

**LEAST SQUARES AND IVX LIMIT THEORY IN SYSTEMS
OF PREDICTIVE REGRESSIONS WITH GARCH
INNOVATIONS**

By

Tassos Magdalinos

November 9, 2016

RESEARCH INSTITUTE FOR ECONOMETRICS

DISCUSSION PAPER NO. 5-16



Research Institute for Econometrics

מכון מחקר לאקונומטריקה

DEPARTMENT OF ECONOMICS

BAR-ILAN UNIVERSITY

RAMAT-GAN 5290002, ISRAEL

<http://econ.biu.ac.il/en/node/2473>

Least Squares and IVX Limit Theory in Systems
of Predictive Regressions with GARCH
innovations¹

Tassos Magdalinos
University of Southampton, UK

9 November 2016

¹The author acknowledges research support by the ESRC under grant RES-061-25-0517 and by the British Academy under grant SG151171.

Abstract

The paper examines the effect of conditional heteroskedasticity to least squares inference in stochastic regression models. We show that a regressor signal of exact order $O_p^e(n^{1+\alpha})$ for arbitrary $\alpha > 0$ is sufficient to eliminate stationary GARCH effects from the limit distributions of least squares based estimators and self-normalised test statistics. The above order dominates the $O_p^e(n)$ signal of stationary regressors but is dominated by the $O_p^e(n^2)$ signal of I(1) regressors, thereby showing that least squares invariance to GARCH effects is not an exclusively I(1) phenomenon but extends to processes with persistence degree arbitrarily close to stationarity. The theory validates standard inference for self normalised test statistics based on: (i) the OLS estimator when $\alpha \in (0, 1)$; (ii) the IVX estimator (Phillips and Magdalinos, 2009; Kostakis, Magdalinos and Stamatogiannis 2015a) when $\alpha > 0$, when the innovation sequence of the system is a stationary vec-GARCH process. An adjusted version of the IVX testing procedure is shown to also accommodate stationary regressors and produce standard chi-squared inference under conditional heteroskedasticity in the innovations across the full range $\alpha \geq 0$.

Keywords: Central limit theory, Conditional Heteroskedasticity, Mixed Normality, Wald test

AMS 1991 subject classification: 62M10; *JEL classification:* C22

1. Introduction

The effect of conditional heteroskedasticity in autoregressive and stochastic regression models has been a topic of intense research activity since the introduction of ARCH and GARCH processes by Engle (1982) and Bollerslev (1986). Limit theory for stationary autoregressive moving average (ARMA) time series with conditionally heteroskedastic innovations has been developed by Weiss (1986) and Pantula (1989) in the case of ARCH innovations and Ling and McAleer (2003) in the case of vector-valued processes with GARCH innovations. Early work on least squares estimation of non stationary autoregressions with ARCH(1) and GARCH(1,1) innovations can be found in Pantula (1989) and Ling and Li (1997a). Asymptotic theory for quasi maximum likelihood estimation has been developed for both stationary and non stationary times series with GARCH innovations: see Ling and Li (1997b), Ling and Li (1998), Ling and McAleer (2003) and references therein.

The current literature on least squares estimation of autoregressive processes with conditionally heteroskedastic innovations reports that the presence of GARCH effects in the limit distributions of the OLS estimator and the associated t and Wald test statistics depends on the stationarity properties of the autoregressive process. In the case of stationary autoregressions, both the convergence rate and the limit distribution of the OLS estimator are affected: the standard \sqrt{n} -consistency rate requires finite fourth moments (a condition that imposes restrictions on the GARCH coefficients) and, even when the \sqrt{n} rate is achieved, the asymptotic variance of the OLS estimator depends on the GARCH parameters in a way that invalidates standard t and Wald hypothesis tests. The situation is different for models with nonstationary time series, where GARCH innovations make no contribution to the limit distribution of the OLS estimator and the usual Dickey-Fuller type t and Wald tests are asymptotically valid. This asymptotic invariance continues to apply in models with near-integrated time series with local to unity roots of the form $\rho_n = 1 + c/n$, where n is the sample size (Phillips, 1987a, Chan and Wei, 1987) and their vector-valued extensions with autoregressive matrix of the form $R_n = I + C/n$ (Phillips, 1988).

This I(0)-I(1) dichotomy has a signal-to-noise ratio interpretation: a near-integrated regressor has sufficient signal, of order $O_p(n^2)$, to asymptotically eliminate stationary GARCH effects; such elimination cannot be achieved by the weaker $O_p(n)$ signal of a stationary regressor, resulting to the contribution of GARCH effects to the least squares limit distribution. This insight raises the issue of the existence of a minimal order of regression signal required to eliminate GARCH effects, leading naturally to the investigation of intermediate regression signals arising from near stationary time series. The class of near stationary time series, introduced by Phillips and Magdalinos (2007a & 2007b) in the case of scalar autoregressions and Magdalinos and Phillips (2009) in the case of vector autoregressions and systems of regression equations, has intermediate I(0)-I(1) persistence rate driven by an autoregressive root of the form $\rho_n = 1 + c/n^\alpha$, where $c < 0$ and $\alpha \in (0, 1)$. The signal generated by such processes is of

order $O_p(n^{1+\alpha})$ and varies continuously with the exponent α , establishing boundaries with nearly integrated processes as $\alpha \rightarrow 1$ and stationary processes as $\alpha \rightarrow 0$.

The present work develops a limit theory for near-stationary predictive regression systems with general GARCH innovations, assuming covariance stationarity of the GARCH process. We show that, for any $\alpha \in (0, 1)$, GARCH effects are eliminated from the limit distribution of the least squares estimator, thereby establishing that any regressor signal strictly dominating the $O_p(n)$ signal of stationary processes is sufficient to asymptotically eliminate GARCH effects. The OLS estimator has an identical Gaussian limit distribution to that established by Magdalinos and Phillips (2009) under conditionally homoskedastic innovations and the usual Wald statistic (without heteroskedasticity correction) for testing restrictions on the regression coefficient matrix has a standard chi-squared limit distribution. To our knowledge, this is the first result of its kind, with standard Gaussian and chi-squared asymptotics applying respectively to the OLS estimator and the Wald statistic in a stochastic regression model with conditionally heteroskedastic innovations.

The development of least squares limit theory for the case of near stationary regressors is the key step towards extending the validity of the IVX endogenous instrumentation procedure, introduced by Phillips and Magdalinos (2009) and further developed by Kostakis, Magdalinos and Stamatogiannis (2015a) (henceforth PM (2009) and KMS (2015a)) to accommodate the presence of conditional heteroskedasticity in the innovations. In the current predictive regression context, KMS (2015a) show that the IVX procedure is robust to different types of persistence, including purely stationary ($\alpha = 0$), near stationary ($\alpha \in (0, 1)$) and near integrated ($\alpha \geq 1$) time series regressors. In this paper, the method is shown to be robust to GARCH effects near stationary and near integrated systems. In predictive regression with purely stationary regressors, KMS (2015a) show that the IVX and OLS procedures are asymptotically equivalent, so the IVX estimator inherits the usual GARCH effects present in the asymptotic variance of the least squares estimator. A White (1980) type of correction is shown to make the IVX procedure operational for all persistence regimes $\alpha \geq 0$ under conditional heteroskedasticity.

The paper is organised as follows. Section 2 outlines a general modelling framework for a system of predictive regressions with unknown persistence properties and conditionally heteroskedastic innovations of a general covariance stationary vec-GARCH type. Section 3 develops a limit theory for the OLS estimator in the near stationary $\alpha \in (0, 1)$ case and shows that GARCH effects are asymptotically eliminated and do not affect least squares based estimation and hypothesis testing procedures. Section 4 develops a limit theory for the IVX estimator and the associated Wald statistic for systems of predictive regressions of arbitrary integration order and GARCH innovations. Section 5 provides some further discussion of the results and Section 6 includes all proofs.

2. Predictive regression with GARCH innovations

We consider the triangular system of predictive regressions (cf. Magdalinos and Phillips, 2009, 2009b; Kostakis, Magdalinos and Stamatogiannis, 2015)

$$y_t = \mu + Ax_{t-1} + \varepsilon_t, \quad (1)$$

$$x_t = R_n x_{t-1} + u_t, \quad (2)$$

where A is an $m \times r$ coefficient matrix and

$$R_n = I_r + \frac{C}{n^\alpha} \quad \text{for some } \alpha \geq 0 \quad (3)$$

where n denotes the sample size. We distinguish between the following classes of regressor processes:

- (Pi) *Near-integrated regressors, if $\alpha \geq 1$ in (3).*
- (Pii) *Near-stationary regressors, if $\alpha \in (0, 1)$ in (3) and C is a negative stable matrix¹.*
- (Piii) *Stationary regressors, if $\alpha = 0$ in (3) and $R = I_r + C$ satisfies $\|R\| < 1$.*

The system is initialized at some $x_0 = o_p(n^{\alpha/2})$. The stochastic properties of the innovation sequences ε_t and u_t are that satisfy the following condition.

Assumption INNOV. *Let $(\eta_t)_{t \in \mathbb{Z}}$ be a sequence of independent and identically distributed random vectors with $\mathbb{E}(\eta_1) = 0$, $\mathbb{E}(\eta_1 \eta_1') = I_{m+r}$ and $\eta_t = [\eta'_{\varepsilon t}, \eta'_{u t}]'$ with $\eta_{\varepsilon t} \in \mathbb{R}^m$ and $\eta_{u t} \in \mathbb{R}^r$.*

- (i) *The sequence $(\varepsilon_t)_{t \in \mathbb{Z}}$ in (3) is a strictly stationary process admitting the following vec-GARCH(p, q) representation:*

$$\varepsilon_t = H_t^{1/2} \eta_{\varepsilon t}, \quad \text{vech}(H_t) = \varphi + \sum_{i=1}^q A_i \text{vech}(\varepsilon_{t-i} \varepsilon_{t-i}') + \sum_{k=1}^p B_k \text{vech}(H_{t-k}) \quad (4)$$

where φ is a constant vector, A_i, B_k are positive semidefinite matrices for all i, k , and the spectral radius of the matrix $\Gamma = \sum_{i=1}^q A_i + \sum_{k=1}^p B_k$ satisfies $\rho(\Gamma) < 1$.

- (ii) *The sequence u_t in (2) is a linear process*

$$u_t = \sum_{j=0}^{\infty} C_j e_{t-j} \quad \sum_{j=0}^{\infty} j^{1/2} \|C_j\| < \infty \quad (5)$$

¹A square matrix is called negative stable if all its eigenvalues are strictly negative.

where $(C_j)_{j \geq 0}$ a sequence of constant matrices such that $C(1) = \sum_{j=0}^{\infty} C_j$ has full rank and $C_0 = I_r$ and the sequence $(e_t)_{t \in \mathbb{Z}}$ in (5) a strictly stationary vec-GARCH(p, q) process:

$$e_t = \tilde{H}_t^{1/2} \eta_{et}, \quad \text{vech}(\tilde{H}_t) = \tilde{\varphi} + \sum_{i=1}^{\tilde{q}} \tilde{A}_i \text{vech}(e_{t-i} e'_{t-i}) + \sum_{k=1}^{\tilde{p}} \tilde{B}_k \text{vech}(\tilde{H}_{t-k}) \quad (6)$$

where $\tilde{\varphi}$ is a constant vector, \tilde{A}_i, \tilde{B}_k are positive semidefinite matrices for all i, k , and the spectral radius of the matrix $\tilde{\Gamma} = \sum_{i=1}^{\tilde{q}} \tilde{A}_i + \sum_{k=1}^{\tilde{p}} \tilde{B}_k$ satisfies $\rho(\tilde{\Gamma}) < 1$.

(iii) $\mathbb{E} \|\varepsilon_1\|^4 < \infty$ and $\mathbb{E} \|e_1\|^4 < \infty$.

Assumption INNOV accounts for conditionally heteroskedastic innovations with finite fourth moments of a very general form: the vec-GARCH process in (4) and (6) is the most general multivariate GARCH specification (see Chapter 11 of Francq and Zakoian (2010)). The positive semidefinite condition on the matrices A_i, B_k of (4) and the condition on the spectral radius of their sum are part of the standard Boussama (2006) conditions for the existence of a stationary ergodic solution of the vec-GARCH process; see Theorem 11.5 of Francq and Zakoian (2010). The independence of the sequence $(\eta_t)_{t \in \mathbb{Z}}$ and the specification of the innovation processes in (4) and (6) imply that ε_t and e_t are martingale difference sequences with respect to $\mathcal{F}_t := \sigma(\eta_t, \eta_{t-1}, \dots)$ satisfying

$$\mathbb{E}_{\mathcal{F}_{t-1}}(\varepsilon_t \varepsilon'_t) = H_t \quad \text{and} \quad \mathbb{E}_{\mathcal{F}_{t-1}}(e_t e'_t) = \tilde{H}_t. \quad (7)$$

The summability condition (5) is standard in the literature on short-memory linear processes (see Phillips and Solo, 1992). Using the Beveridge Nelson (BN) decomposition, we obtain the following representation for u_t

$$u_t = C(1) e_t - \Delta \tilde{e}_t, \quad \text{for } \tilde{e}_t = \sum_{j=0}^{\infty} \tilde{C}_j e_{t-j}, \quad \tilde{C}_j = \sum_{k=j+1}^{\infty} C_k, \quad (8)$$

where $\sum_{j=0}^{\infty} \|\tilde{C}_j\|^2 < \infty$ is assured by the summability condition in (5) by Lemma 2.1 of Phillips and Solo (1992). Consequently, \tilde{e}_t is a strictly stationary ergodic process satisfying $\mathbb{E} \|\tilde{e}_1\|^2 < \infty$ and the ergodic theorem yields

$$\frac{1}{n} \sum_{t=1}^n \tilde{e}_t u'_t \rightarrow_{a.s.} \mathbb{E}(\tilde{e}_1 u'_1) = \sum_{j=1}^{\infty} \Gamma_u(j) \quad (9)$$

as $n \rightarrow \infty$, where $\Gamma_u(\cdot)$ denotes the autocovariance matrix $\Gamma_u(j) = \mathbb{E}(u_1 u'_{1-j})$.

