left(\frac{1}{T}\right) \end{aligned}$$

because $X_{1iT} = O_p(\sqrt{T})$ and $\mathbb{E}\left[\left|\frac{\tilde{\epsilon}_{1iT}}{\sqrt{T}}\right|\right] = \frac{1}{\sqrt{T}}\mathbb{E}[|\tilde{\epsilon}_{1iT}|] = o(1)$. Further, using $\rho_T = 1 + \frac{c}{T}$ and $\Delta X_{1it} = (\rho_T - 1)X_{1it-1} + u_{1it}$ we obtain:

$$-\frac{1}{T}\sum_{t=2}^T\Delta X_{1it}\tilde{\epsilon}_{1it} = -\frac{1}{T}\sum_{t=2}^Tu_{1it}\tilde{\epsilon}_{1it} - \frac{c}{T^2}\sum_{t=2}^TX_{1it-1}\tilde{\epsilon}_{1it} = -\frac{1}{T}\sum_{t=2}^Tu_{1it}\tilde{\epsilon}_{1it} + O_p\left(\frac{1}{T}\right)$$

Here, $\frac{c}{T^2} \sum_{t=2}^T X_{1it-1} \tilde{\epsilon}_{1it} = \frac{c}{T^2} \sum_{t=2}^T X_{1it-1} K_{1i}(1) \epsilon_{1it} + \frac{c}{T^2} \sum_{t=2}^T X_{1it-1} \Delta \tilde{\epsilon}_{1it} = O_p\left(\frac{1}{T}\right)$ as $\tilde{\epsilon}_{1it}$ is a linear process, so we can apply the BN decomposition. The second term includes the difference, thus it is asymptotically negligible. Finally, $\frac{1}{T} \sum_{t=2}^T u_{1it} \tilde{\epsilon}_{1it} \xrightarrow{P} \Lambda_{i,11}$, where $\Lambda_i = \mathbb{E}[\tilde{\epsilon}_{it} u_{it}^\top]$ and $\Lambda_{i,11}$ is its first diagonal element for $i = 1, \dots, N$ in this notation. \square

b) Similarly to part a) and using the same expression for X_{1it} , we analyze two terms after using the BN decomposition:

$$\frac{1}{T} \sum_{t=2}^T X_{1it-1} u_{2it} = \frac{1}{T} \sum_{t=2}^T X_{1it-1} C_{2i}(1) \epsilon_{2it} - \frac{1}{T} \sum_{t=2}^T X_{1it-1} \Delta \tilde{\epsilon}_{2it}$$

i)

$$\begin{aligned} \frac{1}{T} \sum_{t=2}^T X_{1it-1} C_{2i}(1) \epsilon_{2it} &= C_{1i}(1) C_{2i}(1) \frac{1}{T} \sum_{t=2}^T \tilde{X}_{1it-1} \epsilon_{2it} + \frac{1}{T} \sum_{t=2}^T (X_{1it-1} - C_{1i}(1) \tilde{X}_{1it-1}) C_{2i}(1) \epsilon_{2it} \\ &= C_{1i}(1) C_{2i}(1) \frac{1}{T} \sum_{t=2}^T \tilde{X}_{1it-1} \epsilon_{2it} + \frac{1}{T} \sum_{t=2}^T \left(-\tilde{\epsilon}_{1it-1} + \frac{c}{T} \sum_{j=1}^{t-1} \rho_T^{t-j} \tilde{\epsilon}_{1ij-1} \right) C_{2i}(1) \epsilon_{2it} \end{aligned}$$

Defining the same smallest sigma-algebra \mathcal{F}_{it-1} with information on vector ϵ_{it} up to $t-1$, we show that $\mathbb{E}[\tilde{\epsilon}_{1it-1} \epsilon_{2it}] = \mathbb{E}[\tilde{\epsilon}_{1it-1} \mathbb{E}_{it-1}[\epsilon_{2it}]] = 0$ and:

$$\begin{aligned} \mathbb{E} \left[\left(\frac{1}{T} \sum_{t=2}^T \tilde{\epsilon}_{1it-1} \epsilon_{2it} \right)^2 \right] &= \frac{1}{T^2} \sum_{t=2}^T \mathbb{E}[\tilde{\epsilon}_{1it-1}^2 \mathbb{E}_{it-1}[\epsilon_{2it}^2]] + \frac{1}{T^2} \sum_{\substack{s,t=2, \\ t>s}}^T \mathbb{E}[\tilde{\epsilon}_{1it-1} \tilde{\epsilon}_{1is-1} \epsilon_{2is} \mathbb{E}_{it-1}[\epsilon_{2it}]] \\ &+ \frac{1}{T^2} \sum_{\substack{s,t=2, \\ s>t}}^T \mathbb{E}[\tilde{\epsilon}_{1it-1} \tilde{\epsilon}_{1is-1} \epsilon_{2it} \mathbb{E}_{is-1}[\epsilon_{2is}]] = \frac{1}{T^2} \sum_{t=2}^T \sigma_{i,22}^2 \mathbb{E}[\tilde{\epsilon}_{1it-1}^2] = O\left(\frac{1}{T}\right) \end{aligned}$$

as $T \rightarrow \infty$, which implies that $\frac{1}{T} \sum_{t=2}^T \tilde{\epsilon}_{1it-1} \epsilon_{2it} \xrightarrow{P} 0$. By the same argument of asymptotically equivalent behavior to X_{1it-1} and independence from ϵ_{2it} , we have:

$$\frac{1}{T^2} \sum_{t=2}^T \sum_{j=1}^{t-1} \rho_T^{t-j} \tilde{\epsilon}_{1ij-1} \epsilon_{2it} = O_p\left(\frac{1}{T}\right)$$

as $T \rightarrow \infty$.

Again, the first term converges according to the functional theory in [Phillips \(1987\)](#):

$$C_{1i}(1) C_{2i}(1) \frac{1}{T} \sum_{t=2}^n \tilde{X}_{1it-1} \epsilon_{2it} \Rightarrow C_{1i}(1) C_{2i}(1) \int_0^1 J_{1ic}(r) dB_{2i}(r)$$

ii) Asymptotic behavior of the second term is similar:

$$\begin{aligned}
\frac{1}{T} \sum_{t=2}^T X_{1it-1} \Delta \tilde{\epsilon}_{2it} &= -\frac{1}{T} \sum_{t=2}^T \Delta X_{1it} \tilde{\epsilon}_{2it} + \frac{1}{T} \sum_{t=2}^n (X_{1it} \tilde{\epsilon}_{2it} - X_{1it-1} \tilde{\epsilon}_{2it-1}) \\
&= -\frac{1}{T} \sum_{t=2}^T \Delta X_{1it} \tilde{\epsilon}_{2it} + \frac{1}{T} X_{1iT} \tilde{\epsilon}_{2iT} = -\frac{1}{T} \sum_{t=2}^T \Delta X_{1it} \tilde{\epsilon}_{2it} + O_p\left(\frac{1}{T}\right) \\
&= -\frac{1}{T} \sum_{t=2}^T u_{1it} \tilde{\epsilon}_{2it} - \frac{c}{T^2} \sum_{t=2}^T X_{1it-1} \tilde{\epsilon}_{2it} + O_p\left(\frac{1}{T}\right) = -\frac{1}{T} \sum_{t=2}^n u_{1it} \tilde{\epsilon}_{2it} + O_p\left(\frac{1}{T}\right)
\end{aligned}$$

where $\frac{c}{T^2} \sum_{t=2}^T X_{1it-1} \tilde{\epsilon}_{2it} = \frac{c}{T^2} \sum_{t=2}^T X_{1it-1} K_{1i}(1) \epsilon_{1it} + \frac{c}{T^2} \sum_{t=2}^T X_{1it-1} \Delta \tilde{\epsilon}_{1it} = O_p\left(\frac{1}{T}\right)$. Also, $\frac{1}{T} \sum_{t=2}^T u_{1it} \tilde{\epsilon}_{2it} \xrightarrow{P} \Lambda_{i,21}$, where $\Lambda_i = \mathbb{E}[\tilde{\epsilon}_{it} u_{it}^T]$ and Λ_{21} is its first element in the second row for $i = 1, \dots, N$ in this notation. \square

*The proofs for the cases of $\frac{1}{T} \sum_{t=2}^T X_{2it-1} u_{1it}$ and $\frac{1}{T} \sum_{t=2}^T X_{2it-1} u_{1it}$ are analogous and the bias terms will be $\Lambda_{i,12}$ and $\Lambda_{i,22}$, respectively.

Lemma B

- a) $\frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} \Delta \tilde{\epsilon}_{ij} = o_p(1)$ as $T \rightarrow \infty$
- b) $\frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-(T-j)-1} \Delta \tilde{\epsilon}_{ij} = o_p(1)$ as $T \rightarrow \infty$

Here, $\tilde{\epsilon}_{ij} \in \mathbb{R}^2$ and covers each individual i .

Proof of Lemma B

a) At first, we re-write the sum to eliminate the difference and use the definition of θ_T :

$$\begin{aligned}
\frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} \Delta \tilde{\epsilon}_{ij} &= \frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} \tilde{\epsilon}_{ij} - \frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} \tilde{\epsilon}_{ij-1} \\
&= \frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} \tilde{\epsilon}_{ij} - \left(\frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-1-j} \tilde{\epsilon}_{ij} + o_p(1) \right) \\
&= \frac{1}{\sqrt{k_T}} (1 - \theta_T^{-1}) \sum_{j=1}^T \theta_T^{-j} \tilde{\epsilon}_{ij} + o_p(1) \\
&= -\frac{1}{\sqrt{k_T}} \theta_T^{-1} \left(-\frac{b}{k_T} \right) \sum_{j=1}^T \theta_T^{-j} \tilde{\epsilon}_{ij} + o_p(1)
\end{aligned}$$

The first term converges to 0 in $\mathcal{L}1$ norm:

$$\begin{aligned}
\mathbb{E} \left[\left\| \frac{1}{\sqrt{k_T}} \theta_T^{-1} \left(\frac{b}{k_T} \right) \sum_{j=1}^T \theta_T^{-j} \tilde{\epsilon}_{ij} \right\| \right] &\leq \frac{1}{\sqrt{k_T}} \theta_T^{-1} \left(\frac{b}{k_T} \right) \sum_{j=1}^T \theta_T^{-j} \mathbb{E} [\|\tilde{\epsilon}_{ij}\|] \\
&= \mathbb{E} [\|\tilde{\epsilon}_{i1}\|] \frac{1}{\sqrt{k_T}} \theta_T^{-1} \left(\frac{b}{k_T} \right) \sum_{j=1}^T \theta_T^{-j} \\
&= \mathbb{E} [\|\tilde{\epsilon}_{i1}\|] \frac{1}{\sqrt{k_T}} \theta_T^{-1} \left(\frac{b}{k_T} \right) \frac{1 - \theta_T^{-T}}{\theta_T - 1} \\
&= \mathbb{E} [\|\tilde{\epsilon}_{i1}\|] \theta_T^{-1} \frac{1}{\sqrt{k_T}} (1 + o(1)) = o(1)
\end{aligned}$$

as $T \rightarrow \infty$ due to covariance stationarity and $\mathbb{E} [\|\tilde{\epsilon}_{i1}\|] < \infty$. The same result holds in scalar case replacing vector norm with an absolute value. \square

b) Similarly to part a):

$$\begin{aligned}
\frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-(T-j)-1} \Delta \tilde{\epsilon}_{ij} &= \frac{1}{\sqrt{k_T}} \theta_T^{-T} \sum_{j=1}^T \theta_T^{j-1} \tilde{\epsilon}_{ij} - \left(\frac{1}{\sqrt{k_T}} \theta_T^{-T-1} \sum_{j=1}^T \theta_T^{j-1} \tilde{\epsilon}_{ij} + o_p(1) \right) \\
&= \frac{1}{\sqrt{k_T}} (1 - \theta_T^{-1}) \theta_T^{-T} \sum_{j=1}^T \theta_T^{j-1} \tilde{\epsilon}_{ij} + o_p(1) \\
&= -\frac{1}{\sqrt{k_T}} \theta_T^{-1} (1 - \theta_T) \theta_T^{-T} \sum_{j=1}^T \theta_T^{j-1} \tilde{\epsilon}_{ij} + o_p(1) \\
&= \frac{1}{\sqrt{k_T}} \theta_T^{-2} \frac{b}{k_T} \theta_T^{-T} \sum_{j=1}^T \theta_T^j \tilde{\epsilon}_{ij} + o_p(1)
\end{aligned}$$