Denoting the demeaned regression matrices in the system (1)–(2) by $\underline{Y} = \left(\underline{y}'_1, \dots, \underline{y}'_n \right)'$, $\underline{X} = \left(\underline{x}'_0, \dots, \underline{x}'_{n-1} \right)'$, where $\underline{y}'_t = y'_t - \bar{y}'_n$, $\underline{x}'_t = x'_t - \bar{x}'_{n-1}$, $\bar{y}_n = n^{-1} \sum_{t=1}^n y_t$ and $\bar{x}_{n-1} = n^{-1} \sum_{t=1}^n x_{t-1}$, the OLS estimator of A in (1) is given by

$$\hat{A}_n = \underline{Y}' \underline{X} (\underline{X}' \underline{X})^{-1}. \quad (10)$$

The effect of GARCH innovations on the asymptotic theory of the least squares estimator of A is known to differ according to persistence class of the regressor x_t in (2). For stationary processes in class (iii), $\sqrt{n} \text{vec}(\hat{A}_n - A)$ is asymptotically zero mean Gaussian with non-standard asymptotic variance that depends on the GARCH parameters and the fourth moment of the innovations. As a result, the usual self-normalised hypothesis tests will be invalid and a White (1980) type of correction is necessary to obtain correctly sized t and Wald tests. The situation is very different for the near-I(1) processes of class (i), where the non-standard limit distributions of $n \text{vec}(\hat{A}_n - A)$ in the unit root and local to unity cases (Phillips 1987, 1988; Chan and Wei, 1987) are invariant to the presence of GARCH effects and the associated Dickey-Fuller type t and Wald tests remain valid without corrections for conditional heteroskedasticity. This dichotomy has a signal-to-noise ratio interpretation: stationary GARCH effects in the noise of the system (1)–(2) are asymptotically eliminated by the strong signal $\sum_{t=1}^n x_{t-1} x'_{t-1} = O_p(n^2)$ of a near-integrated process in class (i). On the other hand, the weaker $O_p(n)$ signal of a stationary process in class (iii) is not sufficient to eliminate GARCH effects from the noise. Given the vast discrepancy in the order of magnitude of the above signals, a natural question is the existence of a "minimal" order of magnitude for the signal of x_t to asymptotically eliminate GARCH effects. An affirmative answer requires the development of a limit distribution theory for the least squares estimator in the intermediate case of near-stationary regressors of class (ii), undertaken in the next section.

3. Least squares limit theory for near-stationary systems with GARCH innovations

We develop a limit theory for the centred least squares regression estimate

$$n^{\frac{1+\alpha}{2}} (\hat{A}_n - A) = \left(\frac{1}{n^{\frac{1+\alpha}{2}}} \sum_{t=1}^n \varepsilon_t x'_{t-1} \right) \left(\frac{1}{n^{1+\alpha}} \sum_{t=1}^n x_{t-1} x'_{t-1} \right)^{-1} + O_p \left(\frac{1}{n^{\frac{1-\alpha}{2}}} \right) \quad (11)$$

for regressors x_t belonging to the class P(ii) of near stationary processes. The asymptotically negligible term above arises from estimating the intercept in (1) and employing the demeaned series for y_t and x_{t-1} for the construction of the OLS estimator \hat{A}_n ; in the case of stationary and near stationary regressors this demeaning is

eliminated asymptotically². Our approach follows Magdalinos and Phillips (2009) in the sense that we derive a law of large numbers and a martingale central limit theorem, respectively, for the denominator and numerator of the matrix quotient (11) and use this to extract the limit theory. The main technical issue is to obtain the probability limit of the quadratic variation of the martingale transform in the numerator of (11) when ε_t is a vec-GARCH process defined in (4). An approximation to this quadratic variation is achieved by reducing the problem to the existence of a stable solution to a stochastic recurrence relation involving products of innovations and covariates, see Lemma 3.4. Stability of the solution permits standard martingale approximation arguments (Lemma 3.5) that resolve the asymptotics in (11).

To fix ideas, we establish some notation for the recursive equations that we employ in the development of the asymptotic theory. Given the matrices $A_1, \dots, A_q, B_1, \dots, B_p$ in (4), define

$$C_i = \begin{cases} A_i + B_i, & \text{if } i \leq p \wedge q \\ A_i, & \text{if } p < i \leq q \\ B_i, & \text{if } q < i \leq p \end{cases} \quad (12)$$

$$\Gamma_i = I_{r^2} \otimes C_i, \quad \Gamma_{n,i} = R_n^i \otimes R_n^i \otimes C_i \quad (13)$$

and consider the stochastic difference equations:

$$Y(j) = \sum_{l=1}^{\kappa} \Gamma_l Y(j-l) + v(j) \quad (14)$$

$$Y_n(j) = \sum_{l=1}^{\kappa} \Gamma_{n,l} Y_n(j-l) + v_n(j) \quad (15)$$

for $j \geq 1$ and $\kappa := q \vee p$. The companion matrix associated with (15) is given by

$$M_{n,\kappa} = \begin{bmatrix} \Gamma_{n,1} & \Gamma_{n,2} & \dots & \Gamma_{n,\kappa-1} & \Gamma_{n,\kappa} \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & I & 0 \end{bmatrix} \quad (16)$$

where all identity matrices are of order $r^2 m(m+1)/2 \times r^2 m(m+1)/2$. The companion matrix associated with (14), denoted by M_κ , has the same form as the matrix in (16) with $\Gamma_{n,i}$ replaced by Γ_i for all $i \in \{1, \dots, \kappa\}$. It is relatively straightforward to show from first principles that the stationarity conditions for the GARCH process ensure the stability of the solution of (14) and (15).

²See the proof of Theorem 3.7 in the Appendix for details and the proof of (11).

3.1 Lemma. *Under Assumption INNOV, the spectral radius and norm of the companion matrices $M_{n,\kappa}$ and M_κ defined in (16) satisfy:*

- (i) $\rho(M_\kappa) < 1$ and $\sum_{j=0}^{\infty} \|M_\kappa^j\| < \infty$
- (ii) $\rho(M_{n,\kappa}) < 1$ for all but finitely many n , and $\sup_{n \geq 1} \sum_{j=0}^{\infty} \|M_{n,\kappa}^j\| < \infty$.

We begin by providing some useful results involving the regressor x_t and its approximant $\zeta_{n,t}$ arising from the BN decomposition (8).

3.2 Lemma. *Under Assumption INNOV, the process*

$$\zeta_{n,t} = \sum_{j=0}^{t-1} R_n^j C(1) e_{t-j} \quad (17)$$

satisfies the following:

- (i) $\max_{1 \leq t \leq n} \mathbb{E} \|n^{-\alpha/2} \zeta_{n,t}\|^4 = O(1)$.
- (ii) $L_n^{-1} n^{(1+\alpha)/2} \max_{1 \leq t \leq n} \|\zeta_{n,t}\| \rightarrow_p 0$ for an arbitrary sequence $L_n \rightarrow \infty$.
- (iii) $\max_{1 \leq m \leq n} \|n^{-1} \sum_{t=2}^m (\zeta_{n,t-1} \otimes e_t)\|_{L_2} \rightarrow 0$
- (iv) $n^{-(1+\alpha)/2} \left\| \sum_{t=1}^n (x_{t-1} \otimes \varepsilon_t) - \sum_{t=1}^n (\zeta_{n,t-1} \otimes \varepsilon_t) \right\| = o_p(1)$.

We now consider the sample moment matrix $\sum_{t=1}^n x_t x_t'$. The following result shows that $n^{-1-\alpha} \sum_{t=1}^n x_t x_t'$ has the same probability limit as in the case of a near-stationary regressor generated by a conditionally homoskedastic martingale difference e_t , see Magdalinos and Phillips (2009a). Denote the autocovariance matrix of u_t by $\Gamma_u(j) = \mathbb{E}(u_1 u_{1-j}')$ and the associated long run covariance $\Omega_{uu} = \sum_{j=-\infty}^{\infty} \Gamma_u(j)$. Positive definiteness of Ω_{uu} and the negative-stable property of C imply that the matrix

$$V_C = \int_0^{\infty} e^{rC} \Omega_{uu} e^{rC'} dr \quad (18)$$

is well defined and positive definite.

3.3 Lemma.

- (i) $n^{-1-\alpha} \left\| \sum_{t=1}^n x_{t-1} x_{t-1}' - \sum_{t=1}^n \zeta_{n,t-1} \zeta_{n,t-1}' \right\| = o_p(1)$
- (ii) $n^{-1-\alpha} \sum_{t=1}^n \zeta_{n,t-1} \zeta_{n,t-1}' \rightarrow_p V_C$ where V_C is the matrix in (18).

Lemma 3.4 shows that stationary GARCH effects are eliminated from the first order asymptotics of the denominator of the matrix quotient (11), for a near stationary regressor x_t of arbitrary order. The asymptotic development so far was based on unconditional moment bounds and truncation based on Lemmata 3.1 and 3.2: the GARCH specification (4) has not been employed. Obtaining the limit distribution of the martingale transform in the numerator of the matrix quotient (11), asymptotically equivalent to

$$N_n = \frac{1}{n^{(1+\alpha)/2}} \sum_{t=1}^n (\zeta_{n,t-1} \otimes \varepsilon_t) \quad (19)$$

in view of Lemma 3.2(iv), is more challenging. We show that the predictable quadratic variation of N_n

$$\langle N \rangle_n = \frac{1}{n^{1+\alpha}} \sum_{t=1}^n (\zeta_{n,t-1} \zeta'_{n,t-1} \otimes \mathbb{E}_{\mathcal{F}_{t-1}} \varepsilon_t \varepsilon'_t) = \frac{1}{n^{1+\alpha}} \sum_{t=1}^n (\zeta_{n,t-1} \zeta'_{n,t-1} \otimes H_t) \quad (20)$$

with H_t defined in (4), can be approximated by

$$V_n = \frac{1}{n^{1+\alpha}} \sum_{t=1}^n (\zeta_{n,t-1} \zeta'_{n,t-1} \otimes \Sigma_{\varepsilon\varepsilon}), \quad \Sigma_{\varepsilon\varepsilon} = \mathbb{E} \varepsilon_1 \varepsilon'_1 \quad (21)$$

with approximation error expressed in terms of the solution of a stochastic recurrence relation arising from (4), and that the stability of this solution implies the asymptotic negligibility of the approximation error.

The next result shows how the error of approximating the quadratic variation $\langle N \rangle_n$ in (20) by V_n in (21) can be estimated by using a bound that depends on the solutions of (14) and (15).

3.4 Lemma. *Consider the vector-valued processes*

$$\Sigma_t(j) = \text{vec}(e_t e'_t - \Sigma_{ee}) \otimes \text{vech}(H_{t+j}) \quad (22)$$

$$S_{n,t}(j) = R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes \text{vech}(H_{t+j}) \quad (23)$$

with H_t defined in (4) and

$$w_t = \text{vech}(\varepsilon_t \varepsilon'_t - H_t). \quad (24)$$

- (i) For each $j \geq 1$ and fixed t , $\Sigma_t(j)$ satisfies (14) with innovations $v(j) \equiv v_t(j)$ given by

$$v_t(j) = \text{vec}(e_t e'_t - \Sigma_{ee}) \otimes \varphi + \sum_{l=1}^q (I \otimes A_l) [\text{vec}(e_t e'_t - \Sigma_{ee}) \otimes w_{t+j-l}]. \quad (25)$$

- (ii) For each $j \geq 1$ and fixed t, n , $S_{n,t}(j)$ satisfies (15) with innovations $v_n(j) \equiv v_{n,t}(j)$ given by

$$v_{n,t}(j) = R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes \varphi + \sum_{i=1}^q (I_{r^2} \otimes A_i) [R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes w_{t+j-i}]. \quad (26)$$

- (iii) Given the sequences $\langle N \rangle_n$ and V_n in (20) and (21), the following bound applies:

$$\|\langle N \rangle_n - V_n\| \leq b(\sigma_n + s_n) + o_p(1) \quad (27)$$

as $n \rightarrow \infty$, where

$$\sigma_n = \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} \|R_n\|^{2(j-1)} \left\| \sum_{t=1}^{n-j} \Sigma_t(j) \right\|, \quad s_n = \frac{1}{n^{1+\alpha}} \left\| \sum_{j=1}^{n-1} \sum_{t=2}^{n-j} S_{n,t}(j) \right\| \quad (28)$$

and $b \in (0, \infty)$ is a uniform constant.

By Lemma 3.4, the processes in (22) and (23) can be expressed as the companion form solutions of the stochastic recurrence relations (14) and (15), see (64) and (65) in the Appendix. The leading terms of these solutions consist of "moving averages" of the martingale difference sequences (25) and (26) weighted by powers of the companion matrices M_κ and $M_{n,\kappa}$ respectively. The stability property of the latter allows to employ standard martingale arguments to demonstrate that the bounding sequences σ_n and s_n in (27) are asymptotically negligible.

3.5 Lemma.

- (i) The bounding sequences in (27) satisfy $\sigma_n \rightarrow_p 0$ and $s_n \rightarrow_p 0$.
- (ii) The martingale transform in (19) satisfies the conditional Lindeberg condition

$$\mathcal{L}_n(\delta) = \sum_{t=2}^n \mathbb{E}_{\mathcal{F}_{t-1}} (\xi_{nt}^2 \mathbf{1}\{|\xi_{nt}| > \delta\}) \rightarrow_p 0 \quad \delta > 0$$

$$\text{with } \xi_{nt} = n^{-\frac{1+\alpha}{2}} \|\varepsilon_t\| \|\zeta_{nt-1}\|.$$

Part (i) of the above lemma implies that the predictable quadratic variation of the martingale transform in (19) with ε_t following the vec-GARCH process (4) can be approximated by its counterpart when ε_t is conditionally homoskedastic. Combined with the Lindeberg condition of part (ii), a standard martingale central limit theorem applies to the numerator of the matrix quotient (11), and shows that the asymptotic

variance of the OLS estimator \hat{A}_n is invariant to GARCH effects. The asymptotic distribution of the associated Wald statistic

$$W_n = \left(H \text{vec} \hat{A}_n - h \right)' \left\{ H \left[(\underline{X}' \underline{X})^{-1} \otimes \hat{\Sigma}_{\varepsilon\varepsilon} \right] H' \right\}^{-1} \left(H \text{vec} \hat{A}_n - h \right) \quad (29)$$

for testing linear restrictions on the coefficient matrix

$$H_0 : H \text{vec} (A) = h, \quad (30)$$

where H is a known $q \times mr$ matrix with rank q and h is a known vector, follows directly from that of \hat{A}_n . Since the ε_t sequence is uncorrelated, $\hat{\Sigma}_{\varepsilon\varepsilon}$ in (29) is a simple parametric estimator $\hat{\Sigma}_{\varepsilon\varepsilon} = n^{-1} \sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_t'$ based on the residuals of (1): $\hat{\varepsilon}_t = y_t - \bar{y}_n - \hat{A}_n (x_{t-1} - \bar{x}_{n-1})$. These results are summarised below.

3.6 Theorem. *Consider the system of predictive regressions (1)-(3) with $\alpha \in (0, 1)$, C a negative stable matrix and ε_t, u_t satisfying Assumption INNOV. The following limits apply as $n \rightarrow \infty$:*

- (i) $n^{-(1+\alpha)/2} \sum_{t=1}^n (x_{t-1} \otimes \varepsilon_t) \Rightarrow N(0, V_C \otimes \Sigma_{\varepsilon\varepsilon})$
- (ii) $n^{(1+\alpha)/2} \text{vec}(\hat{A}_n - A) \Rightarrow N(0, V_C^{-1} \otimes \Sigma_{\varepsilon\varepsilon})$
- (iii) $W_n \Rightarrow \chi^2(q)$, under (30)

where q is the rank of H in (30), V_C is defined in (18) and $\Sigma_{\varepsilon\varepsilon} = \mathbb{E} \varepsilon_1 \varepsilon_1'$.