Again, the first term converges to 0 in $\mathcal{L}1$ norm:

$$\begin{aligned}
\mathbb{E} \left[\left\| \frac{1}{\sqrt{k_T}} \theta_T^{-2} \frac{b}{k_T} \theta_T^{-T} \sum_{j=1}^T \theta_T^j \tilde{\epsilon}_{ij} \right\| \right] &\leq \frac{1}{\sqrt{k_T}} \theta_T^{-2} \frac{b}{k_T} \theta_T^{-T} \sum_{j=1}^T \theta_T^j \mathbb{E} [\|\tilde{\epsilon}_{ij}\|] \\
&= \mathbb{E} [\|\tilde{\epsilon}_{i1}\|] \frac{1}{\sqrt{k_T}} \theta_T^{-1} \frac{b}{k_T} \frac{1 - \theta_T^{-T}}{\theta_T - 1} \\
&= \mathbb{E} [\|\tilde{\epsilon}_{i1}\|] \frac{1}{\sqrt{k_T}} \theta_T^{-1} (1 + o(1)) = o(1)
\end{aligned}$$

because of covariance stationarity and $\mathbb{E} [\|\tilde{\epsilon}_{i1}\|] < \infty$. The same result holds in a scalar case replacing vector norm with an absolute value. \square

Lemma C

- a) $\frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t}^T \theta_T^{t-j-1} u_{2ij} u_{2it} = o_p(1)$ as $T \rightarrow \infty$
b) $\frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t}^T \theta_T^{t-j-1} u_{2ij} u_{1it} = o_p(1)$ as $T \rightarrow \infty$

Proof of Lemma C

a) By splitting the double sum

$$\frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t}^T \theta_T^{t-j-1} u_{2ij} u_{2it} = \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \theta_T^{-1} u_{2it}^2 + \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t+1}^T \theta_T^{t-j-1} u_{2ij} u_{2it}$$

Invoking Proposition 1 (b), both terms converge to 0 in $\mathcal{L}1$ norm:

$$\mathbb{E} \left[\left| \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \theta_T^{-1} u_{2it}^2 \right| \right] \leq \frac{\theta_T^{-T-1}}{k_T} \sum_{t=1}^T \mathbb{E}[u_{2it}^2] = \frac{\theta_T^{-T-1}}{k_T} T \Omega_{i.22} = o(1)$$

as $T \rightarrow \infty$ due to covariance stationarity.

$$\begin{aligned} \mathbb{E} \left[\left| \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t+1}^T \theta_T^{t-j-1} u_{2ij} u_{2it} \right| \right] &\leq \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t+1}^T \theta_T^{t-j-1} \mathbb{E}[|u_{2ij}| |u_{2it}|] \\ &\leq \mathbb{E}[|u_{2i1}|^2] \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t+1}^T \theta_T^{t-j-1} \\ &= \mathbb{E}[|u_{2i1}|^2] \frac{\theta_T^{-T-2}}{k_T} \sum_{t=1}^T \frac{1 - \theta_T^{-T}}{1 - \theta_T^{-1}} \\ &= \mathbb{E}[|u_{2i1}|^2] \frac{\theta_T^{-T-1}}{k_T} T \frac{1 - \theta_T^{-T}}{\theta_T - 1} \\ &= \mathbb{E}[|u_{2i1}|^2] \frac{\theta_T^{-T-1}}{b} T (1 + o(1)) = o(1) \end{aligned}$$

as $T \rightarrow \infty$ using the Cauchy-Schwartz inequality, covariance stationarity and the fact that θ_T^{-T-1} decays exponentially. \square

b) By splitting the double sum

$$\frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t}^T \theta_T^{t-j-1} u_{2ij} u_{1it} = \frac{\theta_T^{-T-1}}{k_T} \sum_{t=1}^T u_{2it} u_{1it} + \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t+1}^T \theta_T^{t-j-1} u_{2ij} u_{1it}$$

Again, both terms converge to 0 in $\mathcal{L}1$ norm:

$$\begin{aligned} \mathbb{E} \left[\left| \frac{\theta_T^{-T-1}}{k_T} \sum_{t=1}^T u_{2it} u_{1it} \right| \right] &\leq \frac{\theta_T^{-T-1}}{k_T} \sum_{t=1}^T \mathbb{E}[|u_{2it}| |u_{1it}|] \\ &\leq \sqrt{\mathbb{E}[|u_{2i1}|^2] \mathbb{E}[|u_{1i1}|^2]} \frac{\theta_T^{-T-1}}{k_T} T = o(1) \end{aligned}$$

as $T \rightarrow \infty$ using the Cauchy-Schwartz inequality and covariance stationarity.

$$\begin{aligned} \mathbb{E} \left[\left| \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t+1}^T \theta_T^{t-j-1} u_{2ij} u_{1it} \right| \right] &\leq \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t+1}^T \theta_T^{t-j-1} \mathbb{E}[|u_{2ij}| |u_{1it}|] \\ &\leq \sqrt{\mathbb{E}[|u_{2i1}|^2] \mathbb{E}[|u_{1i1}|^2]} \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \sum_{j=t+1}^T \theta_T^{t-j-1} \\ &= \sqrt{\mathbb{E}[|u_{2i1}|^2] \mathbb{E}[|u_{1i1}|^2]} \frac{\theta_T^{-T-2}}{k_T} \sum_{t=1}^T \frac{1 - \theta_T^{-T}}{1 - \theta_T^{-1}} \\ &= \sqrt{\mathbb{E}[|u_{2i1}|^2] \mathbb{E}[|u_{1i1}|^2]} \frac{\theta_T^{-T-1}}{k_T} T \frac{1 - \theta_T^{-T}}{\theta_T - 1} \\ &= \sqrt{\mathbb{E}[|u_{2i1}|^2] \mathbb{E}[|u_{1i1}|^2]} \frac{\theta_T^{-T-1}}{b} T (1 + o(1)) = o(1) \end{aligned}$$

as $T \rightarrow \infty$ using the Cauchy-Schwartz inequality, covariance stationarity and the fact that θ_T^{-T-1} decays exponentially.

Lemma D

$[X_{1iT} \ Y_{siT}] \xrightarrow{D} [X_{1i}(b) \ Y_{si}(b)]$ **jointly** for $s = 1, 2$, where:

$$\begin{aligned} X_{1iT} &= \frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} u_{1ij} \\ Y_{siT} &= \frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-(T-j)-1} u_{sij} \\ X_{1i}(b) &\stackrel{D}{=} Y_{1i}(b) \stackrel{D}{=} \mathcal{N}\left(0, \frac{1}{2b} C_{1i}^2(1) \sigma_{i.11}^2\right) \\ Y_{2i}(b) &\stackrel{D}{=} \mathcal{N}\left(0, \frac{1}{2b} C_{2i}^2(1) \sigma_{i.22}^2\right) \end{aligned}$$

The joint convergence comes from the asymptotic independence of the Normals.

Proof of Lemma D

To demonstrate joint convergence in distribution, we will invoke Cramer-Wold device (see Proposition 6.3.1 in [Brockwell and Davis \(2013\)](#)) in order to show convergence of a linear combination to a Gaussian random variable. Additionally, if the variance of the result is additively seperable, both variables in the linear combination are independent by the additive property of Gaussian random variables.

Define $C = [m \ h]^\top$ and $B_{si} = [X_{1iT}, Y_{siT}]^\top \in \mathbb{R}^2$. Then the linear combination to invoke Cramer-Wold device is $Z_{siT} = C^\top B_{si}$ for $s = 1, 2$.

i) For $s = 1$, using the BN decomposition and Lemma B, Z_{1iT} can be written as:

$$\begin{aligned}
Z_{1iT} &= \frac{1}{\sqrt{k_T}} \sum_{j=1}^T m \theta_T^{-j} u_{1ij} + \frac{1}{\sqrt{k_T}} \sum_{j=1}^T h \theta_T^{-(T-j)-1} u_{1ij} = \frac{1}{\sqrt{k_T}} \sum_{j=1}^T m \theta_T^{-j} C_{1i}(1) \epsilon_{1ij} \\
&\quad - \frac{m}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} \Delta \tilde{\epsilon}_{1ij} + \frac{1}{\sqrt{k_T}} \sum_{j=1}^T h \theta_T^{-(T-j)-1} C_{1i}(1) \epsilon_{1ij} \\
&\quad - \frac{h}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-(T-j)-1} \Delta \tilde{\epsilon}_{1ij} \\
&= \sum_{j=1}^T \frac{1}{\sqrt{k_T}} \left(m \theta_T^{-j} + h \theta_T^{-(T-j)-1} \right) C_{1i}(1) \epsilon_{1ij} + o_p(1) \\
&= \sum_{j=1}^T \xi_{1iTj} + o_p(1)
\end{aligned}$$

Due to Lemma B. Clearly, $\mathbb{E} \left[\sum_{j=1}^T \xi_{1iTj} \right] = 0$ and the asymptotic variance is the following:

$$\begin{aligned}
\text{Var} \left[\sum_{j=1}^T \xi_{1iTj} \right] &= \sum_{j=1}^T \frac{1}{k_T} (m \theta_T^{-j} + h \theta_T^{-(T-j)-1})^2 C_{1i}^2(1) \sigma_{i.11}^2 \\
&= \left(\frac{1}{k_T} \sum_{j=1}^T m^2 \theta_T^{-2j} + 2 \frac{1}{k_T} T \theta_T^{-T-1} m h + \frac{1}{k_T} \sum_{j=1}^T h^2 \theta_T^{-2(T-j)-2} \right) \Gamma_{i.11} \\
&= \left(\frac{m^2}{k_T} \frac{1 - \theta_T^{-2T}}{\theta_T^2 - 1} + o(1) + \frac{h^2}{k_T} \frac{1 - \theta_T^{-2T}}{\theta_T^2 - 1} \right) \Gamma_{i.11} \\
&\rightarrow \left(\frac{m^2 + h^2}{2b} \right) \Gamma_{i.11}
\end{aligned}$$

as $T \rightarrow \infty$ using the fact that ϵ_{1ij} are IID and Proposition 1 (c).

To demonstrate convergence in distribution to a random Normal variable, we check the Lindeberg condition. For bounding, we will use the fact that $(x + y)^2 \leq 2(x^2 + y^2)$. Fix arbitrary $\eta > 0$ and $K \in (0, \infty)$ to uniformly bound the right hand side of the asymptotic variance equation:

$$\begin{aligned}
& \sum_{j=1}^T \mathbb{E} \left[\left| \zeta_{1iTj} \right|^2 \mathbf{1} \left\{ \left| \zeta_{1iTj} \right| > \eta \right\} \right] \\
&= \frac{1}{k_T} \sum_{j=1}^T \mathbb{E} \left[\left| \left(m\theta_T^{-j} + h\theta_T^{-(T-j)-1} \right) C_{1i}(1) \epsilon_{1ij} \right|^2 \mathbf{1} \left\{ \left| \left(m\theta_T^{-j} + h\theta_T^{-(T-j)-1} \right) C_{1i}(1) \epsilon_{1ij} \right| > \eta \sqrt{k_T} \right\} \right] \\
&= \frac{|C_{1i}(1)|^2}{k_T} \sum_{j=1}^T \left(m\theta_T^{-j} + h\theta_T^{-(T-j)-1} \right)^2 \mathbb{E} \left[\left| \epsilon_{1ij} \right|^2 \mathbf{1} \left\{ \left(m\theta_T^{-j} + h\theta_T^{-(T-j)-1} \right)^2 \left| \epsilon_{1ij} \right|^2 > \eta^2 k_T \right\} \right] \\
&\leq \frac{2|C_{1i}(1)|^2}{k_T} \sum_{j=1}^T \left(m^2 \theta_T^{-2j} + h^2 \theta_T^{-2(T-j)-2} \right) \mathbb{E} \left[\left| \epsilon_{1ij} \right|^2 \mathbf{1} \left\{ \left(m\theta_T^{-j} + h\theta_T^{-(T-j)-1} \right)^2 \left| \epsilon_{1ij} \right|^2 > \eta^2 k_T \right\} \right] \\
&\leq K \max_{1 \leq j \leq T} \mathbb{E} \left[\left| \epsilon_{1ij} \right|^2 \mathbf{1} \left\{ \left(m\theta_T^{-j} + h\theta_T^{-(T-j)-1} \right)^2 \left| \epsilon_{1ij} \right|^2 > \eta^2 k_T \right\} \right] \\
&\leq K \max_{1 \leq j \leq T} \mathbb{E} \left[\left| \epsilon_{1ij} \right|^2 \mathbf{1} \left\{ 2 \left(m^2 \theta_T^{-2j} + h^2 \theta_T^{-2(T-j)-2} \right) \left| \epsilon_{1ij} \right|^2 > \eta^2 k_T \right\} \right] \\
&\leq K \mathbb{E} \left[\left| \epsilon_{1i1} \right|^2 \mathbf{1} \left\{ 2 \left(m^2 + h^2 \right) \left| \epsilon_{1i1} \right|^2 > \eta^2 k_T \right\} \right] \\
&= K \mathbb{E} \left[\left| \epsilon_{1i1} \right|^2 \mathbf{1} \left\{ \left| \epsilon_{1i1} \right|^2 > \frac{\eta^2 k_T}{2(m^2 + h^2)} \right\} \right] = o(1)
\end{aligned}$$

as $T \rightarrow \infty$ since $\left| \epsilon_{1i1} \right|^2$ is integrable. □

ii) For $s = 2$, using the BN decomposition and Lemma B, Z_{2iT} can be written as:

$$\begin{aligned}
Z_{2iT} &= \frac{1}{\sqrt{k_T}} \sum_{j=1}^T m\theta_T^{-j} u_{1ij} + \frac{1}{\sqrt{k_T}} \sum_{j=1}^T h\theta_T^{-(T-j)-1} u_{2ij} = \frac{1}{\sqrt{k_T}} \sum_{j=1}^T m\theta_T^{-j} C_{1i}(1)\epsilon_{1ij} \\
&\quad - \frac{m}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} \Delta \tilde{\epsilon}_{1ij} + \frac{1}{\sqrt{k_T}} \sum_{j=1}^T h\theta_T^{-(T-j)-1} C_{2i}(1)\epsilon_{2ij} \\
&\quad - \frac{h}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-(T-j)-1} \Delta \tilde{\epsilon}_{2ij} \\
&= \sum_{j=1}^T \frac{1}{\sqrt{k_T}} \left(m\theta_T^{-j} C_{1i}(1)\epsilon_{1ij} + h\theta_T^{-(T-j)-1} C_{2i}(1)\epsilon_{2ij} \right) + o_p(1) \\
&= \sum_{j=1}^T \xi_{2iTj} + o_p(1)
\end{aligned}$$

due to Lemma B. Clearly, $\mathbb{E} \left[\sum_{j=1}^T \xi_{2iTj} \right] = 0$. The asymptotic variance is similar when we take the covariance terms into account:

$$\begin{aligned}
\text{Var} \left[\sum_{j=1}^T \xi_{1iTj} \right] &= \frac{1}{k_T} \sum_{j=1}^T m^2 \theta_T^{-2j} C_{1i}^2(1) \sigma_{i,11}^2 + \frac{1}{k_T} \sum_{j=1}^T h^2 \theta_T^{-2(T-j)-2} C_{2i}^2(1) \sigma_{i,22}^2 \\
&\quad + 2C_{1i}(1)C_{2i}(1) \frac{T\theta_T^{-T-1} \sigma_{i,21}}{k_T} \\
&= \frac{1}{k_T} \sum_{j=1}^T m^2 \theta_T^{-2j} \Gamma_{i,11} + \frac{1}{k_T} \sum_{j=1}^T h^2 \theta_T^{-2(T-j)-2} \Gamma_{i,22} + o(1) \\
&= \left(\frac{m^2}{k_T} \frac{1 - \theta_T^{-2T}}{\theta_T^2 - 1} \right) \Gamma_{i,11} + \left(\frac{h^2}{k_T} \frac{1 - \theta_T^{-2T}}{\theta_T^2 - 1} \right) \Gamma_{i,22} + o(1) \rightarrow \frac{m^2 \Gamma_{i,11}}{2b} + \frac{h^2 \Gamma_{i,22}}{2b}
\end{aligned}$$

as $T \rightarrow \infty$ by Proposition 1 (c).

We show the asymptotic normality by the Lindeberg condition. We introduce $M, N \in (0, \infty)$ to uniformly bound the two sequences. Also, because m^2, h^2 are any finite constants and $C_{si}^2(1) < \infty$ for $s = 1, 2$, we can find L such that $L \geq m^2 C_{1i}^2(1)$ and $L \geq h^2 C_{2i}^2(1)$. Then:

$$\begin{aligned}
& \sum_{j=1}^T \mathbb{E} \left[\left| \xi_{2iTj} \right|^2 \mathbf{1} \left\{ \left| \xi_{2iTj} \right| > \eta \right\} \right] \\
&= \frac{1}{k_T} \sum_{j=1}^T \mathbb{E} \left[\left| \left(m\theta_T^{-j} C_{1i}(1) \epsilon_{1ij} + h\theta_T^{-(T-j)-1} C_{2i}(1) \epsilon_{2ij} \right) \right|^2 \mathbf{1} \left\{ \left| \xi_{2iTj} \right| > \eta \sqrt{k_T} \right\} \right] \\
&= \frac{1}{k_T} \sum_{j=1}^T \mathbb{E} \left[\left| \left(m\theta_T^{-j} C_{1i}(1) \epsilon_{1ij} + h\theta_T^{-(T-j)-1} C_{2i}(1) \epsilon_{2ij} \right) \right|^2 \mathbf{1} \left\{ \left| \xi_{2iTj} \right|^2 > \eta^2 k_T \right\} \right] \\
&\leq 2 \frac{C_{1i}^2(1)}{k_T} \sum_{j=1}^T m^2 \theta_T^{-2j} \mathbb{E} \left[\left| \epsilon_{1ij} \right|^2 \mathbf{1} \left\{ \left| \left(m\theta_T^{-j} C_{1i}(1) \epsilon_{1ij} + h\theta_T^{-(T-j)-1} C_{2i}(1) \epsilon_{2ij} \right) \right|^2 > \eta^2 k_T \right\} \right] \\
&\quad + 2 \frac{C_{2i}^2(1)}{k_T} \sum_{j=1}^T h^2 \theta_T^{-2(T-j)-2} \mathbb{E} \left[\left| \epsilon_{2ij} \right|^2 \mathbf{1} \left\{ \left| \left(m\theta_T^{-j} C_{1i}(1) \epsilon_{1ij} + h\theta_T^{-(T-j)-1} C_{2i}(1) \epsilon_{2ij} \right) \right|^2 > \eta^2 k_T \right\} \right] \\
&\leq 2 \frac{C_{1i}^2(1)}{k_T} \sum_{j=1}^T m^2 \theta_T^{-2j} \mathbb{E} \left[\left| \epsilon_{1ij} \right|^2 \mathbf{1} \left\{ 2 \left(m^2 \theta_T^{-2j} C_{1i}^2(1) \epsilon_{1ij}^2 + h^2 \theta_T^{-2(T-j)-2} C_{2i}^2(1) \epsilon_{2ij}^2 \right) > \eta^2 k_T \right\} \right] \\
&\quad + 2 \frac{C_{2i}^2(1)}{k_T} \sum_{j=1}^T h^2 \theta_T^{-2(T-j)-2} \mathbb{E} \left[\left| \epsilon_{2ij} \right|^2 \mathbf{1} \left\{ 2 \left(m^2 \theta_T^{-2j} C_{1i}^2(1) \epsilon_{1ij}^2 + h^2 \theta_T^{-2(T-j)-2} C_{2i}^2(1) \epsilon_{2ij}^2 \right) > \eta^2 k_T \right\} \right] \\
&\leq M \max_{1 \leq j \leq T} \mathbb{E} \left[\left| \epsilon_{1ij} \right|^2 \mathbf{1} \left\{ 2 \left(m^2 \theta_T^{-2j} C_{1i}^2(1) \epsilon_{1ij}^2 + h^2 \theta_T^{-2(T-j)-2} C_{2i}^2(1) \epsilon_{2ij}^2 \right) > \eta^2 k_T \right\} \right] \\
&\quad + N \max_{1 \leq j \leq T} \mathbb{E} \left[\left| \epsilon_{2ij} \right|^2 \mathbf{1} \left\{ 2 \left(m^2 \theta_T^{-2j} C_{1i}^2(1) \epsilon_{1ij}^2 + h^2 \theta_T^{-2(T-j)-2} C_{2i}^2(1) \epsilon_{2ij}^2 \right) > \eta^2 k_T \right\} \right] \\
&\leq M \mathbb{E} \left[\left| \epsilon_{1i1} \right|^2 \mathbf{1} \left\{ 2 \left(m^2 C_{1i}^2(1) \epsilon_{1i1}^2 + h^2 C_{2i}^2(1) \epsilon_{2i1}^2 \right) > \eta^2 k_T \right\} \right] \\
&\quad + N \mathbb{E} \left[\left| \epsilon_{2i1} \right|^2 \mathbf{1} \left\{ 2 \left(m^2 C_{1i}^2(1) \epsilon_{1i1}^2 + h^2 C_{2i}^2(1) \epsilon_{2i1}^2 \right) > \eta^2 k_T \right\} \right] \\
&\leq M \mathbb{E} \left[\left| \epsilon_{1i1} \right|^2 \mathbf{1} \left\{ L \left(\epsilon_{1i1}^2 + \epsilon_{2i1}^2 \right) > \frac{\eta^2 k_T}{2} \right\} \right] + N \mathbb{E} \left[\left| \epsilon_{2i1} \right|^2 \mathbf{1} \left\{ L \left(\epsilon_{1i1}^2 + \epsilon_{2i1}^2 \right) > \frac{\eta^2 k_T}{2} \right\} \right] \\
&= M \mathbb{E} \left[\left| \epsilon_{1i1} \right|^2 \mathbf{1} \left\{ \left| \epsilon_{1i1} \right|^2 + \left| \epsilon_{2i1} \right|^2 > \frac{\eta^2 k_T}{2L} \right\} \right] + N \mathbb{E} \left[\left| \epsilon_{2i1} \right|^2 \mathbf{1} \left\{ \left| \epsilon_{1i1} \right|^2 + \left| \epsilon_{2i1} \right|^2 > \frac{\eta^2 k_T}{2L} \right\} \right] = o(1)
\end{aligned}$$