Analogous results apply to the OLS estimator \hat{R}_n of the vector autoregressive process in (2), with martingale difference innovations $u_t = e_t$ that satisfy the vec-GARCH specification (6). The associated Wald statistic for testing $H_0 : H_R \text{vec}(R_n) = h_R$, where H_R is a known $q \times r^2$ matrix with rank q and h_R is a known vector, is given by

$$W_n^R = \left(H_R \text{vec} \hat{R}_n - h_R \right)' \left\{ H_R \left[(X' X)^{-1} \otimes \hat{\Sigma}_{uu} \right] H_R' \right\}^{-1} \left(H_R \text{vec} \hat{R}_n - h_R \right) \quad (31)$$

where $\hat{\Sigma}_{uu} = n^{-1} \sum_{t=1}^n \hat{u}_t \hat{u}_t'$ is based on the (2) residuals $\hat{u}_t = x_t - R_n x_{t-1}$.

3.7 Theorem. *Consider the vector autoregression (2)-(3) with $\alpha \in (0, 1)$, C a negative stable matrix and strictly stationary ergodic innovations $u_t = e_t = \tilde{H}_t^{1/2} \eta_{et}$ with \tilde{H}_t generated by (6). Under Assumption INNOV, the following limits apply as $n \rightarrow \infty$:*

- (i) $n^{-(1+\alpha)/2} \sum_{t=1}^n (x_{t-1} \otimes u_t) \Rightarrow N(0, V_C \otimes \Sigma_{uu})$
- (ii) $n^{(1+\alpha)/2} \text{vec}(\hat{R}_n - R_n) \Rightarrow N(0, V_C^{-1} \otimes \Sigma_{uu})$
- (iii) $W_n^R \Rightarrow \chi^2(q)$,

where q is the rank of H_R and V_C is defined in (18).

3.8 Remarks.

- (i) Theorems 3.6 and 3.7 provide a full characterisation of the effect of GARCH innovations in stochastic regression models by considering regressors with signal that is intermediate to the $O_p(n^2)$ signal of I(1) processes and the $O_p(n)$ signal of I(0) processes. We show that a regression signal of order

$$\sum_{t=1}^n x_{t-1}x'_{t-1} = O_p(nL_n) \quad (32)$$

where $L_n \rightarrow \infty$ at arbitrary rate is sufficient to asymptotically eliminate stationary GARCH effects from the distribution of the least squares estimator and the associated self-normalised test statistics. The implication is that the elimination of GARCH effects from least squares regression asymptotics is not an exclusively I(1) phenomenon: it occurs when the regressors exhibit persistence of any degree, including near-stationary regressors that are arbitrarily close to stationarity. Note that the polynomial rate given to the sequence L_n above serves solely the purpose of facilitating the presentation by employing existing notation of PM (2009) and KMS (2015a): all mathematical arguments carry through trivially by replacing the rate n^α with an arbitrary sequence $L_n \rightarrow \infty$ with $L_n/n \rightarrow 0$.

- (ii) The result has an intuitive signal to noise interpretation: the $O_p(n)$ signal in stationary regression is not sufficiently strong to asymptotically remove the effects of conditional heteroskedasticity in the noise; this only becomes possible when the regression signal is strengthened to (32), while the order of the conditionally heteroskedastic innovations remains I(0).

4. IVX limit theory with GARCH innovations

Having characterised the asymptotic behaviour of the least squares estimator in near stationary systems with conditionally heteroskedastic innovations, we turn to the issue of conducting inference in the predictive regression system (1)-(3) when the order of regressor persistence is unknown. A robust methodology that produces standard inference for testing restrictions on the matrix A of coefficients in (1) across all persistence regimes P(i)-P(iii) based on an endogenous instrumentation procedure, termed IVX, has been proposed by PM (2009) and further developed in the current predictive regression context by KMS (2015a). In this paper, we investigate the extent to which the above procedure is valid under conditionally heteroskedastic innovations.

To fix ideas, instruments are constructed by differencing the regressor x_t and a new process

$$\tilde{z}_t = R_{nz}\tilde{z}_{t-1} + \Delta x_t \quad \tilde{z}_0 = 0 \quad (33)$$

is generated according to an artificial autoregressive matrix

$$R_{nz} = I_r + \frac{C_z}{n^\beta}, \quad \beta \in (0, 1), \quad C_z < 0, \quad (34)$$

with specified persistence degree β . The matrix A of coefficients in (1) is then estimated by a standard instrumental variable estimator that employs the instruments in (33):

$$\tilde{A}_{IVX} = \underline{Y}' \tilde{Z} \left(\underline{X}' \tilde{Z} \right)^{-1}. \quad (35)$$

The asymptotic development of the previous section is the key to the determination of the asymptotic properties of the above IVX estimator under GARCH effects. The asymptotic behaviour of the numerator of the matrix quotient in (35) is driven by the martingale transform

$$\tilde{N}_n = \frac{1}{n^{(1+\alpha\wedge\beta)/2}} \sum_{t=1}^n (\tilde{z}_{t-1} \otimes \varepsilon_t) \quad (36)$$

with instrument process \tilde{z}_t behaving asymptotically like a near stationary process of the type P(ii): when $\beta < \alpha$ (in which case the instruments are less persistent than the regressors) \tilde{z}_{t-1} can be replaced asymptotically in (36) by z_{t-1} , where

$$z_t = R_{nz} z_{t-1} + u_t \quad (37)$$

a n^β -near-stationary process; when $\beta > \alpha$, employing more persistent instruments than the regressor in (2) results to \tilde{z}_t behaving asymptotically as the regressor x_t , a necessarily near-stationary process by the choice of β in (34), in which case \tilde{N}_n in (36) is asymptotically equivalent to N_n in (19); see Lemma 3.1(i) and Lemma 3.5(i)³ of PM (2009) and Lemma B2(iv) of KMS (2015b). We conclude that the limit distribution of \tilde{N}_n in (36) can be derived directly from Theorem 3.7(i). Denote a strictly stationary ergodic version of x_t when $\alpha = 0$ by

$$x_{0,t} = \sum_{j=0}^{\infty} R^j u_{t-j}, \quad R = I_r + C, \quad \|R\| < 1, \quad (38)$$

and the constant matrices

$$V_{C_z} = \int_0^\infty e^{rC_z} \Omega_{uu} e^{rC_z} dr \quad \text{and} \quad \mathbb{V} = \int_0^\infty e^{rC} V_C e^{rC} dr. \quad (39)$$

³In the context of this paper, the statement of Lemma 3.5(i) of PM (2009) is valid for all $\alpha \in (0, \beta)$ since the innovation sequence $u_{0t} \equiv \varepsilon_t$ is a martingale difference.

4.1 Theorem. Consider the model (1)–(3) under Assumption INNOV with instruments \tilde{z}_t defined by (34) and (33). Let B_u be a r -variate Brownian motion with covariance matrix Ω_{uu} , $J_C(t) = \int_0^t e^{C(t-s)} dB_u(s)$ be an Ornstein-Uhlenbeck process and let

$$\underline{B}_u(t) = B_u(t) - \int_0^1 B_u(t) dt, \quad \underline{J}_C(t) = J_C(t) - \int_0^1 J_C(t) dt$$

denote the demeaned versions of B_u and J_C . The following limit theory as $n \rightarrow \infty$ applies for the estimator \tilde{A}_{IVX} in (35):

- (i) when $\beta < \alpha \wedge 1$, $n^{\frac{1+\beta}{2}} \text{vec}(\tilde{A}_{IVX} - A) \Rightarrow MN\left(0, \left(\tilde{\Psi}_{uu}^{-1}\right)' C_z V_{C_z} C_z \tilde{\Psi}_{uu}^{-1} \otimes \Sigma_{\varepsilon\varepsilon}\right)$
- (ii) when $\alpha \in (0, \beta)$, $n^{\frac{1+\alpha}{2}} \text{vec}(\tilde{A}_{IVX} - A) \Rightarrow N\left(0, V_C^{-1} \otimes \Sigma_{\varepsilon\varepsilon}\right)$
- (iii) when $\alpha = \beta > 0$, $n^{\frac{1+\alpha}{2}} \text{vec}(\tilde{A}_{IVX} - A) \Rightarrow N\left(0, \mathbb{V}^{-1} C^{-1} V_C C^{-1} (\mathbb{V}')^{-1} \otimes \Sigma_{\varepsilon\varepsilon}\right)$
- (iv) when $\alpha = 0$, $\sqrt{n} \text{vec}(\tilde{A}_{IVX} - A) \Rightarrow N(0, V_0)$

where $x_{0,t}$ is defined in (38), the matrices V_C , V_{C_z} and \mathbb{V} are defined in (18) and (39),

$$V_0 = \left([\mathbb{E}x_{0,1}x'_{0,1}]^{-1} \otimes I_m \right) \mathbb{E} \left(x_{0,1}x'_{0,1} \otimes \varepsilon_2\varepsilon'_2 \right) \left([\mathbb{E}x_{0,1}x'_{0,1}]^{-1} \otimes I_m \right) \quad (40)$$

and $\tilde{\Psi}_{uu}$ is a random matrix given by $\tilde{\Psi}_{uu} = \Omega_{uu} + \int_0^1 \underline{J}_C dJ'_C$ when $\alpha \geq 1$ with $C = 0$ when $\alpha > 1$, and $\tilde{\Psi}_{uu} = \Omega_{uu} + V_C C$ when $\alpha < 1$.

4.2 Corollary. Under Assumption INNOV, the IVX-Wald statistic

$$\tilde{W}_n = \left(H \text{vec} \tilde{A}_{IVX} - h \right)' \left\{ H \left[(\underline{X}' P_{\underline{Z}} \underline{X})^{-1} \otimes \hat{\Sigma}_{\varepsilon\varepsilon} \right] H' \right\}^{-1} \left(H \text{vec} \tilde{A}_{IVX} - h \right) \quad (41)$$

for testing the hypothesis (30) has a $\chi^2(q)$ asymptotic distribution when $\alpha > 0$.

The only class of predictor variables not covered by Corollary 4.2 is that of purely stationary autoregressions P(iii) with conditionally heteroskedastic innovations. This is by no means surprising since, in the above case, the IVX-Wald test statistic is asymptotically equivalent to a standard OLS-Wald statistic which is known to have a non-standard limit distribution under conditionally heteroskedastic innovations. When x_t is a stationary process and the innovation sequence ε_t in (1) is conditionally heteroskedastic, the asymptotic variance of $n^{-1/2} \sum_{t=1}^n (x_{t-1} \otimes \varepsilon_t)$ is given by $\Upsilon = \mathbb{E} (x_{t-1}x'_{t-1} \otimes \varepsilon_t\varepsilon'_t)$ and does not factorise to $\mathbb{E} (x_{t-1}x'_{t-1}) \otimes \Sigma_{\varepsilon\varepsilon}$ as in the case when ε_t are conditionally homoskedastic; consequently, the matrix $n(\underline{X}'\underline{X})^{-1} \otimes \hat{\Sigma}_{\varepsilon\varepsilon}$ is no longer a consistent estimator of the asymptotic variance of the (asymptotically

equivalent) OLS and IVX estimators, so both the OLS and IVX based Wald statistics will fail to be asymptotically $\chi^2(q)$. The standard limit distribution can be recovered by introducing a White (1980) type of correction in the Wald statistic, which requires consistent estimation of $n\Upsilon$ when x_t is a stationary process. In order to preserve the robustness of the IVX procedure to the persistence properties of x_t , we employ the estimator

$$\hat{\Upsilon}_n = \sum_{t=1}^n (\tilde{z}_{t-1} \tilde{z}'_{t-1} \otimes \hat{\varepsilon}_t \hat{\varepsilon}'_t) \quad (42)$$

where \tilde{z}_t are the IVX instruments in (33) and $\hat{\varepsilon}_t$ are the OLS residuals from (1). The corrected IVX-Wald statistic takes the form

$$\hat{W}_n = \left(H \text{vec} \tilde{A}_{IVX} - h \right)' \left(H \hat{Q}_n H' \right)^{-1} \left(H \text{vec} \tilde{A}_{IVX} - h \right) \quad (43)$$

$$\hat{Q}_n = \left[\left(\tilde{Z}' \underline{X} \right)^{-1} \otimes I_m \right] \hat{\Upsilon}_n \left[\left(\underline{X}' \tilde{Z} \right)^{-1} \otimes I_m \right] \quad (44)$$

The next result characterises the asymptotic behaviour of $\hat{\Upsilon}_n$ in (42) and confirms that it provides an appropriate conditional heteroskedasticity adjustment to the IVX-Wald statistic.

4.3 Lemma. *Under Assumption INNOV, the following hold as $n \rightarrow \infty$:*

- (i) $\max_{1 \leq t \leq n} \mathbb{E} \left\| n^{-(\alpha \wedge \beta)/2} \tilde{z}_t \right\|^4 = O(1)$.
- (ii) $n^{-1-\alpha \wedge \beta} \hat{\Upsilon}_n \rightarrow_p \Phi$, where

$$\Phi = \begin{cases} V_{C_z} \otimes \Sigma_{\varepsilon\varepsilon}, & \beta < \alpha \\ V_C \otimes \Sigma_{\varepsilon\varepsilon}, & 0 < \alpha < \beta \\ \mathbb{E} (x_{0,1} x'_{0,1} \otimes \varepsilon_2 \varepsilon'_2) & \alpha = 0 \end{cases}$$

where $x_{0,t}$ is defined in (38), and the matrices V_C and V_{C_z} are defined in (18) and (39).

4.4 Theorem. *Under Assumption INNOV, the corrected IVX-Wald statistic \hat{W}_n in (43) for testing the hypothesis (30) has a $\chi^2(q)$ asymptotic distribution when $\alpha \geq 0$.*

4.5 Remarks.

- (i) Theorem 4.1 and its corollary show that the limit distribution of the standard IVX-Wald statistic \tilde{W}_n is invariant to the presence of conditional heteroskedasticity in the innovations for all regressors that exhibit some degree of persistence $\alpha > 0$. GARCH effects are present in the limit distribution only in the case where the regressor x_t is a stable autoregressive process with $\alpha = 0$. These

results are a direct consequence of the asymptotic development in Section 3 and the fact that any degree of persistence $\alpha > 0$ is sufficient to eliminate GARCH effects in near stationary systems of regression equations: intuitively, an IVX instrument \tilde{z}_t behaves asymptotically as a near-stationary process (z_t if $\beta < \alpha$ and x_t if $\beta > \alpha > 0$), the martingale transform \tilde{N}_n in (36) will behave asymptotically as its near stationary counterpart (19), and will thus have sufficient signal to eliminate GARCH effects from the limit distribution. For the same reason, the limit distribution of the standard IVX-Wald test is distorted by the presence of GARCH innovations when $\alpha = 0$, since \tilde{z}_t behaves like the stationary process x_t .

- (ii) Theorem 4.4 shows that a simple adjustment to the IVX Wald test statistic extends the validity of the IVX approach in the presence of GARCH innovations across the whole range $\alpha \geq 0$ of data generating mechanisms considered in classes P(i)-P(iii). These classes define regressors with diverse stochastic properties, ranging from pure stationarity to unit root nonstationarity and include the intermediate local to unity and near stationary persistence regimes. The adjustment differs from a standard conditional heteroskedasticity correction in that $\mathbb{E}(x_{t-1}x'_{t-1} \otimes \varepsilon_t\varepsilon'_t)$ is estimated by using the IVX instruments instead of the regressors, in order to ensure the robustness of the corrected IVX-Wald statistic in (43) to regressors with degree of persistence $\alpha > \beta$.
- (iii) KMS (2015a) have proposed a finite sample correction to the IVX Wald test statistic \tilde{W}_n in (41) that exhibits better finite sample properties while being asymptotically equivalent to \tilde{W}_n . The conditional heteroskedasticity adjustment employed to \tilde{W}_n can also be employed to the IVX-Wald test statistic of KMS (2015a), resulting to the adjusted version having a $\chi^2(q)$ limit distribution for all $\alpha \geq 0$.

5. Discussion

The paper provides a complete characterisation of the asymptotic properties of least squares regression methods in the presence of conditional heteroskedasticity in the innovations that take the form of a covariance stationary vec-GARCH process. Existing results on stochastic regression with conditionally heteroskedastic innovations lead to different conclusions depending on the integration properties of the regressors. Least squares limit theory with I(1) processes is invariant to the presence of conditional heteroskedasticity and the usual Dickey-Fuller type of limit distributions apply. On the other hand, GARCH effects appear in the first order asymptotics of the OLS estimator and the associated self-normalised statistics generated by I(0) regressors. Approached as a signal-to-noise ratio problem, a natural question that arises is the degree of regression signal required in order to asymptotically eliminate conditional

heteroskedasticity from the noise. The paper provides a simple and intuitive answer: any signal that dominates the $O_p^e(n)$ signal of a stationary regressor is sufficient. Consequently, GARCH effects appear in least squares limit theory only in the case of stationary regressors: for near-stationary and local to unity regressors, the OLS estimator has same the limit distribution that applies under conditionally homoskedastic innovations, given in Magdalinos and Phillips (2009) and Phillips (1988) respectively.

The asymptotic invariance of least squares methods to GARCH effects in the innovations in the case of regressors that are not exactly $I(0)$ carries over to the IVX procedure of PM (2009) and KMS (2015a), where the IVX-Wald test statistic is shown to have a standard chi-squared limit distribution. The advantage of this method is that, unlike least squares, the limit distribution is robust to regressor persistence. To accommodate $I(0)$ regressors in the presence of conditional heteroskedasticity in the innovations, we introduce a White-type correction based on the endogenously generated IVX instruments rather than the regressor in order preserve the method's robustness property. This adjusted IVX-Wald test statistic is shown to have a standard chi-squared limit distribution under all persistence regimes and stationary GARCH innovations, validating the IVX procedure under conditional heteroskedasticity.

6. Technical appendix and proofs

We denote by $\|M\| = \max \left\{ \sqrt{\lambda} : \lambda \in \sigma(M'M) \right\}$ and $\|M\|_F = (tr M'M)^{1/2}$ the spectral and Frobenius matrix norms and by $\sigma(A)$ and $\rho(A)$ the spectrum and the spectral radius of a square matrix A .

Proof of Lemma 3.1. It is sufficient to show that all non-zero eigenvalues of M_κ lie inside the open unit disk $\{z \in \mathbb{C} : |z| < 1\}$. Suppose that $\lambda \in \mathbb{C} \setminus \{0\}$ is an arbitrary eigenvalue of M_κ . Letting

$$G_\kappa(\lambda) = I_s - \frac{1}{\lambda} \Gamma_1 - \dots - \frac{1}{\lambda^\kappa} \Gamma_\kappa,$$

with $s = r^2 m(m+1)/2$, and using the standard formula for the determinant of a partitioned matrix (e.g. 5.30 of Abadir and Magnus (2005)) and induction on κ we obtain

$$\det(M_\kappa - \lambda I_{\kappa s}) = (-\lambda)^{\kappa s} \det G_\kappa(\lambda). \quad (45)$$

The identity (45) implies that any non-zero eigenvalue λ of M_κ satisfies

$$\det G_\kappa(\lambda) = 0. \quad (46)$$

Denoting by \bar{M} the conjugate transpose of a square complex matrix M , the real (also called Hermitian) part of $G_\kappa(\lambda)$ is given by

$$\begin{aligned}
\mathcal{R}[G_\kappa(\lambda)] &= \frac{1}{2} [G_\kappa(\lambda) + \bar{G}_\kappa(\lambda)] \\
&= \frac{1}{2} \left[2I_s - \left(\frac{1}{\lambda} + \frac{1}{\bar{\lambda}} \right) \Gamma_1 - \dots - \left(\frac{1}{\lambda^\kappa} + \frac{1}{\bar{\lambda}^\kappa} \right) \Gamma_\kappa \right] \\
&= \frac{1}{2} \left[2I_s - \frac{2 \operatorname{Re}(\lambda)}{|\lambda|^2} \Gamma_1 - \frac{2 \operatorname{Re}(\lambda^2)}{|\lambda|^4} \Gamma_2 - \dots - \frac{2 \operatorname{Re}(\lambda^\kappa)}{|\lambda|^{2\kappa}} \Gamma_\kappa \right] \\
&= I_m - \sum_{i=1}^{\kappa} \Gamma_i + \sum_{j=1}^{\kappa} \left[1 - \frac{\operatorname{Re}(\lambda^j)}{|\lambda|^{2j}} \right] \Gamma_j. \tag{47}
\end{aligned}$$

Note that $\mathcal{R}[G_\kappa(\lambda)]$ is a real symmetric matrix. The conditions $\Gamma_i \geq 0$ and $\rho_0 < 1$ imply that the matrix $I_m - \sum_{i=1}^{\kappa} \Gamma_i$ is positive definite: ρ_0 is the largest eigenvalue of $\sum_{i=1}^{\kappa} \Gamma_i$ (since $\sum_{i=1}^{\kappa} \Gamma_i \geq 0$), so $1 - \rho_0$ is the smallest eigenvalue of $I_m - \sum_{i=1}^{\kappa} \Gamma_i$. Moreover, for arbitrary $\lambda \in \mathbb{C} \setminus \{0\}$ and $j \in \mathbb{N}$,

$$|\lambda| \geq 1 \implies |\lambda|^{2j} \geq |\lambda|^j = |\lambda^j| \geq |\operatorname{Re}(\lambda^j)| \implies 1 - \frac{\operatorname{Re}(\lambda^j)}{|\lambda|^{2j}} \geq 0$$

which implies that the second sum on the right of (47) is a positive semidefinite matrix. Since $I_m - \sum_{i=1}^{\kappa} \Gamma_i > 0$, (47) implies that $\mathcal{R}[G_\kappa(\lambda)]$ is a positive definite matrix for all $\lambda \in \mathbb{C}$ satisfying $|\lambda| \geq 1$. Positivity of $\mathcal{R}[G_q(\lambda)]$ implies the inequality

$$|\det[G_q(\lambda)]| \geq \det \mathcal{R}[G_q(\lambda)] > 0$$

for all $\lambda \in \mathbb{C}$ satisfying $|\lambda| \geq 1$ (see e.g. Exercise 11(b), page 106 of Serre (2010)). We conclude that (46) is violated when $|\lambda| \geq 1$, so M_κ cannot have an eigenvalue with $|\lambda| \geq 1$. Hence, $\rho(M_\kappa) < 1$.

To show the second assertion, denote by $(M_\kappa^j)_{kl}$ the kl element of the matrix M_κ^j for $j \geq 1$. By Corollary 5.6.13 of Horn and Johnson (2013) for arbitrary $\delta > 0$ there exists $b(\delta) > 0$ such that

$$\max_{1 \leq k \leq \kappa s} \max_{1 \leq l \leq \kappa s} |(M_\kappa^j)_{kl}| \leq b(\delta) [\rho(M_\kappa) + \delta]^j \quad \text{for all } j \geq 1.$$

Since $\rho(M_\kappa) < 1$, we may choose $\delta \in (0, 1 - \rho(M_\kappa))$, which implies that $\lambda_\delta := \rho(M_\kappa) + \delta \in (0, 1)$. For this choice of δ ,

$$\|M_\kappa^j\|_F \leq \kappa s \max_{1 \leq k \leq \kappa s} \max_{1 \leq l \leq \kappa s} |(M_\kappa^j)_{kl}| \leq \kappa s b(\delta) \lambda_\delta^j$$

giving

$$\sum_{j=0}^{\infty} \|M_\kappa^j\|_F \leq \kappa s b(\delta) \sum_{j=0}^{\infty} \lambda_\delta^j = \frac{\kappa s b(\delta)}{1 - \lambda_\delta}.$$

This shows part (i). For part (ii), $M_{n,\kappa} \rightarrow M_\kappa$ as $n \rightarrow \infty$ so, by continuity of the eigenvalues of a matrix as a function of the matrix elements, $\rho(M_{n,\kappa}) \rightarrow \rho(M_\kappa)$ and $\|M_{n,\kappa}\| \rightarrow \|M_\kappa\|$ as $n \rightarrow \infty$. Since $\rho(M_\kappa) < 1$, convergence of $\rho(M_{n,\kappa})$ implies that $\rho(M_{n,\kappa}) < 1$ for all but finitely many n . Also, since $\sum_{j=0}^{\infty} \|M_\kappa^j\| < \infty$, $\|M_{n,\kappa}\| \rightarrow \|M_\kappa\|$ and dominated convergence yield

$$\lim_{n \rightarrow \infty} \sum_{j=0}^{\infty} \|M_{n,\kappa}^j\| = \sum_{j=0}^{\infty} \|M_\kappa^j\| < \infty$$

which implies that $\sup_{n \geq 1} \sum_{j=0}^{\infty} \|M_{n,\kappa}^j\| < \infty$ (since convergent real sequences are bounded).

Proof of Lemma 3.2. For part (i), denoting $F = C(1)$ for brevity, the identity $\|x\|^4 = \text{tr}\{(x \otimes x)(x' \otimes x')\}$ for any vector x , yields

$$\begin{aligned} & \mathbb{E} \left(\frac{\|\zeta_{n,t}\|^4}{n^{\alpha/2}} \right) \\ &= \frac{1}{n^{2a}} \text{tr} \left\{ \mathbb{E} (\zeta_{n,t} \otimes \zeta_{n,t}) (\zeta'_{n,t} \otimes \zeta'_{n,t}) \right\} \\ &= \frac{1}{n^{2a}} \text{tr} \sum_{i,j,k,l=1}^t (R_n^{t-j} F \otimes R_n^{t-i} F) \mathbb{E} (e_j e'_k \otimes e_i e'_l) (F' R_n^{t-k'} \otimes F' R_n^{t-l'}) \\ &= \frac{1}{n^{2a}} \sum_{i,j,k,l=1}^t \mathbb{E} \text{tr} \left[(F' R_n^{t-k'} R_n^{t-j} F \otimes F' R_n^{t-l'} R_n^{t-i} F) (e_j e'_k \otimes e_i e'_l) \right] \\ &= \frac{1}{n^{2a}} \sum_{i,j,k,l=1}^t [\text{vec}(F' R_n^{t-j'} R_n^{t-k} F \otimes F' R_n^{t-i'} R_n^{t-l} F)]' \mathbb{E} \text{vec}(e_j e'_k \otimes e_i e'_l) \quad (48) \end{aligned}$$

by using the identity $\text{tr}(A'B) = [\text{vec}(A)]' \text{vec}(B)$. Since, for any $r \times r$ matrices K, L the $r^4 \times 1$ vectors $\text{vec}(K \otimes L)$ and $\text{vec}(K) \otimes \text{vec}(L)$ consist of the same elements $\{K_{ij}L_{kl} : 1 \leq i, j, k, l \leq r\}$ in different order of appearance, there exists a $r^4 \times r^4$ permutation matrix Π such that $\text{vec}(K \otimes L) = \Pi [\text{vec}(K) \otimes \text{vec}(L)]$, giving

$$\text{vec}(e_j e'_k \otimes e_i e'_l) = \Pi [\text{vec}(e_j e'_k) \otimes \text{vec}(e_i e'_l)] = \Pi (e_k \otimes e_j \otimes e_l \otimes e_i). \quad (49)$$

The Kronecker product on the right of (49) is a vector that consists of the same elements (in different order of appearance) for any reordering of the indices k, j, l, i ; since the product of two permutation matrices is a permutation matrix, we can rearrange the order $\{e_k, e_j, e_l, e_i\}$ in the Kronecker (49) by changing the permutation matrix Π . By the Cauchy-Schwarz inequality and equivalence of norms in finite dimensional spaces, we obtain that

$$\begin{aligned} |[\text{vec}(A)]' \text{vec}(B)| &\leq \|\text{vec}(A)\| \|\text{vec}(B)\| = \|A\|_F \|\text{vec}(B)\| \\ &\leq b \|A\| \|\text{vec}(B)\| \end{aligned}$$

for all matrices A, B of the same order and some finite constant $b > 0$. Applying the above bound to (48), and using the fact that $\|\Pi\| = 1$:

$$\begin{aligned} \mathbb{E} \left(\frac{\|\zeta_{n,t}\|^4}{n^{\alpha/2}} \right) &\leq \frac{1}{n^{2a}} \sum_{i,j,k,l=1}^t \|F' R_n^{t-j'} R_n^{t-k} F\|_F \|F' R_n^{t-i'} R_n^{t-l} F\|_F \|\mathbb{E}(e_k \otimes e_j \otimes e_l \otimes e_i)\| \\ &\leq \frac{b \|F\|^4}{n^{2a}} \sum_{i,j,k,l=1}^t \rho_n^{t-j} \rho_n^{t-i} \rho_n^{t-k} \rho_n^{t-l} \|\mathbb{E}(e_k \otimes e_j \otimes e_l \otimes e_i)\| \end{aligned} \quad (50)$$

where $\rho_n = \|R_n\|$ and b is a uniform bounding constant (taking possibly different values).