as $T \rightarrow \infty$ because $\|\epsilon_{i1}\|^2$ is integrable, which means that $\mathbb{E} \left[\|\epsilon_{i1}\|^2 \mathbf{1} \left\{ \|\epsilon_{i1}\|^2 > a \right\} \right] = o(1)$ as $a \rightarrow \infty$. Using the simple Euclidean norm, this implies that:

$$\begin{aligned}
& \mathbb{E} \left[\left(\left| \epsilon_{1i1} \right|^2 + \left| \epsilon_{2i1} \right|^2 \right) \mathbf{1} \left\{ \left| \epsilon_{1i1} \right|^2 + \left| \epsilon_{2i1} \right|^2 > a \right\} \right] = \mathbb{E} \left[\left| \epsilon_{1i1} \right|^2 \mathbf{1} \left\{ \left| \epsilon_{1i1} \right|^2 + \left| \epsilon_{2i1} \right|^2 > a \right\} \right] \\
&\quad + \mathbb{E} \left[\left| \epsilon_{2i1} \right|^2 \mathbf{1} \left\{ \left| \epsilon_{1i1} \right|^2 + \left| \epsilon_{2i1} \right|^2 > a \right\} \right] = o(1)
\end{aligned}$$

as $a \rightarrow \infty$. \square

Lemma E

- a) $\frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T X_{2it-1} u_{2it} \xrightarrow{D} C_{2i}^2(1) X_{2i}(b) Y_{2i}(b)$
b) $\frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T X_{2it-1} u_{1it} \xrightarrow{D} C_{1i}(1) C_{2i}(1) X_{2i}(b) Y_{1i}(b)$
c) $\frac{\theta_T^{-2T}}{k_T^2} \sum_{t=1}^T X_{2it-1}^2 \xrightarrow{D} \frac{1}{2b} C_{2i}^2(1) X_{2i}^2(b)$

as $T \rightarrow \infty$. Here, $X_{2i}(b) \stackrel{D}{=} Y_{2i}(b) = \mathcal{N}(0, \frac{\sigma_{i22}^2}{2b})$ and $Y_{1i}(b) = \mathcal{N}(0, \frac{\sigma_{i11}^2}{2b})$.

Proof of Lemma E

a)

$$\begin{aligned} \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T X_{2it-1} u_{2it} &= \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \left(\sum_{j=1}^{t-1} \theta_T^{t-j-1} u_{2ij} + \theta_T^{t-1} X_{2i0} \right) u_{2it} \\ &= \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \left(\sum_{j=1}^{t-1} \theta_T^{t-j-1} u_{2ij} \right) u_{2it} + \frac{X_{2i0}}{\sqrt{k_T}} \frac{1}{\sqrt{k_T}} \sum_{t=1}^T \theta_T^{-(T-t)-1} u_{2it} \\ &= \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \left(\sum_{j=1}^T \theta_T^{t-j-1} u_{2ij} \right) u_{2it} - \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \left(\sum_{j=t}^T \theta_T^{t-j-1} u_{2ij} \right) u_{2it} + o_p(1) \\ &= \left(\frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} u_{2ij} \right) \left(\frac{1}{\sqrt{k_T}} \sum_{t=1}^T \theta_T^{-(T-t)-1} u_{2it} \right) + o_p(1) \xrightarrow{D} C_{2i}^2(1) X_{2i}(b) Y_{2i}(b) \end{aligned}$$

as $T \rightarrow \infty$ invoking Lemma C (a) and using joint convergence established in Lemma D. \square

b) Analogous to part a):

$$\begin{aligned} \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T X_{2it-1} u_{1it} &= \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \left(\sum_{j=1}^{t-1} \theta_T^{t-j-1} u_{2ij} + \theta_T^{t-1} X_{2i0} \right) u_{1it} \\ &= \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \left(\sum_{j=1}^{t-1} \theta_T^{t-j-1} u_{2ij} \right) u_{1it} + \frac{X_{2i0}}{\sqrt{k_T}} \frac{1}{\sqrt{k_T}} \sum_{t=1}^T \theta_T^{-(T-t)-1} u_{1it} \\ &= \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \left(\sum_{j=1}^T \theta_T^{t-j-1} u_{2ij} \right) u_{1it} - \frac{\theta_T^{-T}}{k_T} \sum_{t=1}^T \left(\sum_{j=t}^T \theta_T^{t-j-1} u_{2ij} \right) u_{1it} + o_p(1) \\ &= \left(\frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} u_{2ij} \right) \left(\frac{1}{\sqrt{k_T}} \sum_{t=1}^T \theta_T^{-(T-t)-1} u_{1it} \right) + o_p(1) \xrightarrow{D} C_{1i}(1) C_{2i}(1) X_{2i}(b) Y_{1i}(b) \end{aligned}$$

as $T \rightarrow \infty$ invoking Lemma C (b) and using joint convergence established in Lemma

D. □

c) Squaring both sides of $X_{2it} = \theta_T X_{2it-1} + u_{2it}$, adding and subtracting X_{2it-1}^2 , solving for it and summing over t we obtain:

$$\begin{aligned} \frac{\theta_T^{-2T}}{k_T^2} \sum_{t=1}^T X_{2t-1}^2 &= \frac{1}{k_T(\theta_T^2 - 1)} \left(\frac{\theta_T^{-2T}}{k_T} X_{2T}^2 - \frac{\theta_T^{-2T}}{k_T} X_{2i0}^2 - 2 \frac{\theta_T^{-2T+1}}{k_T} \sum_{t=1}^T X_{2it-1} u_{2it} - \frac{\theta_T^{-2T}}{k_T} \sum_{t=1}^T u_{2it}^2 \right) \\ &= \frac{1}{k_T(\theta_T^2 - 1)} \left(\frac{\theta_T^{-2T}}{k_T} X_{2T}^2 - 2 \frac{\theta_T^{-2T+1}}{k_T} \sum_{t=1}^T X_{2it-1} u_{2it} - \frac{\theta_T^{-2T}}{k_T} \sum_{t=1}^T u_{2it}^2 \right) + o_p(1) \end{aligned}$$

The last two terms in brackets are $o_p(1)$ as $T \rightarrow \infty$ because:

i)

$$\frac{\theta_T^{-2T}}{k_T} \sum_{t=1}^T u_{2it}^2 = \frac{\theta_T^{-2T}}{k_T} T \frac{1}{T} \sum_{t=1}^T u_{2it}^2 = o_p(1)$$

due to Proposition 1 (b) and $\frac{1}{T} \sum_{t=1}^T u_{2it}^2 \xrightarrow{P} \Omega_{i,22}$.

ii)

$$\begin{aligned} \frac{\theta_T^{-2T+1}}{k_T} \sum_{t=1}^T X_{2it-1} u_{2it} &= \frac{\theta_T^{-2T+1}}{k_T} \sum_{t=1}^T \sum_{j=1}^{t-1} \theta_T^{t-j-1} u_{2ij} u_{2it} + \frac{\theta_T^{-T+1} X_{2i0}}{\sqrt{k_T}} \frac{1}{\sqrt{k_T}} \sum_{t=1}^T \theta_T^{-(T-t)-1} u_{2it} \\ &= \frac{\theta_T^{-2T+1}}{k_T} \sum_{t=1}^T \sum_{j=1}^{t-1} \theta_T^{t-j-1} u_{2ij} u_{2it} + o_p(1) O_p(1) \end{aligned}$$

as $T \rightarrow \infty$. The first term converges to 0 in $\mathcal{L}1$ norm:

$$\begin{aligned} \mathbb{E} \left[\left| \frac{\theta_T^{-2T+1}}{k_T} \sum_{t=1}^T \sum_{j=1}^{t-1} \theta_T^{t-j-1} u_{2ij} u_{2it} \right| \right] &\leq \frac{\theta_T^{-2T+1}}{k_T} \sum_{t=1}^T \sum_{j=1}^{t-1} \theta_T^{t-j-1} \mathbb{E}(|u_{2ij}| |u_{2it}|) \\ &\leq \mathbb{E}[|u_{2i1}|^2] \frac{\theta_T^{-2T+1}}{k_T} \sum_{t=1}^T \sum_{j=1}^{t-1} \theta_T^{t-j-1} \\ &= \mathbb{E}[|u_{2i1}|^2] \frac{\theta_T^{-2T+1}}{k_T} \sum_{t=1}^T \theta_T^{t-1} \frac{1 - \theta_T^{-(t-1)}}{\theta_T - 1} \\ &= \mathbb{E}[|u_{2i1}|^2] \frac{\theta_T^{-2T+1}}{k_T(\theta_T - 1)} \frac{1 - \theta_T^T}{1 - \theta_T} - \frac{\mathbb{E}[|u_{2i1}|^2]}{b} \theta_T^{-2T+1} T \\ &= \frac{\mathbb{E}[|u_{2i1}|^2]}{b} \theta_T^{-T+1} k_T - \frac{\mathbb{E}[|u_{2i1}|^2]}{b} \theta_T^{-2T+1} k_T + o(1) = o(1) \end{aligned}$$

as $T \rightarrow \infty$ due to exponential decay of θ_T^{-2T+1} . Finally:

$$\frac{1}{k_T(\theta_T^2 - 1)} \frac{\theta_T^{-2T}}{k_T} X_{2T}^2 = \frac{1}{k_T(\theta_T^2 - 1)} \left(\frac{1}{\sqrt{k_T}} \sum_{j=1}^n \theta_T^{-j} u_{2ij} + o_p(1) \right)^2 \xrightarrow{D} \frac{1}{2b} C_{2i}^2(1) X_{2i}^2(b)$$

as $T \rightarrow \infty$ by Lemma D, CMT and Proposition 1 (c). \square

Main Results to Bridge Local to Unity and Mildly Explosive Behavior

Lemma F

This Lemma is a direct generalization of the results bridging local to unity and mildly explosive asymptotics in [Phillips and Lee \(2015\)](#). The proof is provided for completeness.

a) $J_{jic}(r)$ is independent from $X_{ji}(b)$ and $Y_{ji}(b)$ for $j = 1, 2$ and $i = 1, \dots, N$.

b) For all $s, r > 0$ the following joint convergence applies:

$\left[\frac{X_{1i|Tr}}{\sqrt{T}}, \frac{X_{2i|Ts}}{\sqrt{k_T \theta_T^{Ts}}} \right] \xrightarrow{D} [C_{1i}(1) J_{1ic}(r), X_{2i}(b)]$ as $T \rightarrow \infty$ and $[\cdot]$ is a function picking the nearest integer value of the argument.

c) $\frac{1}{Tk_T \theta_T^T} \sum_{t=1}^T X_{1it-1} X_{2it-1} = o_p(1)$ as $T \rightarrow \infty$.