The martingale difference property of e_t implies that $\mathbb{E}(e_k \otimes e_j \otimes e_l \otimes e_i) = 0$ when

$$\max\{i, j, k, l\} > \max\{\{i, j, k, l\} \setminus \max\{i, j, k, l\}\},$$

so the expectation in (50) is non-zero in the following cases: (i) all elements of $\{i, j, k, l\}$ are equal; (ii) three elements of $\{i, j, k, l\}$ are equal and strictly greater than the remaining element; (iii) the elements of $\{i, j, k, l\}$ are pairwise equal; (iv) two elements of $\{i, j, k, l\}$ are equal and strictly greater than the remaining two unequal elements. In cases (i)-(iii), the result is immediate since there are at most 2 sums in (50). For example, in case (ii), the right side of (50) is bounded by

$$\begin{aligned} &\frac{\|F\|^4}{n^{2a}} \left(\sum_{j=1}^{\infty} \rho_n^j \right) \left(\sum_{k=1}^{\infty} \rho_n^{3k} \right) \max_{i,l \geq 1} \mathbb{E}(\|e_l\| \|e_i\|^3) \\ &\leq O(1) \max_{i,l \geq 1} [\mathbb{E}(\|e_l\|^2 \|e_i\|^2)]^{1/2} (\mathbb{E} \|e_1\|^4)^{1/2} \\ &\leq O(1) \mathbb{E} \|e_1\|^4 \end{aligned}$$

because $\sum_{i=1}^{\infty} \rho_n^{mi} = O(n^\alpha)$ for any fixed m . In case (iv) there will be 3 sums in (50), so the above bound is too crude. In this case, we may write expectation in (50)

$$\mathbb{E}(e_k \otimes e_j \otimes e_l \otimes e_i) = \mathbb{E}(e_k \otimes e_j \otimes e_i \otimes e_i), \quad i > j > k$$

without loss of generality (by varying the permutation matrix Π in (49) and noting that $\|\Pi\| = 1$), so (50) yields

$$\begin{aligned} \mathbb{E} \left(\frac{\|\zeta_{n,t}\|^4}{n^{\alpha/2}} \right) &\leq \frac{b \|F\|^4}{n^{2a}} \sum_{i=1}^t \rho_n^{2(t-i)} \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} \rho_n^{t-j} \rho_n^{t-k} \|\mathbb{E}(e_k \otimes e_j \otimes \text{vec}(e_i e_i'))\| \\ &\leq \frac{b \|D_r\| \|F\|^4}{n^{2a}} \sum_{i=1}^t \rho_n^{2(t-i)} \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} \rho_n^{t-j} \rho_n^{t-k} \|Y_{k,j}(i)\| \end{aligned} \quad (51)$$

where D_r denotes the $r^2 \times r(r+1)/2$ duplication matrix (Chapter 11 of Abadir and Magnus, 2005) that satisfies $\text{vec}(K) = D_r \text{vech}(K)$ for a symmetric $r \times r$ matrix K and

$$Y_{k,j}(i) := \mathbb{E}(e_k \otimes e_j \otimes \text{vech}(e_i e_i')) = \mathbb{E}\left(e_k \otimes e_j \otimes \text{vech}\left(\tilde{H}_i\right)\right)$$

since $e_t = \tilde{H}_t^{1/2} \eta_{et}$. Applying (6) and recalling that $\mathbb{E}(e_k \otimes e_j) = 0$ since $k < j$, we obtain

$$\begin{aligned} Y_{k,j}(i) &= \sum_{l=1}^{\tilde{q} \wedge (i-j-1)} \left(I_{r_2} \otimes \tilde{A}_l\right) \mathbb{E}\left[e_k \otimes e_j \otimes \text{vech}\left(\tilde{H}_{i-l}\right)\right] \\ &+ \sum_{k=1}^{\tilde{p} \wedge (i-j-1)} \left(I_{r_2} \otimes \tilde{B}_k\right) \mathbb{E}\left[e_k \otimes e_j \otimes \text{vech}\left(\tilde{H}_{i-k}\right)\right] \end{aligned}$$

because the law of iterated expectations gives

$$\begin{aligned} \mathbb{E}\left[e_k \otimes e_j \otimes \text{vech}\left(e_{i-l} e_{i-l}'\right)\right] &= \mathbb{E}\left[e_k \otimes e_j \otimes \text{vech} \mathbb{E}_{\mathcal{F}_{i-l-1}}\left(e_{i-l} e_{i-l}'\right)\right] \\ &= \mathbb{E}\left[e_k \otimes e_j \otimes \text{vech}\left(\tilde{H}_{i-l}\right)\right] \end{aligned}$$

the above expectation is non-zero only if $i-l > j$ and $i-k > j$. We conclude that, for fixed k, j , $Y_{k,j}(i)$ satisfies the recurrence relation

$$Y_{k,j}(i) = \sum_{l=1}^{\tilde{\kappa}} \tilde{\Gamma}_l Y_{k,j}(i-l), \text{ for } i > j; Y_{k,j}(i) = 0 \text{ for } i \leq j \quad (52)$$

where $\tilde{\Gamma}_l = I_{r_2} \otimes \tilde{C}_l$ where \tilde{C}_l is defined in (12) with A_l replaced by \tilde{A}_l and B_l replaced by \tilde{B}_l and $\tilde{\kappa} = \tilde{q} \vee \tilde{p}$. Letting

$$\tilde{Y}_{k,j}(i) = [Y_{k,j}(i)', Y_{k,j}(i-1)', \dots, Y_{k,j}(i)(i - \tilde{\kappa} + 1)']'$$

we can write (52) in companion form:

$$\tilde{Y}_{k,j}(i) = \tilde{M}_\kappa \tilde{Y}_{k,j}(i-1) \quad i \geq j$$

where \tilde{M}_κ is the matrix M_κ defined in (16) with Γ_i replaced by $\tilde{\Gamma}_i$. Recursion yields

$$\tilde{Y}_{k,j}(i) = \tilde{M}_\kappa^{i-j-1} \tilde{Y}_{k,j}(j+1) = \tilde{M}_\kappa^{i-j-1} Y_{k,j}(j+1)$$

since $Y_{k,j}(l) = 0$ for all $l \leq j$. We conclude that

$$\begin{aligned}
\|Y_{k,j}(i)\| &\leq \|\tilde{Y}_{k,j}(i)\| \leq \|\tilde{M}_\kappa^{i-j-1}\| \|Y_{k,j}(j+1)\| \\
&= \|\tilde{M}_\kappa^{i-j-1}\| \left\| \mathbb{E} \left(e_k \otimes e_j \otimes \text{vech} \left(\tilde{H}_{j+1} \right) \right) \right\| \\
&= \|\tilde{M}_\kappa^{i-j-1}\| \left\| \mathbb{E} \left(e_k \otimes e_j \otimes \text{vech} \mathbb{E}_{\mathcal{F}_j} (e_{j+1} e'_{j+1}) \right) \right\| \\
&= \|\tilde{M}_\kappa^{i-j-1}\| \left\| \mathbb{E} \left(e_k \otimes e_j \otimes \text{vech} (e_{j+1} e'_{j+1}) \right) \right\| \\
&\leq b \|\tilde{M}_\kappa^{i-j-1}\| \mathbb{E} (\|e_k\| \|e_j\| \|e_{j+1}\|^2) \\
&\leq b \|\tilde{M}_\kappa^{i-j-1}\| \left\{ \mathbb{E} (\|e_k\|^2 \|e_{j+1}\|^2) \mathbb{E} (\|e_j\|^2 \|e_{j+1}\|^2) \right\}^{1/2} \\
&\leq b \|\tilde{M}_\kappa^{i-j-1}\| \mathbb{E} \|e_1\|^4
\end{aligned}$$

and substituting into (51) yields:

$$\begin{aligned}
\mathbb{E} \left(\frac{\|\zeta_{n,t}\|^4}{n^{\alpha/2}} \right) &\leq \frac{b}{n^{2a}} \sum_{i=1}^t \rho_n^{2(t-i)} \sum_{j=1}^{i-1} \rho_n^{t-j} \|\tilde{M}_\kappa^{i-j-1}\| \sum_{k=1}^{j-1} \rho_n^{t-k} \mathbb{E} \|e_1\|^4 \\
&\leq \frac{b}{n^{2a}} \sum_{i=1}^t \rho_n^{2(t-i)} \sum_{j=1}^{i-1} \|\tilde{M}_\kappa^{i-j-1}\| \rho_n^{t-j+1} \sum_{k=1}^{j-1} \rho_n^{j-1-k} \mathbb{E} \|e_1\|^4 \\
&\leq \frac{b}{n^{2a}} \sum_{i=0}^{\infty} \rho_n^{2i} \sum_{k=1}^{\infty} \rho_n^k \sum_{j=1}^{\infty} \|\tilde{M}_\kappa^j\| \mathbb{E} \|e_1\|^4 \\
&= O(1)
\end{aligned}$$

uniformly in t . This shows part (i).

For part (ii), since for any $\delta > 0$

$$\mathbb{P} \left(\max_{1 \leq t \leq n} \|\zeta_{n,t}\| > \delta L_n^{-1} n^{\frac{1+\alpha}{2}} \right) \leq \frac{1}{\delta^2} L_n^2 \mathbb{E} \left[\frac{1}{n^{1+\alpha}} \max_{1 \leq t \leq n} \|\zeta_{n,t}\|^2 \right]$$

and $L_n \rightarrow \infty$ at arbitrarily slow rate, it is sufficient to show that

$$\frac{1}{n^{1+\alpha}} \mathbb{E} \left[\max_{1 \leq t \leq n} \|\zeta_{n,t}\|^2 \right] \rightarrow 0. \tag{53}$$

To prove (53), letting

$$\tilde{e}_{n,j} = e_j \mathbf{1} \{ \|e_j\| > m_n \}$$

for an arbitrary sequence $m_n \rightarrow \infty$, we can write

$$\begin{aligned}
\|\zeta_{n,t}\|^2 &= \left\| \sum_{j=1}^t R_n^{t-j} C(1) e_j \right\|^2 \\
&\leq \|C(1)\|^2 \left(\sum_{j=1}^t \|R_n\|^{t-j} \|e_j\| \right)^2 \\
&= \|C(1)\|^2 \left(\sum_{j=1}^t \|R_n\|^{t-j} [\|e_j\| \mathbf{1}\{\|e_j\| \leq m_n\} + \|\tilde{e}_{n,j}\|] \right)^2 \\
&\leq \|C(1)\|^2 \left(b_1 m_n n^\alpha + \sum_{j=1}^t \|R_n\|^{t-j} \|\tilde{e}_{n,j}\| \right)^2 \\
&\leq 2 \|C(1)\|^2 \left\{ b_1^2 m_n^2 n^{2\alpha} + \left(\sum_{j=1}^t \|R_n\|^{t-j} \|\tilde{e}_{n,j}\| \right)^2 \right\} \\
&\leq 2 \|C(1)\|^2 \left\{ b_1^2 m_n^2 n^{2\alpha} + \sum_{j=1}^t \|R_n\|^{2(t-j)} \sum_{j=1}^t \|\tilde{e}_{n,j}\|^2 \right\} \\
&\leq 2 \|C(1)\|^2 \left\{ b_1^2 m_n^2 n^{2\alpha} + b_2 n^\alpha \sum_{j=1}^n \|\tilde{e}_{n,j}\|^2 \right\}
\end{aligned}$$

for fixed constants $b_1, b_2 \in (0, \infty)$. Choosing $m_n \rightarrow \infty$ such that $m_n^2/n^{1-\alpha} \rightarrow 0$ and taking expectations yields

$$\frac{1}{n^{1+\alpha}} \mathbb{E} \left[\max_{1 \leq t \leq n} \|\zeta_{n,t}\|^2 \right] \leq O\left(\frac{m_n^2}{n^{1-\alpha}}\right) + b_2 \mathbb{E} (\|e_1\|^2 \mathbf{1}\{\|e_1\| > m_n\}) \rightarrow 0$$

showing (53). For part (iii), the martingale property and part (i) give

$$\begin{aligned}
\mathbb{E} \left\| \frac{1}{n} \sum_{t=1}^n (\zeta_{n,t-1} \otimes e_t) \right\|^2 &= \frac{1}{n^2} \sum_{t=1}^n \mathbb{E} \left(\|\zeta_{n,t-1}\|^2 \|e_t\|^2 \right) \\
&\leq \frac{1}{n} \left[\max_{t \leq n} \mathbb{E} \left(\|\zeta_{n,t-1}\|^4 \right) \right]^{1/2} [\mathbb{E} (\|e_1\|^4)]^{1/2} \\
&= O\left(\frac{1}{n^{1-\alpha}}\right).
\end{aligned}$$

For part (iv), $x_0 = o_p(n^{\alpha/2})$ and the BN decomposition gives

$$\begin{aligned} \frac{1}{n^{(1+\alpha)/2}} \sum_{t=1}^n (x_{t-1} \otimes \varepsilon_t) &= \frac{1}{n^{(1+\alpha)/2}} \sum_{t=1}^n \left(\sum_{j=0}^{t-2} R_n^j u_{t-j-1} \otimes \varepsilon_t \right) + o_p\left(\frac{1}{n^{(1-\alpha)/2}}\right) \\ &= \frac{1}{n^{(1+\alpha)/2}} \sum_{t=1}^n \left[\left(\zeta_{n,t-1} - \sum_{j=0}^{t-2} R_n^j \Delta \tilde{e}_{t-j-1} \right) \otimes \varepsilon_t \right] + o_p\left(\frac{1}{n^{(1-\alpha)/2}}\right) \end{aligned}$$

Since $\sum_{j=0}^{t-2} R_n^j \Delta \tilde{e}_{t-j-1} = R_n^{t+1} \tilde{e}_1 - \tilde{e}_{t-2} - \sum_{j=0}^{t-2} (\Delta R_n^{j+1}) \tilde{e}_{t-j-1}$, and

$$\left\| n^{-(1+\alpha)/2} \sum_{t=1}^n (R_n^{t+1} \tilde{e}_1 \otimes \varepsilon_t) \right\|_{L_1} = O(n^{-(1-\alpha)/2})$$

by the Cauchy-Schwarz inequality, it is enough to show that the martingale arrays

$$m_{1n} = \frac{1}{n^{(1+\alpha)/2}} \sum_{t=1}^n (\tilde{e}_{t-2} \otimes \varepsilon_t), \quad m_{2n} = \frac{1}{n^{(1+\alpha)/2}} \frac{1}{n^\alpha} \sum_{t=1}^n \left[\left(\sum_{j=0}^{t-2} R_n^j \tilde{e}_{t-j-1} \right) \otimes \varepsilon_t \right]$$

are both $o_p(1)$. The Cauchy-Schwarz inequality gives

$$\mathbb{E} \|m_{1n}\|^2 = \frac{1}{n^{1+\alpha}} \sum_{t=1}^n \mathbb{E} (\|\tilde{e}_{t-2}\|^2 \|\varepsilon_t\|^2) \leq \frac{1}{n^\alpha} \{ \mathbb{E} \|\tilde{e}_1\|^4 \mathbb{E} \|\varepsilon_1\|^4 \}^{1/2}$$

and the Cauchy-Schwarz inequality followed by the Minkowski inequality gives

$$\begin{aligned} \mathbb{E} \|m_{2n}\|^2 &\leq \frac{1}{n^{1+3\alpha}} \sum_{t=1}^n \left\{ \mathbb{E} \left\| \sum_{j=0}^{t-2} R_n^j \tilde{e}_{t-j-1} \right\|^4 \right\}^{1/2} \{ \mathbb{E} \|\varepsilon_1\|^4 \}^{1/2} \\ &\leq \frac{\|\varepsilon_1\|_{L_4}^2}{n^{1+3\alpha}} \sum_{t=1}^n \left\{ \left(\sum_{j=0}^{t-2} \{ \|R_n\|^{4j} \mathbb{E} \|\tilde{e}_1\|^4 \}^{1/4} \right)^4 \right\}^{1/2} \\ &\leq \frac{\|\varepsilon_1\|_{L_4}^2 \|\tilde{e}_1\|_{L_4}^2}{n^{3\alpha}} \left(\sum_{j=0}^{n-2} \|R_n\|^j \right)^2 = O\left(\frac{1}{n^\alpha}\right). \end{aligned}$$