Proof of Lemma F

a) Clearly $J_{1ic}(r)$ is just a functional of $B_{1i}(r)$, which is a Gaussian process. Also, $X_{2i}(b)$ is a Gaussian random variable, hence it is sufficient to check the asymptotic covariance between

$$\frac{1}{\sqrt{T}} \sum_{j=1}^T u_{1ij} = \frac{1}{\sqrt{T}} \sum_{j=1}^T C_{1i}(1) \epsilon_{1ij} + o_p(1)$$

and

$$\frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} u_{2ij} = \frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} C_{2i}(1) \epsilon_{2j} + o_p(1)$$

by Lemma B (a). These terms converge to the objects under consideration. For conve-

nience omitting the scaling constants, we obtain:

$$\begin{aligned}
& \mathbb{E} \left[\left(\frac{1}{\sqrt{T}} \sum_{j=1}^T \epsilon_{1ij} \right) \left(\frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-j} \epsilon_{2ij} \right) \right] \\
&= \frac{\sigma_{i.12}}{\sqrt{Tk_T}} \sum_{j=1}^T \theta_T^{-j} = \frac{\sigma_{i.12}}{\sqrt{Tk_T}} \frac{1}{\theta_T} \frac{1 - \theta_T^{-T}}{1 - \theta_T^{-1}} \\
&= \frac{\sigma_{i.12}}{b} \sqrt{\frac{k_T}{T}} \{1 + o(1)\} = o(1)
\end{aligned}$$

as $T \rightarrow \infty$ using the definition of θ_T . This establishes independence between $X_{2i}(b)$ and $B_{1i}(1)$, but the same holds for any $r \in [0, 1]$. Independence between $B_{1i}(1)$ and $Y_{2i}(b)$ is established similarly:

$$\begin{aligned}
& \mathbb{E} \left[\left(\frac{1}{\sqrt{T}} \sum_{j=1}^T \epsilon_{1ij} \right) \left(\frac{1}{\sqrt{k_T}} \sum_{j=1}^T \theta_T^{-(T-j)-1} \epsilon_{2ij} \right) \right] \\
&= \frac{\sigma_{i.12}}{\sqrt{Tk_T}} \sum_{j=1}^T \theta_T^{-(T-j)-1} = \frac{\sigma_{i.12}}{\sqrt{Tk_T}} \theta_T^{-T} \frac{\theta_T^T - 1}{\theta_T - 1} \\
&= \frac{\sigma_{i.12}}{b} \sqrt{\frac{k_T}{T}} \{1 + o(1)\} = o(1)
\end{aligned}$$

as $T \rightarrow \infty$. Note that by the same argument $B_{1i}(1)$ is independent from $Y_{1i}(b)$ and $B_{2i}(1)$ will be independent from $X_{2i}(b)$, $Y_{2i}(b)$ and $Y_{1i}(b)$ - only $\sigma_{i.12}$ will be substituted for another member of $\Sigma_{i\epsilon}$ in the proof. \square

b) The marginal convergence of $\frac{X_{1i|T_T|}}{\sqrt{T}}$ is established by the functional theory in [Phillips \(1987\)](#) in connection to Proposition 1 (a). To proceed, define an integer sequence $L_T \rightarrow \infty$ such that $\frac{k_T}{L_T} = o(1)$. Then, clearly by Lemma E, $\frac{X_{2L_T}}{\sqrt{k_T \theta_T^{L_T}}} \xrightarrow{D} C_{2i}(1) X_{2i}(b)$. Next, observe that by the similar mechanics used in Lemma B, one can obtain:

$$\begin{aligned}
& \frac{1}{\sqrt{k_T}} \sum_{j=L_T+1}^n \theta_T^{-j} \Delta \tilde{\epsilon}_{2ij} = \frac{1}{\sqrt{k_T}} \sum_{j=L_T+1}^T \theta_T^{-j} \tilde{\epsilon}_{2ij} - \frac{1}{\sqrt{k_T}} \sum_{j=L_T+1}^T \theta_T^{-1-j} \tilde{\epsilon}_{2ij} + o_p(1) \\
&= \frac{1}{\sqrt{k_T}} (1 - \theta_T^{-1}) \sum_{j=L_T+1}^T \theta_T^{-j} \tilde{\epsilon}_{2ij} + o_p(1) \\
&= \theta_T^{-1} \frac{b}{k_T^{\frac{2}{3}}} \sum_{j=L_T+1}^T \theta_T^{-j} \tilde{\epsilon}_{2ij} + o_p(1)
\end{aligned}$$

Now, the first term converges to 0 in $\mathcal{L}1$ norm:

$$\begin{aligned}
& \mathbb{E} \left[\left| \theta_T^{-1} \frac{b}{k_T^{\frac{3}{2}}} \sum_{j=L_T+1}^T \theta_T^{-j} \tilde{\epsilon}_{2ij} \right| \right] \leq \theta_T^{-1} \frac{b}{k_T^{\frac{3}{2}}} \sum_{j=L_T+1}^T \theta_T^{-j} \mathbb{E}[|\tilde{\epsilon}_{2ij}|] \\
& = \mathbb{E}[|\tilde{\epsilon}_{2i1}|] \theta_T^{-1} \frac{b}{k_T^{\frac{3}{2}}} \frac{1}{\theta_T^{L_T+1}} \frac{1 - \theta_T^{-T+L_T}}{1 - \theta_T^{-1}} \\
& = \mathbb{E}[|\tilde{\epsilon}_{2i1}|] \theta_T^{-1} \frac{1}{\sqrt{k_T}} (\theta_T^{-L_T} - \theta_T^{-T}) = o(1)
\end{aligned}$$

as $T \rightarrow \infty$. Given this, we receive the following:

$$\frac{X_{2iT}}{\sqrt{k_T \theta_T^T}} = \frac{X_{2iL_T}}{\sqrt{k_T \theta_T^{L_T}}} + \frac{1}{\sqrt{k_T}} \sum_{j=L_T+1}^T \theta_T^{-j} u_{2ij} = \frac{X_{2iL_T}}{\sqrt{k_T \theta_T^{L_T}}} + \frac{1}{\sqrt{k_T}} \sum_{j=L_T+1}^T \theta_T^{-j} C_{2i}(1) \epsilon_{2ij} + o_p(1)$$

While the first term converges in distribution according to Lemma D, the second one converges to 0 in $\mathcal{L}2$ norm:

$$\begin{aligned}
& \mathbb{E} \left[\left| \frac{1}{\sqrt{k_T}} \sum_{j=L_T+1}^T \theta_T^{-j} C_{2i}(1) \epsilon_{2ij} \right|^2 \right] = \frac{C_{2i}^2(1) \sigma_{i.22}^2}{k_T} \sum_{j=L_T+1}^T \theta_T^{-2ij} \\
& = \frac{C_{2i}^2(1) \sigma_{i.22}^2}{k_T} \frac{1}{\theta_T^{2L_T+2}} \frac{1 - \theta_T^{-2T+2L_T}}{1 - \theta_T^{-2}} \\
& = \frac{C_{2i}^2(1) \sigma_{i.22}^2}{k_T (\theta_T^2 - 1)} (\theta_T^{-2L_T} - \theta_T^{-2T}) = o(1)
\end{aligned}$$

as $T \rightarrow \infty$. Now simply let $L_T = \lfloor Ts \rfloor$ for $s > 0$ which preserves $\frac{k_T}{L_T} = o(1)$ and $\frac{X_{2i\lfloor ns \rfloor}}{\sqrt{k_T \theta_T^{\lfloor ns \rfloor}}} \xrightarrow{D} C_{2i}(1) X_{2i}(b)$. Joint convergence is established by independence from part a). \square

c) As in [Phillips and Lee \(2015\)](#) we, by Skorokhod Representation Theorem, choose an alternative probability space such that:

$\left[\frac{X_{1i\lfloor Tr \rfloor}}{\sqrt{T}}, \frac{X_{2i\lfloor Ts \rfloor}}{\sqrt{k_T \theta_T^{\lfloor Ts \rfloor}}} \right] \xrightarrow{P} [C_{1i}(1) J_{1ic}(r), C_{2i}(1) X_{2i}(b)]$ as $T \rightarrow \infty$ because we need convergence to random variables. Picking L_T such that $\frac{L_T}{T} = o(1)$ we obtain:

$$\begin{aligned}
\frac{1}{Tk_T\theta_T^T} \sum_{t=1}^T X_{1it-1}X_{2it-1} &= \frac{1}{\sqrt{Tk_T\theta_T^T}} \sum_{t=L_T+1}^T \frac{X_{1it-1}}{\sqrt{T}} \frac{X_{2it-1}}{\sqrt{k_T\theta_T^{t-1}}} \theta_T^{t-1} \\
&+ \frac{\theta_T^{L_T}}{\sqrt{Tk_T\theta_T^T}} \sum_{t=1}^{L_T} \frac{X_{1it-1}}{\sqrt{T}} \frac{X_{2it-1}}{\sqrt{k_T\theta_T^{t-1}}} \frac{\theta_T^{t-1}}{\theta_T^{L_T}} \\
&= \frac{C_{2i}(1)X_{2i}(b)}{\sqrt{Tk_T\theta_T^T}} \sum_{t=L_T+1}^T C_{1i}(1)J_{1ic}\left(\frac{t}{T}\right)\theta_T^{t-1}\{1+o_p(1)\} + O_p\left(\frac{L_T\theta_T^{L_T}}{\sqrt{Tk_T\theta_T^T}}\right) \\
&= \frac{C_{2i}(1)X_{2i}(b)}{\sqrt{Tk_T\theta_T^T}} \sum_{t=1}^T C_{1i}(1)J_{1ic}\left(\frac{t}{T}\right)\theta_T^{t-1} + o_p(1)
\end{aligned}$$

Now, the remaining term, excluding $X_{2i}(b)$, converges to 0 in $\mathcal{L}2$ norm. Fixing $M > 0$ to bound the finite covariance of the O-U process (since the variance is finite), we obtain:

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{1}{\sqrt{Tk_T\theta_T^T}} \sum_{t=1}^T C_{1i}(1)J_{1ic}\left(\frac{t}{T}\right)\theta_T^{t-1}\right|^2\right] &= \frac{1}{Tk_T\theta_T^{2T}} \frac{C_{1i}^2(1)}{\theta_T^2} \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}\left[J_{1ic}\left(\frac{t}{T}\right)J_{1ic}\left(\frac{s}{T}\right)\right]\theta_T^{t+s} \\
&\leq \frac{1}{Tk_T\theta_T^{2T}} M \frac{C_{1i}^2(1)}{\theta_T^2} \sum_{t=1}^T \sum_{s=1}^T \theta_T^{t+s} \\
&= \frac{1}{Tk_T\theta_T^{2T}} M \frac{C_{1i}^2(1)}{\theta_T^2} \left(\sum_{t=1}^T \theta_T^t\right)^2 \\
&= \frac{1}{Tk_T\theta_T^{2T}} M C_{1i}^2(1) \frac{(\theta_T^T - 1)^2}{(\theta_T - 1)^2} \\
&= \frac{1}{Tk_T\theta_T^{2T}} M C_{1i}^2(1) \frac{(k_T(\theta_T^T - 1))^2}{b^2} \\
&\leq \frac{1}{Tk_T\theta_T^{2T}} M' \frac{k_T^2\theta_T^{2T}}{b^2} = M' \frac{k_T}{Tb^2} = o(1)
\end{aligned}$$

as $T \rightarrow \infty$. Here, M' is another finite constant for bounding. Hence, $\frac{1}{Tk_T\theta_T^T} \sum_{t=1}^T X_{1it-1}X_{2it-1} = o_p(1)$ because the result holds in the original sample space. \square

Proof of Theorem 3.1

Inserting (2.1) into (3.1), expanding and post-multiplying by \mathbf{D}_T we obtain:

$$(\hat{\mathbf{R}}_T - \mathbf{R}_T)\mathbf{D}_T = \left(\sum_{i=1}^N \sum_{t=2}^T u_{it} X_{it-1}^\top \mathbf{D}_T^{-1}\right) \left(\sum_{i=1}^N \mathbf{D}_T^{-1} \sum_{t=2}^T X_{it-1} X_{it-1}^\top \mathbf{D}_T^{-1}\right)^{-1}$$

By splitting the limit, we consider two terms separately:

i)

$$\begin{aligned} \sum_{i=1}^N \sum_{t=2}^T u_{it} X_{it-1}^\top \mathbf{D}_T^{-1} &= \begin{bmatrix} \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T X_{1it-1} u_{1it} & \sum_{i=1}^N \frac{1}{k_T \theta_T^2} \sum_{t=2}^T X_{2it-1} u_{1it} \\ \sum_{i=1}^N \frac{1}{T} \sum_{t=2}^T X_{1it-1} u_{2it} & \sum_{i=1}^N \frac{1}{k_T \theta_T^2} \sum_{t=2}^T X_{2it-1} u_{2it} \end{bmatrix} \\ \Rightarrow &\begin{bmatrix} \sum_{i=1}^N C_{1i}^2(1) \int_0^1 J_{1ic}(r) dB_{1i}(r) + \sum_{i=1}^N \Lambda_{i,11} & \sum_{i=1}^N C_{2i}(1) C_{1i}(1) X_{2i}(b) Y_{1i}(b) \\ \sum_{i=1}^N C_{1i}(1) C_{2i}(1) \int_0^1 J_{1ic}(r) dB_{2i}(r) + \sum_{i=1}^N \Lambda_{i,21} & \sum_{i=1}^N C_{2i}^2(1) X_{2i}(b) Y_{2i}(b) \end{bmatrix} \end{aligned}$$

as $T \rightarrow \infty$ by Lemma A and Lemma E. Joint convergence is implied by the independence results in Lemma D and Lemma F.

ii)

$$\begin{aligned} \sum_{i=1}^N \mathbf{D}_T^{-1} \sum_{t=2}^T X_{it-1} X_{it-1}^\top \mathbf{D}_T^{-1} &= \begin{bmatrix} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T X_{1it-1}^2 & \sum_{i=1}^N \frac{1}{Tk_T \theta_T^2} \sum_{t=2}^T X_{2it-1} X_{1it-1} \\ \sum_{i=1}^N \frac{1}{Tk_T \theta_T^2} \sum_{t=2}^T X_{2it-1} X_{1it-1} & \sum_{i=1}^N \frac{1}{k_T^2 \theta_T^2} \sum_{t=2}^T X_{2it-1}^2 \end{bmatrix} \\ \Rightarrow &\begin{bmatrix} \sum_{i=1}^N C_{1i}^2(1) \int_0^1 J_{1ic}^2(r) dr & 0 \\ 0 & \frac{1}{2b} \sum_{i=1}^N C_{2i}^2(1) X_{2i}^2(b) \end{bmatrix} \end{aligned}$$

by Proposition 2 (b), Lemma E and Lemma F. Joint convergence is implied by the independence results in Lemma D and Lemma F. The final result in (3.4) follows from repeating the same steps with the asymptotically equivalent \mathbf{F}_T to remove the nuisance $\frac{1}{2b}$, invoking Continuous Mapping Theorem (CMT) and multiplying i) and ii). \square

Proof of Theorem 3.2.

Similarly to Theorem 3.1, inserting (2.1) into (3.17), expanding and transposing we obtain:

$$(\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T)^\top = \left(\sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right)^{-1} \left(\sum_{i=1}^N \left[\sum_{t=2}^T u_{it} X_{it-1}^\top - \sqrt{NT} \hat{\Lambda}_i \right] \right)^\top$$

Using $\text{vec}(\mathbf{AB}) = (\mathbf{I} \otimes \mathbf{A})\text{vec}(\mathbf{B})$ where \mathbf{A} and \mathbf{B} are generic square matrices, we obtain:

$$\begin{aligned} & \text{vec}([\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T]^\top) \\ &= \left(\mathbf{I}_2 \otimes \left[\sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right]^{-1} \right) \text{vec} \left(\left[\sum_{i=1}^N \left[\sum_{t=2}^T u_{it} X_{it-1}^\top - \sqrt{NT} \hat{\Lambda}_i \right] \right]^\top \right) \end{aligned}$$

Using $(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}) = (\mathbf{A} \otimes \mathbf{B})^{-1}$ for invertible square matrices \mathbf{A} and \mathbf{B} , we can split the product above and consider both terms separately:

i)

$$\left(\mathbf{D}_{NT}^{-1} \left[\mathbf{I}_2 \otimes \sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right] \mathbf{D}_{NT}^{-1} \right)^{-1}$$

Putting $\sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \equiv \mathbf{L}$ and using $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$ for the generic square matrices, we obtain:

$$(\mathbf{I}_2 \otimes \mathbf{d}_{NT}^{-1})(\mathbf{I}_2 \otimes \mathbf{L})(\mathbf{I}_2 \otimes \mathbf{d}_{NT}^{-1}) = \mathbf{I}_2 \otimes (\mathbf{d}_{NT}^{-1} \mathbf{L} \mathbf{d}_{NT}^{-1})$$

The identical blocks on the diagonal are the following:

$$\begin{aligned} & \mathbf{d}_{NT}^{-1} \mathbf{L} \mathbf{d}_{NT}^{-1} \\ &= \begin{bmatrix} \frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T X_{1it-1}^2 & \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{Tk_T \theta_T^T} \sum_{t=2}^T X_{2it-1} X_{1it-1} \\ \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{1}{Tk_T \theta_T^T} \sum_{t=2}^T X_{2it-1} X_{1it-1} & \frac{1}{N} \sum_{i=1}^N \frac{1}{k_T^2 \theta_T^2} \sum_{t=2}^T X_{2it-1}^2 \end{bmatrix} \\ & \xrightarrow{P} \begin{bmatrix} \mathcal{S}_2 \int_0^1 \int_0^r e^{2(r-s)c} & 0 \\ 0 & \mathcal{S}_5 \frac{1}{4b^2} \end{bmatrix} \equiv \mathbf{B}^* \end{aligned}$$

as $(T, N)_{seq} \rightarrow \infty$ by Proposition 2 (b), Proposition 3, Lemma E and Lemma F. Here:

$$\mathcal{S}_2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i.11}, \quad \mathcal{S}_5 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i.22}$$

Therefore, by the CMT, as $(T, N)_{seq} \rightarrow \infty$:

$$\left(\mathbf{D}_{NT}^{-1} \left[\mathbf{I}_2 \otimes \sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right] \mathbf{D}_{NT}^{-1} \right)^{-1} \xrightarrow{P} (\mathbf{I}_2 \otimes \mathbf{B}^*)^{-1}$$

ii)

$$\begin{aligned} & \mathbf{D}_{NT}^{-1} \text{vec} \left(\left[\sum_{i=1}^N \left[\sum_{t=2}^T u_{it} X_{it-1}^\top - \sqrt{NT} \hat{\Lambda}_i \right] \right]^\top \right) \\ &= \mathbf{D}_{NT}^{-1} \sum_{i=1}^N \left[\sum_{t=2}^T \text{vec}(X_{it-1} u_{it}^\top) - \sqrt{NT} \text{vec}(\hat{\Lambda}_i^\top) \right] \end{aligned}$$

Writing the block $\mathbf{d}_{NT}^{-1} = \frac{1}{\sqrt{N}} \begin{bmatrix} \frac{1}{T} & 0 \\ 0 & \frac{1}{k_T \theta_T^\top} \end{bmatrix}$, as $T \rightarrow \infty$, first, we obtain⁷:

$$\mathbf{D}_{NT}^{-1} \sum_{i=1}^N \left[\sum_{t=2}^T \text{vec}(X_{it-1} u_{it}^\top) - \sqrt{NT} \text{vec}(\hat{\Lambda}_i^\top) \right] \Rightarrow \frac{1}{\sqrt{N}} \sum_{i=1}^N M_i^*$$

Where the vector M_i is the outcome of Proposition 2, Lemma E and Lemma F:

$$M_i = \text{vec} \begin{bmatrix} C_{1i}^2(1) \int_0^1 J_{1ic}(r) dB_{1i}(r) & C_{1i}(1) C_{2i}(1) \int_0^1 J_{1ic}(r) dB_{2i}(r) \\ C_{2i}(1) C_{1i}(1) X_{2i}(b) Y_{1i}(b) & C_{2i}^2(1) X_{2i}(b) Y_{2i}(b) \end{bmatrix}$$

Here, $\mathbb{E}[M_i] = 0$ because of independence of the products of Normals and the expectation of the integrals is 0. The elements in the limiting covariance matrix follow from the Assumption 3, Proposition 3 and Lemma F (a) (independence between Normal random variables and functionals of Brownian motions):

$$\begin{aligned} \mathcal{M}^* &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E}[M_i^* M_i^{*\top}] \\ &= \begin{bmatrix} \mathcal{S}_1 \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 & \mathcal{S}_3 \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 \\ 0 & \frac{\mathcal{S}_4}{4b^2} & 0 & 0 \\ \mathcal{S}_3 \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 & \mathcal{S}_4 \int_0^1 \int_0^r e^{2(r-s)c} ds dr & 0 \\ 0 & 0 & 0 & \frac{\mathcal{S}_7}{4b^2} \end{bmatrix} \end{aligned}$$

⁷Here, similarly to Theorem 3.1, as $T \rightarrow \infty$, the correcting terms on the mildly explosive side (where we have no additive bias) vanish due to Proposition 1 (b).

which is a positive definite matrix. Here:

$$\begin{aligned}\mathcal{S}_1 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i,11}^2, & \mathcal{S}_3 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i,21} \Gamma_{i,11} \\ \mathcal{S}_4 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i,11} \Gamma_{i,22}, & \mathcal{S}_7 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i,22}^2\end{aligned}$$

Therefore, as $(T, N)_{seq} \rightarrow \infty$, using conditions iii) - vi) in Assumption 2 we in total obtain:

$$\mathbf{D}_{NT} \text{vec}([\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T]^\top) \xrightarrow{D} \mathcal{N}(0, (\mathbf{I}_2 \otimes \mathbf{B}^*)^{-1} \mathcal{M}^* (\mathbf{I}_2 \otimes \mathbf{B}^*)^{-1})$$

The final result in (3.18) follows from repeating the same steps with the asymptotically equivalent \mathbf{F}_{NT} to remove the multiplicative nuisance (and obtain \mathbf{B} and \mathcal{M}) in terms of b after letting $T \rightarrow \infty$, invoking CMT and multiplying i) and ii). \square

Proof of Theorem 4.1.