Proof of Lemma 3.3. To prove the approximation in (i), we use the Phillips and Solo (1992) method of applying the BN decomposition $u_t = C(1)e_t - \Delta \tilde{e}_t$ and the summation by parts formula to ζ_{nt-1} to obtain

$$x_{t-1} = \zeta_{nt-1} + \tilde{w}_{n,t} + R_n^{t-1} x_0 \tag{54}$$

with $\tilde{w}_{n,t} = \tilde{e}_t - R_n^t \tilde{e}_1 - \frac{C}{n^\alpha} \sum_{j=1}^{t-1} R_n^{j-1} \tilde{e}_{t-j}$ and \tilde{e}_t defined in (8). First note that, by the Minkowski inequality,

$$\begin{aligned}
\sup_{t \geq 1} \mathbb{E} \|\tilde{w}_{n,t}\|^2 &\leq 2 \left[2 \sup_{t \geq 1} \mathbb{E} \|\tilde{e}_t\|^2 + \frac{\|C\|^2}{n^{2\alpha}} \sup_{t \geq 1} \mathbb{E} \left(\sum_{j=1}^{t-1} \|R_n\|^{j-1} \|\tilde{e}_{t-j}\| \right)^2 \right] \\
&\leq 4 \mathbb{E} \|\tilde{e}_1\|^2 + \frac{2\|C\|^2}{n^{2\alpha}} \sup_{t \geq 1} \left(\sum_{j=1}^{t-1} \left\{ \|R_n\|^{2(j-1)} \mathbb{E} \|\tilde{e}_{t-j}\|^2 \right\}^{1/2} \right)^2 \\
&\leq 4 \mathbb{E} \|\tilde{e}_1\|^2 + 2 \|C\|^2 \mathbb{E} \|\tilde{e}_1\|^2 \left(\frac{1}{n^\alpha} \sum_{j=1}^{\infty} \|R_n\|^{(j-1)} \right)^2 \\
&= \mathbb{E} (\|\tilde{e}_1\|^2) O(1) = O(1)
\end{aligned}$$

by stationarity of \tilde{e}_t and the fact that $\mathbb{E} \|\tilde{e}_1\|^2 < \infty$ and $\sum_{j=1}^{\infty} \|R_n\|^{j-1} = O(n^\alpha)$. Expanding (54) using the above bound and $x_0 = o_p(n^{\alpha/2})$,

$$\left\| \sum_{t=1}^n x_{t-1} x'_{t-1} - \sum_{t=1}^n \zeta_{nt-1} \zeta'_{nt-1} \right\| \leq 2 \sum_{t=1}^n \|\zeta_{nt-1}\| \|\tilde{w}_{n,t}\| + \sum_{t=1}^n \|\tilde{w}_{n,t}\|^2 + o_p(n^{2\alpha})$$

and part (i) follows since $n^{-1-\alpha} \mathbb{E} \sum_{t=1}^n \|\tilde{w}_{n,t}\|^2 = O(n^{-\alpha})$ and

$$\begin{aligned}
\frac{1}{n^{1+\alpha}} \sum_{t=1}^n \mathbb{E} \|\zeta_{nt-1}\| \|\tilde{w}_{n,t}\| &\leq \frac{1}{n^\alpha} \left(\max_{t \leq n} \mathbb{E} \|\zeta_{nt-1}\|^2 \right)^{1/2} \left(\sup_{t \geq 1} \mathbb{E} \|\tilde{w}_{n,t}\|^2 \right)^{1/2} \\
&= O\left(\frac{1}{n^{\alpha/2}}\right)
\end{aligned}$$

For part (ii), write

$$\frac{1}{n^{1+\alpha}} \sum_{t=1}^n \zeta_{nt-1} \zeta'_{nt-1} = A_n + B_n + B'_n$$

where

$$\begin{aligned}
A_n &= \frac{1}{n^{1+\alpha}} \sum_{t=1}^n \sum_{j=0}^{t-1} R_n^j C(1) e_{t-j} e'_{t-j} C(1)' (R_n^j)' \\
B_n &= \frac{1}{n^{1+\alpha}} \sum_{t=1}^n \sum_{i=0}^{t-1} \sum_{j=i+1}^{t-1} R_n^j C(1) e_{t-j} e'_{t-i} C(1)' (R_n^i)' .
\end{aligned}$$

Noting that $\Omega_{uu} = C(1)\Sigma_{ee}C(1)'$ and

$$\begin{aligned}
A_n &= \frac{1}{n^{1+\alpha}} \sum_{j=0}^{n-1} R_n^j C(1) \sum_{t=j+1}^n e_{t-j} e'_{t-j} C(1)' (R_n^j)' \\
&= \frac{1}{n^{1+\alpha}} \sum_{j=0}^{n-1} R_n^j C(1) \sum_{t=1}^{n-j} (e_t e'_t - \Sigma_{ee}) C(1)' (R_n^j)' \\
&\quad + \frac{1}{n^{1+\alpha}} \sum_{j=0}^{n-1} (n-j) R_n^j \Omega_{uu} (R_n^j)' \\
&= A_{1n} + A_{2n}.
\end{aligned}$$

The ergodic theorem yields

$$\begin{aligned}
\|A_{1n}\| &\leq \frac{\|C(1)\|^2}{n^\alpha} \sum_{j=0}^{n-1} \|R_n\|^{2j} \max_{m \leq n} \left\| \frac{1}{n} \sum_{t=1}^m (e_t e'_t - \Sigma_{ee}) \right\| \\
&= O(1) o_{a.s.}(1) = o_{a.s.}(1).
\end{aligned}$$

Since $\sum_{j=0}^{n-1} j \|R_n\|^j = O(n^{2\alpha})$

$$A_{2n} = \frac{1}{n^\alpha} \sum_{j=0}^{n-1} R_n^j \Omega_{uu} (R_n^j)' + O\left(\frac{1}{n^{1-\alpha}}\right) \rightarrow V_C$$

because

$$\begin{aligned}
\text{vec} \left\{ \frac{1}{n^\alpha} \sum_{j=0}^{n-1} R_n^j \Omega_{uu} (R_n^j)' \right\} &= \frac{1}{n^\alpha} \sum_{j=0}^{n-1} (R_n \otimes R_n)^j \text{vec}(\Omega_{uu}) \\
&= \frac{1}{n^\alpha} (I - R_n \otimes R_n)^{-1} \text{vec}(\Omega_{uu}) + o(1) \\
&= -(C \otimes I - I \otimes C)^{-1} \text{vec}(\Omega_{uu}) + o(1) \\
&= \text{vec}(V_C) + o(1).
\end{aligned}$$

For B_n , we have

$$\begin{aligned}
B_n &= \frac{1}{n^{1+\alpha}} \sum_{i=0}^{n-1} \sum_{t=i+1}^n \sum_{j=i+1}^{t-1} R_n^j C(1) e_{t-j} e'_{t-i} C(1)' (R_n^i)' \\
&= \frac{1}{n^{1+\alpha}} \sum_{i=0}^{n-1} \sum_{t=1}^{n-i} \sum_{j=i+1}^{t+i-1} R_n^j C(1) e_{t+i-j} e'_t C(1)' (R_n^i)' \\
&= \frac{1}{n^{1+\alpha}} \sum_{i=0}^{n-1} R_n^i \sum_{t=1}^{n-i} \left[\sum_{j=1}^{t-1} R_n^j C(1) e_{t-j} \right] e'_t C(1)' (R_n^i)' \\
&= \frac{1}{n^\alpha} \sum_{i=0}^{n-1} R_n^{i+1} \frac{1}{n} \sum_{t=1}^{n-i} \zeta_{n,t-1} e'_t C(1)' (R_n^i)'
\end{aligned}$$

and Lemma 3.2(iii) implies that

$$\|B_n\|_{L_1} \leq \|C(1)\| \max_{t \leq n} \left\| \frac{1}{n} \sum_{t=1}^m \zeta_{n,t-1} e'_t \right\|_{L^2} \frac{1}{n^\alpha} \sum_{i=0}^{n-1} \|R_n\|^{2i} = o(1)$$

showing that $\|B_n\| \rightarrow_p 0$.

Proof of Lemma 3.4. By definition of the process ζ_{nt-1} , we can write

$$\begin{aligned} \langle M \rangle_n - V_n &= \frac{1}{n^{1+\alpha}} \sum_{t=1}^n [\zeta_{nt-1} \zeta'_{nt-1} \otimes (H_t - \Sigma_{\varepsilon\varepsilon})] \\ &= A_n + B_n + B'_n \end{aligned} \quad (55)$$

where

$$A_n = \frac{1}{n^{1+\alpha}} \sum_{t=1}^n \left[\left\{ \sum_{j=1}^{t-1} R_n^{j-1} C(1) e_{t-j} e'_{t-j} C(1)' R_n^{j-1} \right\} \otimes (H_t - \Sigma_{\varepsilon\varepsilon}) \right] \quad (56)$$

$$B_n = \frac{1}{n^{1+\alpha}} \sum_{t=1}^n \left[\left\{ \sum_{j=1}^{t-2} \sum_{i=j+1}^{t-1} R_n^{j-1} C(1) e_{t-j} e'_{t-i} C(1)' R_n^{i-1} \right\} \otimes (H_t - \Sigma_{\varepsilon\varepsilon}) \right]. \quad (57)$$

We first expand the term in (56) by adding and subtracting Σ_{ee} from as follows:

$$\begin{aligned} A_n &= \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} (R_n^{j-1} C(1) \otimes I) \sum_{t=1}^{n-j} [e_t e'_t \otimes (H_{t+j} - \Sigma_{\varepsilon\varepsilon})] (C(1)' R_n^{j-1} \otimes I) \\ &= \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} (R_n^{j-1} C(1) \otimes I) \sum_{t=1}^{n-j} [(e_t e'_t - \Sigma_{ee}) \otimes H_{t+j}] (C(1)' R_n^{j-1} \otimes I) \\ &\quad - \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} (R_n^{n-j-1} C(1) \otimes I) \left[\left\{ \sum_{t=1}^j (e_t e'_t - \Sigma_{ee}) \right\} \otimes \Sigma_{\varepsilon\varepsilon} \right] C(1)' (R_n^{n-j-1} \otimes I) \\ &\quad + \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} R_n^{j-1} \Omega_u R_n^{j-1} \otimes \sum_{t=j+1}^n (H_t - \Sigma_{\varepsilon\varepsilon}) \\ &= A_{1n} - A_{2n} + A_{3n} \end{aligned}$$

in order of appearance. It is easy to show that A_{2n} and A_{3n} are $o_{a.s.}(1)$: since

$$\frac{1}{n} \sum_{t=1}^n (e_t e'_t - \Sigma_{ee}) \rightarrow_{a.s.} 0 \quad \text{and} \quad \frac{1}{n} \sum_{t=1}^n (H_t - \Sigma_{\varepsilon\varepsilon}) \rightarrow_{a.s.} 0 \quad (58)$$

by the ergodic theorem, and

$$\frac{1}{n^\alpha} \sum_{j=1}^{n-1} R_n^{j-1} \Omega_u R_n^{j-1} \rightarrow V_C = \int_0^\infty e^{rC} \Omega_u e^{rC} dr \quad (59)$$

we can write

$$\begin{aligned}
A_{3n} &= \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} R_n^{j-1} \Omega_u R_n^{j-1} \otimes \sum_{t=j+1}^n (H_t - \Sigma_{\varepsilon\varepsilon}) \\
&= \frac{1}{n^\alpha} \sum_{j=1}^{n-1} R_n^{j-1} \Omega_u R_n^{j-1} \otimes \frac{1}{n} \sum_{t=1}^n (H_t - \Sigma_{\varepsilon\varepsilon}) - \frac{1}{n^{1+\alpha}} \sum_{j=2}^{n-1} R_n^{j-1} \Omega_u R_n^{j-1} \otimes \sum_{t=1}^j (H_t - \Sigma_{\varepsilon\varepsilon}) \\
&= A'_{3n} + A''_{3n}
\end{aligned}$$

in order of appearance. By (58) and (59), $A'_{3n} \rightarrow_{a.s.} 0$. For arbitrary $\delta > 0$, there exists $j_0(\delta) \in \mathbb{N}$ such that

$$\left\| \frac{1}{j} \sum_{t=1}^j (H_t - \Sigma_{\varepsilon\varepsilon}) \right\| < \delta \quad a.s. \quad \text{for all } j \geq j_0(\delta),$$

giving, *a.s.*

$$\begin{aligned}
\|A''_{3n}\| &\leq \|\Omega_u\| \frac{1}{n^{1+\alpha}} \sum_{j=2}^{j_0(\delta)-1} \sum_{t=1}^{j-1} \|H_t - \Sigma_{\varepsilon\varepsilon}\| + \delta \|\Omega_u\| \frac{1}{n^\alpha} \sum_{j=j_0(\delta)}^{n-1} \frac{j}{n} \|R_n\|^{2(j-1)} \\
&= O\left(\frac{1}{n^{1+\alpha}}\right) + \delta O(1),
\end{aligned}$$

since $\mathbb{E}\|H_1\| < \infty$ implies that $\mathbb{P}(\|H_t\| < \infty) = \mathbb{P}(\|H_1\| < \infty) = 1$ for all $t < j_0(\delta)$. This shows that $A''_{3n} \rightarrow_{a.s.} 0$. The same argument works for A_{2n} : almost surely,

$$\begin{aligned}
\|A_{2n}\| &\leq \|\Sigma_{\varepsilon\varepsilon}\| \|C(1)\|^2 \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} j \|R_n\|^{2(n-j-1)} \left\| \frac{1}{j} \sum_{t=1}^j (e_t e'_t - \Sigma_{ee}) \right\| \\
&\leq \|\Sigma_{\varepsilon\varepsilon}\| \|C(1)\|^2 \left\{ \frac{1}{n^{1+\alpha}} \sum_{j=1}^{j_0(\delta)-1} \sum_{t=1}^j \|(e_t e'_t - \Sigma_{ee})\| + \delta \frac{1}{n^\alpha} \sum_{j=j_0(\delta)}^{n-1} \frac{j}{n} \|R_n\|^{2(n-j-1)} \right\} \\
&= O\left(\frac{1}{n^{1+\alpha}}\right) + \delta O(1).
\end{aligned}$$

We conclude that $\|A_n - A_{1n}\| = o_p(1)$ and

$$\text{vec}(A_{1n}) = \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} (R_n^{j-1} \otimes R_n^{j-1}) (C(1) \otimes C(1)) \sum_{t=1}^{n-j} \text{vec}[(e_t e'_t - \Sigma_{ee}) \otimes H_{t+j}].$$