We will use:

$$\begin{aligned}W_{NT} &= \frac{\left[a_1^\top \text{vec}(\hat{\mathbf{R}}_T^{FM}) \right]^2}{a_1^\top \left[\mathbf{Q}^{-1} \otimes \mathbf{I}_2 \right] \hat{\mathbf{\Sigma}} \left[\mathbf{Q}^{-1} \otimes \mathbf{I}_2 \right] a_1} \\ &= \frac{\left[a_1^\top \text{vec}(\sqrt{NT} [\hat{\mathbf{R}}_T^{FM}]) \right]^2}{a_1^\top \left[\left(\frac{1}{NT^2} \mathbf{Q} \right)^{-1} \otimes \mathbf{I}_2 \right] \frac{1}{N} \sum_{i=1}^N \hat{\mathbf{\Sigma}}_i \left[\left(\frac{1}{NT^2} \mathbf{Q} \right)^{-1} \otimes \mathbf{I}_2 \right] a_1}\end{aligned}$$

Under the $H_0 : \mathbf{R}_T = \rho_T \mathbf{I}_2$, we have:

$$\begin{aligned}& \text{vec}[\sqrt{NT}(\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T)] \\ &= \text{vec} \left[\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \left[\sum_{t=2}^T u_{it} X_{it-1}^\top - \sqrt{NT} \hat{\Lambda}_i \right] \right) \left(\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right)^{-1} \right]\end{aligned}$$

Now, using the relationship between the vectorization operator and the Kronecker product $\text{vec}(\mathbf{AB}) = (\mathbf{B} \otimes \mathbf{I}_2) \text{vec}(\mathbf{A})$ when $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{2 \times 2}$ we obtain:

$$\begin{aligned}
& \text{vec}[\sqrt{NT}(\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T)] \\
&= \left[\left(\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right)^{-1} \otimes \mathbf{I}_2 \right] \text{vec} \left[\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \left[\sum_{t=2}^T u_{it} X_{it-1}^\top - \sqrt{NT} \hat{\Lambda}_i \right] \right) \right] \\
&= \left[\left(\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right)^{-1} \otimes \mathbf{I}_2 \right] \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \left[\sum_{t=2}^T \text{vec}[u_{it} X_{it-1}^\top] - \sqrt{NT} \text{vec}[\hat{\Lambda}_i] \right] \right)
\end{aligned}$$

We can analyze the limits separately:

i) As first $T \rightarrow \infty$ we obtain the following convergence:

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \left[\sum_{t=2}^T \text{vec}[u_{it} X_{it-1}^\top] - \sqrt{NT} \text{vec}[\hat{\Lambda}_i] \right] \Rightarrow \frac{1}{\sqrt{N}} \sum_{i=1}^N \Pi_i$$

where we define Π_i as:

$$\Pi_i = \text{vec} \begin{bmatrix} C_{1i}^2(1) \int_0^1 J_{1ic}(r) dB_{1i}(r) & C_{1i}(1) C_{2i}(1) \int_0^1 J_{2ic}(r) dB_{1i}(r) \\ C_{1i}(1) C_{2i}(1) \int_0^1 J_{1ic}(r) dB_{2i}(r) & C_{2i}^2(1) \int_0^1 J_{2ic}(r) dB_{2i}(r) \end{bmatrix} \in \mathbb{R}^4$$

Then, as $N \rightarrow \infty$, using the conditions iii) - vi) in Assumption 2 we obtain the final convergence result:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \Pi_i \xrightarrow{D} \mathcal{N}(0, \Xi) \in \mathbb{R}^4$$

where

$$\Xi = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Xi_i$$

Here, $\Xi_i = \mathbb{E}[\Pi_i \Pi_i^\top]$ which is found invoking the results on Itô Integrals in Proposition 3.

ii)

$$\begin{aligned}
& \frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \\
&= \begin{bmatrix} \frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T X_{1it-1}^2 & \frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T X_{1it-1} X_{2it-1} \\ \frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T X_{1it-1} X_{2it-1} & \frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T X_{2it-1}^2 \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} \mathcal{S}_2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr & \mathcal{S}_6 \int_0^1 \int_0^r e^{2(r-s)c} ds dr \\ \mathcal{S}_6 \int_0^1 \int_0^r e^{2(r-s)c} ds dr & \mathcal{S}_5 \int_0^1 \int_0^r e^{2(r-s)c} ds dr \end{bmatrix} = \Theta
\end{aligned}$$

as $(T, N)_{seq} \rightarrow \infty$ applying Proposition 3 and Theorem 3.2 when both columns are local to unity. Here:

$$\begin{aligned}
\mathcal{S}_2 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i.11} \\
\mathcal{S}_5 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i.22} \\
\mathcal{S}_6 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i.21}
\end{aligned}$$

Combining the results and using them in the Wald statistic by the CMT, under H_0 , we obtain the following limiting distribution as $(T, N)_{seq} \rightarrow \infty$:

$$\begin{aligned}
W_{NT} &= \frac{\left[a_1^\top \text{vec}(\sqrt{NT}[\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T]) \right]^2}{a_1^\top \left[\left(\frac{1}{NT^2} \mathbf{Q} \right)^{-1} \otimes \mathbf{I}_2 \right] \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Sigma}}_i \left[\left(\frac{1}{NT^2} \mathbf{Q} \right)^{-1} \otimes \mathbf{I}_2 \right] a_1} \\
&\xrightarrow{D} \frac{\left[\mathcal{N}(0, a_1^\top [\Theta^{-1} \otimes \mathbf{I}_2] \Xi [\Theta^{-1} \otimes \mathbf{I}_2] a_1) \right]^2}{a_1^\top [\Theta^{-1} \otimes \mathbf{I}_2] \Xi [\Theta^{-1} \otimes \mathbf{I}_2] a_1} \sim \chi_1^2
\end{aligned}$$

□

Proof of Theorem 4.2.

Under the $H_1 : \mathbf{R}_T = \text{diag}(\rho_T, \theta_T)$ we can write:

$$\begin{aligned} a_1^\top \text{vec}(\sqrt{NT} \hat{\mathbf{R}}_T^{FM}) &= \sqrt{NT}(\hat{r}_{11}^{FM} - \hat{r}_{22}^{FM}) = \sqrt{NT}(\hat{r}_{11}^{FM} - \rho_T) \\ &\quad - \sqrt{NT}(\hat{r}_{22}^{FM} - \theta_T) + \sqrt{NT}(\rho_T - \theta_T) \\ &= \sqrt{NT}(\hat{r}_{11}^{FM} - \rho_T) + \sqrt{NT}(\hat{r}_{22}^{FM} - \theta_T) + \left(\sqrt{N}c - \frac{\sqrt{NT}b}{k_T} \right) \end{aligned}$$

Here, $\sqrt{NT}(\hat{r}_{11}^{FM} - r_{11}) = O_p(1)$ as $(T, N)_{\text{seq}} \rightarrow \infty$ by Theorem 3.2. Further:

$$\begin{aligned} \sqrt{NT}(\hat{r}_{22}^{FM} - \theta_T) &= \frac{1}{k_T \theta_T^T} \sqrt{NT} k_T \theta_T^T (\hat{r}_{22}^{FM} - r_{22}) \\ &= \sqrt{N} \frac{T}{k_T \theta_T^T} \frac{k_T \theta_T^T \sum_{i=1}^N \left[\sum_{t=2}^T X_{2it-1} u_{2it} - \sqrt{NT} \hat{\Lambda}_{i,22} \right]}{\sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2} \\ &= \sqrt{N} \frac{T}{k_T \theta_T^T} \frac{\sum_{i=1}^N \frac{1}{k_T \theta_T^T} \sum_{t=2}^T X_{2it-1} u_{2it} - \sqrt{N} \sum_{i=1}^N \frac{T}{k_T \theta_T^T} \hat{\Lambda}_{i,22}}{\sum_{i=1}^N \frac{1}{k_T^2 \theta_T^{2T}} \sum_{t=2}^T X_{2it-1}^2} \end{aligned}$$

which, for fixed N , is $O_p\left(\frac{T}{k_T \theta_T^T}\right)$, hence it will vanish under sequential limits. The third term diverges for all $c < 0$ and $b > 0$ as $(T, N)_{\text{seq}} \rightarrow \infty$.

Further, to examine the denominator of W_{NT} , we explore asymptotics of the determinant of the matrix \mathbf{Q} :

$$\begin{aligned} \mathbf{Q} &= \sum_{i=1}^N X_{i,-1}^\top X_{i,-1} = \sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \\ &= \begin{bmatrix} \sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2 & \sum_{i=1}^N \sum_{t=2}^T X_{1it-1} X_{2it-1} \\ \sum_{i=1}^N \sum_{t=2}^T X_{1it-1} X_{2it-1} & \sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2 \end{bmatrix} \end{aligned}$$

Hence, the determinant is the following:

$$\begin{aligned}
\det(\mathbf{Q}) &= \sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2 \sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2 - \left(\sum_{i=1}^N \sum_{t=2}^T X_{1it-1} X_{2it-1} \right)^2 \\
&= \sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2 \sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2 \left(1 - \frac{\left(\sum_{i=1}^N \sum_{t=2}^T X_{1it-1} X_{2it-1} \right)^2}{\sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2 \sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2} \right) \\
&= \sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2 \sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2 \left(1 - \frac{\left(\sum_{i=1}^N \frac{1}{Tk_T \theta_T^T} \sum_{t=2}^T X_{1it-1} X_{2it-1} \right)^2}{\sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T X_{1it-1}^2 \sum_{i=1}^N \frac{1}{k_T^2 \theta_T^{2T}} \sum_{t=2}^T X_{2it-1}^2} \right) \\
&= \left(\sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2 \sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2 \right) Z
\end{aligned}$$

where Z contains the higher order terms then the product of sums in parentheses.