For any matrices $K \in \mathbb{R}^{r \times r}$, $L \in \mathbb{R}^{m \times m}$, the vectors $\text{vec}(K \otimes L)$ and $\text{vec}(K) \otimes \text{vec}(L)$ consist of the same elements $\{K_{ij} L_{kl} : 1 \leq i, j \leq r, 1 \leq k, l \leq m\}$ but appear in different order in the two vectors. Therefore, there exists a $m^2 r^2 \times m^2 r^2$ permutation

matrix Π such that $\text{vec}(K \otimes L) = \Pi [\text{vec}(K) \otimes \text{vec}(L)]$. Using this and the identity $\text{vec}(K) = D_m \text{vech}(K)$ for a symmetric $m \times m$ matrix K , where D_m denotes the $m^2 \times m(m+1)/2$ duplication matrix (Chapter 11 of Abadir and Magnus, 2005), we can write

$$\text{vec}(A_{1n}) = \Pi (I_{r^2} \otimes D_m) \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} (R_n^{j-1} \otimes R_n^{j-1}) (C(1) \otimes C(1)) \sum_{t=1}^{n-j} \Sigma_t(j), \quad (60)$$

with $\Sigma_t(j) = \text{vec}(e_t e_t' - \Sigma_{ee}) \otimes \text{vech}(H_{t+j})$ as in Lemma 3.5. Since $\|\Pi\| = 1$ and $\|D_m\| = \sqrt{2}$ for a permutation matrix Π , (60) yields the asymptotic bound

$$\|\text{vec}(A_{1n})\| \leq \sqrt{2} \|C(1)\|^2 \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} \|R_n\|^{2(j-1)} \left\| \sum_{t=1}^{n-j} \Sigma_t(j) \right\| + o_p(1). \quad (61)$$

The term in (57) can be written as:

$$\begin{aligned} B_n &= \frac{1}{n^{1+\alpha}} \sum_{t=1}^n \sum_{j=1}^{t-2} \sum_{i=j+1}^{t-1} [\{R_n^{j-1} C(1) e_{t-j} e_{t-i}' C(1)' R_n^{i-1}\} \otimes (H_t - \Sigma_{\varepsilon\varepsilon})] \\ &= \frac{1}{n^{1+\alpha}} \sum_{t=1}^n \sum_{j=1}^{t-2} \sum_{i=1}^{t-j-1} [\{R_n^{j-1} C(1) e_{t-j} e_{t-i-j}' C(1)' R_n^{i+j-1}\} \otimes (H_t - \Sigma_{\varepsilon\varepsilon})] \\ &= \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-2} \sum_{t=j+2}^n \sum_{i=1}^{t-j-1} [\{R_n^{j-1} C(1) e_{t-j} e_{t-i-j}' C(1)' R_n^{i+j-1}\} \otimes (H_t - \Sigma_{\varepsilon\varepsilon})] \\ &= \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} \sum_{i=1}^{t-1} [\{R_n^{j-1} C(1) e_t e_{t-i}' C(1)' R_n^{i+j-1}\} \otimes (H_{t+j} - \Sigma_{\varepsilon\varepsilon})] \\ &= \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} \left[\left\{ R_n^{j-1} C(1) e_t \left(\sum_{i=1}^{t-1} R_n^{i-1} C(1) e_{t-i} \right)' R_n^j \right\} \otimes (H_{t+j} - \Sigma_{\varepsilon\varepsilon}) \right] \\ &= \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} [\{R_n^{j-1} C(1) e_t \zeta_{n,t-1}' R_n^j\} \otimes (H_{t+j} - \Sigma_{\varepsilon\varepsilon})] \end{aligned}$$

by definition of the process $\zeta_{n,t}$. Employing the same argument for the vectorisation of a Kronecker product, we deduce that

$$\begin{aligned} \text{vec}(B_n) &= \Pi \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} [\text{vec} \{R_n^{j-1} C(1) e_t \zeta_{n,t-1}' R_n^j\} \otimes \text{vec}(H_{t+j} - \Sigma_{\varepsilon\varepsilon})] \\ &= \Pi (I_r \otimes R_n^{-1} \otimes D_m) \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} [R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes \text{vech}(H_{t+j})] + o_p(1) \\ &= \Pi (I_r \otimes R_n^{-1} \otimes D_m) \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} S_{n,t}(j) + o_p(1) \end{aligned}$$

where $S_{n,t}(j)$ is defined in (23) and the term involving $\Sigma_{\varepsilon\varepsilon}$ is $o_p(1)$ because

$$\begin{aligned} & \left\| \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} [R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t] \right\|_{L_1} \\ & \leq \frac{\|C(1)\|}{n^\alpha} \sum_{j=1}^{n-2} \|R_n\|^{2j} \max_{t \leq n} \left\| \frac{1}{n} \sum_{t=2}^m [\zeta_{n,t-1} \otimes e_t] \right\|_{L_2} = o(1) \end{aligned}$$

by Lemma 3.2(iii). The last expression for $\text{vec}(B_n)$ yields

$$\|\text{vec}(B_n)\| \leq b \frac{1}{n^{1+\alpha}} \left\| \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} S_{n,t}(j) \right\| + o_p(1) \quad (62)$$

for some uniform bounding constant $b > 0$. Combining (61) and (62) shows (27) and part (iii).

It remains to show that $\Sigma_t(j)$ and $S_{n,t}(j)$ satisfy (14) with innovations given by (25) and (26) respectively. For $S_{n,t}(j)$, applying (4) to $\text{vech}(H_{t+j})$ yields

$$\begin{aligned} S_{n,t}(j) &= R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes \text{vech}(H_{t+j}) \\ &= R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes \left\{ \sum_{i=1}^q A_i \text{vech}(H_{t+j-i}) + \sum_{k=1}^p B_k \text{vech}(H_{t+j-k}) \right\} \\ &\quad + R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes \left\{ \sum_{i=1}^q A_i \text{vech}(\varepsilon_{t+j-i} \varepsilon'_{t+j-i} - H_{t+j-i}) \right\} \\ &\quad + R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes \varphi \\ &= \sum_{i=1}^{\kappa} (R_n^i \otimes R_n^i \otimes \Gamma_i) [R_n^{j-i} \zeta_{n,t-1} \otimes R_n^{j-i} C(1) e_t \otimes \text{vech}(H_{t+j-i})] + v_{n,t}(j) \\ &= \sum_{i=1}^{\kappa} (R_n^i \otimes R_n^i \otimes \Gamma_i) S_{n,t}(j-i) + v_{n,t}(j) \end{aligned} \quad (63)$$

where $\kappa = p \vee q$,

$$\begin{aligned} v_{n,t}(j) &= R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes \varphi \\ &\quad + \sum_{i=1}^q (I_{r^2} \otimes A_i) [R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes w_{t+j-i}] \end{aligned}$$

and $w_{t+j-i} = \text{vech}(\varepsilon_{t+j-i} \varepsilon'_{t+j-i} - H_{t+j-i})$. Since the above expression for $v_{n,t}(j)$ coincides with (26) and $\Gamma_{n,i} = R_n^i \otimes R_n^i \otimes \Gamma_i$ by (13), (63) shows part (ii).

Applying (4) to $\text{vech}(H_{t+j})$ in

$$\Sigma_t(j) = \text{vec}(e_t e_t' - \Sigma_{ee}) \otimes \text{vech}(H_{t+j})$$

and proceeding as in (63) shows part (i).

Proof of Lemma 3.5(i). The processes in (22) and (23) have companion form solutions

$$\tilde{\Sigma}_t(j) = M_\kappa^j \tilde{\Sigma}_t(0) + \sum_{l=1}^j M_\kappa^{j-l} \tilde{v}_t(l), \quad j \geq 1 \quad (64)$$

$$\tilde{S}_{n,t}(j) = M_{n,\kappa}^j \tilde{S}_{n,t}(0) + \sum_{l=1}^j M_{n,\kappa}^{j-l} \tilde{v}_{n,t}(l), \quad j \geq 1 \quad (65)$$

for the "stacked" processes

$$\tilde{\Sigma}_t(j) = [\Sigma_t(j)', \Sigma_t(j-1)', \dots, \Sigma_t(j-\kappa+1)']' \quad (66)$$

$$\tilde{S}_{n,t}(j) = [S_{n,t}(j)', S_{n,t}(j-1)', \dots, S_{n,t}(j-\kappa+1)']' \quad (67)$$

and

$$\tilde{v}_t(j) = [v_t(j)', 0, \dots, 0]', \quad \tilde{v}_{n,t}(j) = [v_{n,t}(j)', 0, \dots, 0]'. \quad (68)$$

For part (i), by (66), (68) and the definition of the Euclidian vector norm,

$$\left\| \sum_{t=1}^{n-j} \Sigma_t(j) \right\| \leq \left\| \sum_{t=1}^{n-j} \tilde{\Sigma}_t(j) \right\| \quad \text{and} \quad \left\| \sum_{t=1}^{n-j} \tilde{v}_t(l) \right\| = \left\| \sum_{t=1}^{n-j} v_t(l) \right\|.$$

We can therefore apply the companion form solution (64) to the first term of the bound (27) of Lemma 3.4(iii) to obtain

$$\begin{aligned} \sigma_n &= \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} \|R_n\|^{2(j-1)} \left\| \sum_{t=1}^{n-j} \Sigma_t(j) \right\| \\ &\leq \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} \|R_n\|^{2(j-1)} \left\| \sum_{t=1}^{n-j} \tilde{\Sigma}_t(j) \right\| \\ &\leq \frac{1}{n^\alpha} \sum_{j=1}^{\infty} \|M_\kappa^j\| \frac{1}{n} \sum_{t=1}^n \left\| \tilde{\Sigma}_t(0) \right\| \\ &\quad + \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} \|R_n\|^{2(j-1)} \sum_{l=1}^j \|M_\kappa^{j-l}\| \left\| \sum_{t=1}^{n-j} v_t(l) \right\| \\ &= \sigma_{1n} + \sigma_{2n} \end{aligned} \quad (69)$$

in order of appearance. The first term of (69) satisfies $\mathbb{E} \|\sigma_{1n}\| = O(n^{-\alpha})$ because $\sum_{j=1}^{\infty} \|M_{\kappa}^j\| < \infty$ by Lemma 3.1 and

$$\begin{aligned} \mathbb{E} \left\| \tilde{\Sigma}_t(0) \right\| &= \mathbb{E} \left\{ \sum_{i=0}^{\kappa-1} \|\Sigma_t(-i)\|^2 \right\}^{1/2} \leq \sum_{i=0}^{\kappa-1} \mathbb{E} \|\Sigma_t(-i)\| \\ &\leq \kappa \max_{i < \kappa} \mathbb{E} (\|e_t\|^2 \|H_{t-i}\|) \\ &\leq \kappa \max_{i < \kappa} \left\{ \mathbb{E} \|e_1\|^4 \mathbb{E} \|H_{t-i}\|^2 \right\}^{1/2} \\ &\leq \kappa \left\{ \mathbb{E} \|e_1\|^4 \mathbb{E} \|\varepsilon_1\|^4 \right\}^{1/2} \end{aligned}$$

by the Jensen inequality for conditional expectations. For the second term of (69), letting

$$\omega_t(k) = \text{vec}(e_t e_t' - \Sigma_{ee}) \otimes w_{t+k} \quad (70)$$

and using the expression in (25) we can write

$$\begin{aligned} \sigma_{2n} &\leq \frac{\|\varphi\|}{n^\alpha} \sum_{l=1}^{\infty} \|M_{\kappa}^l\| \sum_{j=1}^{n-1} \|R_n\|^{2(n-j-1)} \frac{j}{n} \left\| \frac{1}{j} \sum_{t=1}^j \text{vec}(e_t e_t' - \Sigma_{ee}) \right\| \\ &\quad + \frac{\|\varphi\|}{n^{1+\alpha}} \sum_{j=1}^{n-1} \|R_n\|^{2(j-1)} \sum_{i=1}^q \|A_i\| \sum_{l=1}^j \|M_{\kappa}^{j-l}\| \left\| \sum_{t=1}^{n-j} \omega_t(l-i) \right\| \end{aligned}$$

The first term on the right is $o_{a.s.}(1)$ by the ergodic theorem; the second term is $O_p(n^{-\alpha})$ when $l \leq i$ because, in this case, it is bounded in L_1 norm by

$$\begin{aligned} \frac{b}{n^{1+\alpha}} \max_{1 \leq l, i \leq q} \sum_{t=1}^n \mathbb{E} \|\omega_t(l-i)\| &\leq \frac{2b}{n^{1+\alpha}} \max_{1 \leq l, i \leq q} \sum_{t=1}^n \mathbb{E} \|e_t\|^2 \|w_{t+l-i}\| \\ &\leq \frac{4b}{n^\alpha} \left\{ \mathbb{E} \|e_1\|^4 \mathbb{E} \|\varepsilon_1\|^4 \right\}^{1/2} \end{aligned}$$

where

$$b = \sum_{j=1}^{\infty} \|M_{\kappa}^j\| \sum_{i, l=1}^q \|A_i\| \|M_{\kappa}^{-l}\|$$

is a finite constant and $\mathbb{E} \|w_1\|^2 \leq 4\mathbb{E} \|\varepsilon_1\|^4$ by the Jensen inequality for conditional expectations. We conclude that

$$\sigma_{2n} \leq \frac{\|\varphi\|}{n^\alpha} \sum_{j=1}^{n-1} \|R_n\|^{2(j-1)} \sum_{i=1}^q \|A_i\| \sum_{l=1}^{j-i} \|M_{\kappa}^{j-i-l}\| \left\| \frac{1}{n} \sum_{t=1}^{n-j} \omega_t(l) \right\| + o_p(1)$$

so the condition

$$\max_{1 \leq j, l \leq n} \left\| \frac{1}{n} \sum_{t=1}^{n-j} \omega_t(l) \right\|_{L_1} \rightarrow 0. \quad (71)$$

is sufficient to show that $\sigma_{2n} \rightarrow_p 0$. The definition of $\omega_t(l)$ in (70) implies that

$$\tilde{\omega}_t(l) = \omega_t(l) \mathbf{1} \{ \|e_t\|^4 \leq L_n \}$$

is an \mathcal{F}_{t+l} -martingale difference sequence for each $l \geq 1$, where the truncating sequence $(L_n)_{n \in \mathbb{N}}$ is chosen to satisfy $L_n \rightarrow \infty$ and $L_n/n \rightarrow 0$. The Lyapounov inequality then gives

$$\begin{aligned} \left\| \frac{1}{n} \sum_{t=1}^{n-j} \tilde{\omega}_t(l) \right\|_{L_1} &\leq \frac{1}{n} \left\| \sum_{t=1}^{n-j} \tilde{\omega}_t(l) \right\|_{L_2} \\ &= \frac{1}{n} \left(\sum_{t=1}^{n-j} \mathbb{E} \|\tilde{\omega}_t(l)\|^2 \right)^{1/2} \\ &\leq b \frac{L_n^{1/2}}{n} \left(\sum_{t=1}^{n-j} \mathbb{E} \|w_{t+l}\|^2 \right)^{1/2} \\ &\leq 2b \frac{L_n^{1/2}}{\sqrt{n}} (\mathbb{E} \|\varepsilon_1\|^4)^{1/2} \rightarrow 0 \end{aligned}$$

uniformly in j, l . Therefore, the Cauchy-Schwarz inequality yields

$$\begin{aligned} \max_{1 \leq j, l \leq n} \left\| \frac{1}{n} \sum_{t=1}^{n-j} \omega_t(l) \right\|_{L_1} &\leq \max_{1 \leq j, l \leq n} \frac{1}{n} \sum_{t=1}^{n-j} \mathbb{E} \|\omega_t(l) \mathbf{1} \{ \|e_t\|^4 > L_n \} \| + o(1) \\ &\leq \max_{1 \leq j, l \leq n} \frac{1}{n} \sum_{t=1}^{n-j} \mathbb{E} [(\|e_t\|^2 + \|\Sigma_{ee}\|^2) \mathbf{1} \{ \|e_t\|^4 > L_n \} \|w_{t+l}\|] + o(1) \\ &\leq \{ \mathbb{E} [(\|e_1\|^4 + \|\Sigma_{ee}\|^4) \mathbf{1} \{ \|e_1\|^4 > L_n \}] E \|w_1\|^2 \}^{1/2} + o(1) \\ &= o(1) \end{aligned}$$

since $\mathbb{E} \|e_1\|^4 < \infty$ and $L_n \rightarrow \infty$. This proves (71) and $\sigma_n \rightarrow_p 0$.