This implies that we can write the scaled inverse in the following way:

$$\begin{aligned}
NT^2 \mathbf{Q}^{-1} &= \frac{NT^2}{\det(\mathbf{Q})} \begin{bmatrix} \sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2 & - \sum_{i=1}^N \sum_{t=2}^T X_{1it-1} X_{2it-1} \\ - \sum_{i=1}^N \sum_{t=2}^T X_{1it-1} X_{2it-1} & \sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{NT^2 \sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2} & - \frac{\sum_{i=1}^N \sum_{t=2}^T X_{1it-1} X_{2it-1}}{NT^2 \sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2 \sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2} \\ - \frac{\sum_{i=1}^N \sum_{t=2}^T X_{1it-1} X_{2it-1}}{\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2 \sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2} & \frac{T^2}{\frac{1}{N} \sum_{i=1}^N \sum_{t=2}^T X_{2it-1}^2} \end{bmatrix} \{1 + o_p(1)\} \\
&= \begin{bmatrix} \frac{1}{\frac{1}{NT^2} \sum_{i=1}^N \sum_{t=2}^T X_{1it-1}^2} & - \frac{\frac{T \theta_T^{-T}}{k_T} \sum_{i=1}^N \frac{1}{Tk_T \theta_T^T} \sum_{t=2}^T X_{1it-1} X_{2it-1}}{\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T X_{1it-1}^2 \sum_{i=1}^N \frac{1}{k_T^2 \theta_T^{2T}} \sum_{t=2}^T X_{2it-1}^2} \\ - \frac{\frac{T \theta_T^{-T}}{k_T} \sum_{i=1}^N \frac{1}{Tk_T \theta_T^T} \sum_{t=2}^T X_{1it-1} X_{2it-1}}{\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=2}^T X_{1it-1}^2 \sum_{i=1}^N \frac{1}{k_T^2 \theta_T^{2T}} \sum_{t=2}^T X_{2it-1}^2} & \frac{\frac{T^2 \theta_T^{-2T}}{k_T^2}}{\frac{1}{N} \sum_{i=1}^N \frac{1}{k_T^2 \theta_T^{2T}} \sum_{t=2}^T X_{2it-1}^2} \end{bmatrix} \\
&\{1 + o_p(1)\}
\end{aligned}$$

Clearly, as $T \rightarrow \infty$ all the elements except for the upper left one converge to zero in probability. As $N \rightarrow \infty$ afterwards, we obtain the following matrix by the CMT:

$$\left(\frac{1}{NT^2} \mathbf{Q} \right)^{-1} \xrightarrow{P} \begin{bmatrix} \frac{1}{S_2 \int_0^1 \int_0^r e^{2(r-s)c} ds dr} & 0 \\ 0 & 0 \end{bmatrix} \equiv \mathbf{M}$$

Where $\mathcal{S}_2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i,11}$. Because we identify the local to unity parameter c from the first series which has the same properties under H_1 , we know that $\frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\epsilon}}_i \rightarrow \boldsymbol{\Xi}$ under H_1 as $(T, N)_{seq} \rightarrow \infty$ as well. Therefore, we can analyze $[\mathbf{M} \otimes \mathbf{I}_2] \boldsymbol{\Xi} [\mathbf{M} \otimes \mathbf{I}_2]$. Since each of the three matrices belongs to $\mathbb{R}^{4 \times 4}$, we can conveniently apply block-multiplication by noticing that $[\mathbf{M} \otimes \mathbf{I}_2]$ contains only one non-zero block (the upper left which is a diagonal matrix). Hence:

$$[\mathbf{M} \otimes \mathbf{I}_2] \boldsymbol{\Xi} [\mathbf{M} \otimes \mathbf{I}_2] = \begin{bmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} = \begin{bmatrix} \mathbf{ABA} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

Because two first \mathbf{A} and \mathbf{B} contain reciprocal integral parts, we can write the final product in the following way:

$$\mathbf{ABA} = \begin{bmatrix} \frac{1}{\mathcal{S}_2} & 0 \\ 0 & \frac{1}{\mathcal{S}_2} \end{bmatrix} \begin{bmatrix} \mathcal{S}_1 & \mathcal{S}_3 \\ \mathcal{S}_3 & \mathcal{S}_4 \end{bmatrix} \begin{bmatrix} \frac{1}{\mathcal{S}_2} & 0 \\ 0 & \frac{1}{\mathcal{S}_2} \end{bmatrix} \left(\int_0^1 \int_0^r e^{2(r-s)c} ds dr \right)^{-1}$$

where

$$\begin{aligned} \mathcal{S}_1 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i,11}^2 \\ \mathcal{S}_3 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i,21} \Gamma_{i,11} \\ \mathcal{S}_4 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Gamma_{i,11} \Gamma_{i,22} \end{aligned}$$

Now, given the structure of $[\mathbf{M} \otimes \mathbf{I}_2] \boldsymbol{\Xi} [\mathbf{M} \otimes \mathbf{I}_2]$ under the H_1 , it is clear that $a_1^T [\mathbf{M} \otimes \mathbf{I}_2] \boldsymbol{\Xi} [\mathbf{M} \otimes \mathbf{I}_2] a_1$ produces its upper left element, which also coincides with the upper left element of \mathbf{ABA} that is equal to $\left(\frac{\mathcal{S}_2}{\mathcal{S}_1} \int_0^1 \int_0^r e^{2(r-s)c} ds dr \right)^{-1}$. Therefore, under H_1 as $(T, N)_{seq} \rightarrow \infty$, the behavior of W_{NT} can be described as:

$$W_{NT} \rightarrow \infty$$

because the numerator of Wald statistic has a dominant term $\left(\frac{\sqrt{NT}b}{k_T} \right)^2$ and the denominator converges to $\left(\frac{\mathcal{S}_2}{\mathcal{S}_1} \int_0^1 \int_0^r e^{2(r-s)c} ds dr \right)^{-1}$. \square

Proposition 4. Under H_1 , W_{NT}^B is asymptotically undefined.

Proof of Proposition 4.

i) We will use:

$$\begin{aligned} W_{NT}^B &= \left[\mathbf{A}^\top rvec(\hat{\mathbf{R}}_T^{FM}) \right]^\top \left(\mathbf{A}^\top \left[\mathbf{I}_2 \otimes \mathbf{Q}^{-1} \right] \hat{\boldsymbol{\Xi}} \left[\mathbf{I}_2 \otimes \mathbf{Q}^{-1} \right] \mathbf{A} \right)^{-1} \left[\mathbf{A}^\top rvec(\hat{\mathbf{R}}_T^{FM}) \right] \\ &= \mathbf{P}^\top \left(\mathbf{A}^\top \left[\mathbf{I}_2 \otimes \left(\frac{1}{NT^2} \mathbf{Q} \right)^{-1} \right] \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\Xi}}_i \left[\mathbf{I}_2 \otimes \left(\frac{1}{NT^2} \mathbf{Q} \right)^{-1} \right] \mathbf{A} \right)^{-1} \mathbf{P} \end{aligned}$$

where $\mathbf{P} = \left[\mathbf{A}^\top rvec(\sqrt{NT} \hat{\mathbf{R}}_T^{FM}) \right]$. The matrix in the quadratic form is different because:

$$rvec(\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T) = \left(\mathbf{I}_2 \otimes \left[\sum_{i=1}^N \sum_{t=2}^T X_{it-1} X_{it-1}^\top \right]^{-1} \right) vec \left(\left[\sum_{i=1}^N \left[\sum_{t=2}^T u_{it} X_{it-1}^\top - \sqrt{NT} \hat{\Lambda}_i \right] \right]^\top \right)$$

but, under the null, $\boldsymbol{\Xi}_i$ will serve as the covariance matrix of the vector in the product above as well, as $(T, N)_{seq} \rightarrow \infty$. This is because covariances iii) and iv) in Proposition 3 (g) are the same, thus transposition above will not have any effect on the final expression of the covariance matrix.

Under the null, \mathbf{P} behaves in the following way:

$$\mathbf{P} = \left[\mathbf{A}^\top rvec(\sqrt{NT} \hat{\mathbf{R}}_T^{FM}) \right] = \left[\mathbf{A}^\top rvec(\sqrt{NT} [\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T]) \right]$$

$$\xrightarrow{D} \mathcal{N}(0, \mathbf{A}^\top \left[\mathbf{I}_2 \otimes \boldsymbol{\Theta}^{-1} \right] \boldsymbol{\Xi} \left[\mathbf{I}_2 \otimes \boldsymbol{\Theta}^{-1} \right] \mathbf{A}) \in \mathbb{R}^3$$

as $(T, N)_{seq} \rightarrow \infty$ by the same logic as in the proof of Theorem 4.1 only using the matrix \mathbf{A} . Because the covariance matrix is non-singular under the H_0 , we obtain:

$$W_{NT}^B \xrightarrow{D} \chi_3^2$$

as $(T, N)_{seq} \rightarrow \infty$ by the CMT.

ii) To investigate H_1 , we examine the following vector first:

$$\begin{aligned}
\mathbf{A}^\top rvec(\sqrt{NT}\hat{\mathbf{R}}_T^{FM}) &= \mathbf{A}^\top rvec(\sqrt{NT}[\hat{\mathbf{R}}_T^{FM} - \mathbf{R}_T]) + \mathbf{A}^\top rvec(\sqrt{NT}\mathbf{R}_T) \\
&= [\sqrt{NT}(\hat{r}_{11}^{FM} - \rho_T) - \sqrt{NT}(\hat{r}_{22}^{FM} - \theta_T) \quad \sqrt{NT}(\hat{r}_{12}^{FM} - r_{12}) \quad \sqrt{NT}(\hat{r}_{21}^{FM} - r_{21})]^\top \\
&\quad + [\sqrt{NT}(\rho_T - \theta_T) \quad r_{12} \quad r_{21}]^\top \\
&= \left[O_p(1) - O_p\left(\frac{T}{k_T\theta_T}\right) \quad O_p\left(\frac{T}{k_T\theta_T}\right) \quad O_p(1) \right]^\top + \left[\left(\sqrt{N}c - \frac{\sqrt{NT}b}{k_T} \right) \quad r_{12} \quad r_{21} \right]^\top
\end{aligned}$$

due to Theorem 4.1 and Theorem 4.2. Here, $\sqrt{NT}(\hat{r}_{22}^{FM} - \theta_T)$ and $\sqrt{NT}(\hat{r}_{12}^{FM} - r_{12})$ are of the same order and they vanish when $T \rightarrow \infty$, while $\sqrt{NT}(\hat{r}_{11}^{FM} - \rho_T)$ and $\sqrt{NT}(\hat{r}_{21}^{FM} - r_{21})$ are $O_p(1)$ as $(T, N)_{seq} \rightarrow \infty$. The rest is divergent as $(T, N)_{seq} \rightarrow \infty$ as $c < 0$, $b > 0$. We are left to check if the middle matrix is invertible.

Note that:

$$\mathbf{I}_2 \otimes \left(\frac{1}{NT^2} \mathbf{Q} \right)^{-1} \xrightarrow{P} \mathbf{I}_2 \otimes \mathbf{M} = \begin{bmatrix} \mathbf{M} & \mathbf{O} \\ \mathbf{O} & \mathbf{M} \end{bmatrix}$$

as $(T, N)_{seq} \rightarrow \infty$, where \mathbf{M} is the same as in Theorem 4.2. Therefore:

$$\mathbf{A}^\top \left[\mathbf{I}_2 \otimes \left(\frac{1}{NT^2} \mathbf{Q} \right)^{-1} \right] \frac{1}{N} \sum_{i=1}^N \hat{\Xi}_i \left[\mathbf{I}_2 \otimes \left(\frac{1}{NT^2} \mathbf{Q} \right)^{-1} \right] \mathbf{A} \xrightarrow{P} \mathbf{A}^\top \begin{bmatrix} \mathbf{MBM} & \mathbf{MCM} \\ \mathbf{MDM} & \mathbf{MEM} \end{bmatrix} \mathbf{A} \equiv \mathbf{A}^\top \mathcal{A} \mathbf{A}$$

as $(T, N)_{seq} \rightarrow \infty$ where the blocks \mathbf{B} , \mathbf{C} , \mathbf{D} and \mathbf{E} are the same blocks with the limiting averages of the kernel estimators in $\hat{\Xi}_i$ as in Theorem 4.2. Because of the structure of each 2×2 block in \mathcal{A} (only upper left element is non-zero), we, for example, have:

$$a_2^\top \mathcal{A} a_1 = a_2^\top \mathcal{A} a_2 = a_2^\top \mathcal{A} a_3 = 0$$

which gives a zero second row, therefore $\mathbf{A}^\top \mathcal{A} \mathbf{A}$ is singular. \square