We turn to the second term of the bound (27) of Lemma 3.4(iii), to prove that $s_n \rightarrow_p 0$. Using the solution (65) and the same argument leading to (69), we obtain

$$\begin{aligned} s_n &\leq \frac{1}{n^{1+\alpha}} \left\| \sum_{j=1}^{n-1} \sum_{t=2}^{n-j} \tilde{S}_{n,t}(j) \right\| \\ &\leq \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-1} \|M_{n,\kappa}^j\| \sum_{t=2}^n \left\| \tilde{S}_{n,t}(0) \right\| + \frac{1}{n^{1+\alpha}} \left\| \sum_{j=1}^{n-1} \sum_{t=2}^{n-j} \sum_{l=1}^j M_{n,\kappa}^{j-l} \tilde{v}_{n,t}(l) \right\| \\ &= s_{1n} + s_{2n} \end{aligned}$$

Since $\sup_{n \geq 1} \sum_{j=1}^{\infty} \|M_{n,\kappa}^j\| < \infty$ by Lemma 3.1,

$$\begin{aligned} \mathbb{E} \|s_{1n}\| &\leq b \frac{1}{n^\alpha} \max_{t \leq n} \mathbb{E} \left\| \tilde{S}_{n,t}(0) \right\| \leq b \max_{t \leq n} \max_{0 \leq i \leq \kappa-1} \frac{1}{n^\alpha} \mathbb{E} (\|e_t\| \|\zeta_{n,t-1}\| \|H_{t-i}\|) \\ &\leq \frac{1}{n^{\alpha/2}} \|H_1\|_{L_2} \|e_1\|_{L_4} \max_{t \leq n} \left\| \frac{\zeta_{n,t-1}}{n^{\alpha/2}} \right\|_{L_4} = O\left(\frac{1}{n^{\alpha/2}}\right) \end{aligned}$$

by Lemma 3.2(i). For s_{2n} , some care is required to use the norm equivalence between $\tilde{v}_{n,t}(l)$ and $v_{n,t}(l)$: standard manipulations yield

$$\begin{aligned} s_{2n} &= \frac{1}{n^{1+\alpha}} \left\| \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} \sum_{l=0}^{j-1} M_{n,\kappa}^l \tilde{v}_{n,t}(j-l) \right\| \\ &= \frac{1}{n^{1+\alpha}} \left\| \sum_{l=0}^{n-3} M_{n,\kappa}^l \sum_{j=l+1}^{n-2} \sum_{t=2}^{n-j} \tilde{v}_{n,t}(j-l) \right\| \\ &\leq \sum_{l=0}^{n-3} M_{n,\kappa}^l \frac{1}{n^{1+\alpha}} \left\| \sum_{j=1}^{n-l-2} \sum_{t=2}^{n-l-j} \tilde{v}_{n,t}(j) \right\| \\ &= \sum_{l=0}^{n-3} M_{n,\kappa}^l \frac{1}{n^{1+\alpha}} \left\| \sum_{j=1}^{n-l-2} \sum_{t=2}^{n-l-j} v_{n,t}(j) \right\|. \end{aligned}$$

Substituting the expression for $v_{n,t}(j)$ in (26) we obtain

$$\begin{aligned} s_{2n} &\leq \sum_{l=0}^{n-3} M_{n,\kappa}^l \sum_{i=1}^q \|A_i\| \frac{1}{n^{1+\alpha}} \left\| \sum_{j=1}^{n-l-2} \sum_{t=2}^{n-l-j} [R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes w_{t+j-i}] \right\| \\ &\quad + \|\varphi\| \|C(1)\| \sum_{l=0}^{n-3} M_{n,\kappa}^l \frac{1}{n^{1+\alpha}} \sum_{j=1}^{n-l-2} \|R_n\|^{2j} \left\| \sum_{t=2}^{n-l-j} (\zeta_{n,t-1} \otimes e_t) \right\| \end{aligned}$$

The last term on the right converges to 0 in L_1 by Lemma 3.2(iii). For the first term, partitioning the sum

$$j \in \{1, \dots, n-l-2\} = \{1, \dots, i\} \cup \{i+1, \dots, n-l-2\}$$

we obtain that

$$s_{2n} \leq s_{3n} + s_{4n} + o_p(1)$$

where

$$s_{3n} = \sum_{l=0}^{n-3} M_{n,\kappa}^l \sum_{i=1}^q \|A_i\| \frac{1}{n^{1+\alpha}} \left\| \sum_{j=i+1}^{n-l-2} \sum_{t=2}^{n-l-j} [R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes w_{t+j-i}] \right\|$$

and

$$\begin{aligned}
\mathbb{E}s_{4n} &\leq \|C(1)\| \sum_{l=0}^{n-3} \|M_{n,\kappa}^l\| \sum_{i=1}^q \|A_i\| \frac{1}{n^{1+\alpha}} \sum_{j=1}^q \sum_{t=2}^{n-l-j} \mathbb{E} \|\zeta_{n,t-1}\| \|e_t\| \|w_{t+j-i}\| \\
&\leq \|C(1)\| \sum_{l=0}^{n-3} \|M_{n,\kappa}^l\| \sum_{i=1}^q \|A_i\| \frac{q}{n^\alpha} \left(\max_{t \leq n} \mathbb{E} \|\zeta_{n,t-1}\|^4 \right)^{1/4} \|e_1\|_{L_4} \|w_1\|_{L_2} \\
&= O\left(\frac{1}{n^{\alpha/2}}\right).
\end{aligned}$$

It remains to show that $s_{3n} \rightarrow_p 0$. The inner double sum in the expression for s_{3n} can be written as

$$\begin{aligned}
&\sum_{j=i+1}^{n-l-2} \sum_{t=2}^{n-l-j} [R_n^j \zeta_{n,t-1} \otimes R_n^j C(1) e_t \otimes w_{t+j-i}] \\
&= \sum_{j=1}^{n-l-i-2} \sum_{t=2}^{n-l-j-i} [R_n^{j+i} \zeta_{n,t-1} \otimes R_n^{j+i} C(1) e_t \otimes w_{t+j}] \\
&= (R_n^i \otimes R_n^i) \sum_{t=3}^{n-l-i} (\xi_{n,t-1} \otimes w_t)
\end{aligned}$$

where

$$\xi_{n,t-1} = \sum_{j=1}^{t-2} (R_n^j \zeta_{n,t-j-1} \otimes R_n^j C(1) e_{t-j})$$

is a \mathcal{F}_{t-1} -martingale difference sequence satisfying

$$\begin{aligned}
\max_{t \leq n} \mathbb{E} \|\xi_{n,t-1}\|^2 &= \|C(1)\|^2 \max_{t \leq n} \sum_{j=1}^{t-2} \|R_n\|^{4j} \mathbb{E} \left(\|\zeta_{n,t-j-1}\|^2 \|e_{t-j}\|^2 \right) \\
&\leq \|C(1)\|^2 \sum_{j=1}^{n-2} \|R_n\|^{4j} \|e_1\|_{L_4}^2 \left(\max_{t \leq n} \mathbb{E} \|\zeta_{n,t}\|^4 \right)^{1/2} \\
&= O(n^{2\alpha}). \tag{72}
\end{aligned}$$

Substituting in the expression for s_{3n} we obtain

$$\begin{aligned}
\mathbb{E}s_{3n} &\leq \frac{1}{n^{1+\alpha}} \max_{i,l} \mathbb{E} \left\| \sum_{t=3}^{n-l-i} (\xi_{n,t-1} \otimes w_t) \right\| \\
&\leq \frac{1}{n^{1+\alpha}} \max_{i,l} \mathbb{E} \left\| \sum_{t=3}^{n-l-i} (\xi_{n,t-1} \otimes w_t) \mathbf{1}_{\{\|H_t\| \leq L_n\}} \right\| \\
&\quad + \frac{1}{n^{1+\alpha}} \max_{i,l} \left\| \sum_{t=3}^{n-l-i} (\xi_{n,t-1} \otimes w_t) \mathbf{1}_{\{\|H_t\| > L_n\}} \right\| \\
&= \epsilon_{1n} + \epsilon_{2n}
\end{aligned}$$

where the sequence $\xi_{n,t-1} \otimes w_t$ is an \mathcal{F}_t -martingale difference sequence

$$\begin{aligned} w_t &= \text{vech}(\varepsilon_t \varepsilon_t' - H_t) = \text{vech} H_t^{1/2} (\eta_{\varepsilon t} \eta_{\varepsilon t}' - I_m) H_t^{1/2} \\ &= D_m^+ \left(H_t^{1/2} \otimes H_t^{1/2} \right) \text{vec}(\eta_{\varepsilon t} \eta_{\varepsilon t}' - I_m) \end{aligned}$$

where D_m^+ is the Moore-Penrose inverse of the duplication matrix D_m , satisfying $\|D_m^+\| = 1$ (e.g. 11.30 in Abadir and Magnus, 2005). Since $\|H_t\|$ is \mathcal{F}_{t-1} -measurable, the martingale difference property is preserved for the truncated sequence

$$(\xi_{n,t-1} \otimes w_t) \mathbf{1} \{ \|H_t\| \leq L_n \},$$

giving

$$\begin{aligned} \epsilon_{1n} &\leq \frac{1}{n^{1+\alpha}} \left\{ \max_{i,l} \sum_{t=3}^{n-l-i} \mathbb{E} \|\xi_{n,t-1}\|^2 \|w_t\|^2 \mathbf{1} \{ \|H_t\| \leq L_n \} \right\}^{1/2} \\ &\leq \frac{B}{n^{1+\alpha}} \left\{ \max_{i,l} \sum_{t=3}^{n-l-i} \mathbb{E} \|\xi_{n,t-1}\|^2 \|H_t\|^2 \mathbf{1} \{ \|H_t\| \leq L_n \} \right\}^{1/2} \\ &\leq \frac{BL_n}{n^{1/2+\alpha}} \left\{ \max_{t \leq n} \mathbb{E} \|\xi_{n,t-1}\|^2 \right\}^{1/2} \\ &= O\left(\frac{L_n}{n^{1/2}}\right) \end{aligned}$$

by (72), where $B = \{2\mathbb{E}(\|\eta_{\varepsilon 1}\|^4 + 1)\}^{1/2}$. Taking $L_n \rightarrow \infty$ with $L_n/n^{1/2} \rightarrow 0$ we can write

$$\begin{aligned} \epsilon_{2n} &\leq \frac{1}{n^{1+\alpha}} \sum_{t=3}^n \mathbb{E} (\|\xi_{n,t-1}\| \|w_t\| \mathbf{1} \{ \|H_t\| > L_n \}) \\ &\leq \frac{2}{n^{1+\alpha}} \sum_{t=3}^n \mathbb{E} (\|\xi_{n,t-1}\| \|H_t\| \mathbf{1} \{ \|H_t\| > L_n \}) \\ &\leq \frac{2}{n^{1+\alpha}} \sum_{t=3}^n \left\{ \mathbb{E} (\|\xi_{n,t-1}\|^2) \mathbb{E} (\|H_t\|^2 \mathbf{1} \{ \|H_t\| > L_n \}) \right\}^{1/2} \\ &\leq \frac{2}{n^\alpha} \left\{ \max_{t \leq n} \mathbb{E} \|\xi_{n,t-1}\|^2 \right\}^{1/2} \left\{ \mathbb{E} (\|H_1\|^2 \mathbf{1} \{ \|H_1\| > L_n \}) \right\}^{1/2} \\ &= O(1) \left\{ \mathbb{E} (\|H_1\|^2 \mathbf{1} \{ \|H_1\| > L_n \}) \right\}^{1/2} = o(1). \end{aligned}$$

This completes the proof of $s_n \rightarrow_p 0$.

Proof of Lemma 3.5(ii). For all $\delta > 0$,

$$\begin{aligned}
\mathcal{L}_n(\delta) &= \frac{1}{n^{1+\alpha}} \sum_{t=2}^n \|\zeta_{nt-1}\|^2 \mathbb{E}_{\mathcal{F}_{t-1}} \left(\|\varepsilon_t\|^2 \mathbf{1} \left\{ \|\varepsilon_t\| \|\zeta_{nt-1}\| > n^{\frac{1+\alpha}{2}} \delta \right\} \right) \\
&\leq \frac{1}{n^{1+\alpha}} \sum_{t=2}^n \|\zeta_{nt-1}\|^2 \mathbf{1} \left\{ \|\zeta_{nt-1}\| > L_n^{-1} n^{\frac{1+\alpha}{2}} \delta \right\} \mathbb{E}_{\mathcal{F}_{t-1}} (\|\varepsilon_t\|^2) \\
&\quad + \frac{1}{n^{1+\alpha}} \sum_{t=2}^n \|\zeta_{nt-1}\|^2 \mathbb{E}_{\mathcal{F}_{t-1}} (\|\varepsilon_t\|^2 \mathbf{1} \{ \|\varepsilon_t\| > L_n \}) \\
&= \mathcal{L}_{1n}(\delta) + \mathcal{L}_{2n}
\end{aligned}$$

for an arbitrary sequence $(L_n)_{n \in \mathbb{N}}$ satisfying

$$L_n \rightarrow \infty \text{ and } \mathbb{P} \left(\max_{t \leq n} \|\zeta_{nt-1}\| > L_n^{-1} n^{\frac{1+\alpha}{2}} \delta \right) \rightarrow 0. \quad (73)$$

For the first term,

$$\mathcal{L}_{1n}(\delta) \leq \mathbf{1} \left\{ \max_{t \leq n} \|\zeta_{nt-1}\| > L_n^{-1} n^{\frac{1+\alpha}{2}} \delta \right\} \frac{1}{n^{1+\alpha}} \sum_{t=2}^n \|\zeta_{nt-1}\|^2 \mathbb{E}_{\mathcal{F}_{t-1}} (\|\varepsilon_t\|^2) \xrightarrow{L_1} 0$$

since

$$\mathbb{E} \mathbf{1} \left\{ \max_{t \leq n} \|\zeta_{nt-1}\| > L_n^{-1} n^{\frac{1+\alpha}{2}} \delta \right\} = \mathbb{P} \left(\max_{t \leq n} \|\zeta_{nt-1}\| > L_n^{-1} n^{\frac{1+\alpha}{2}} \delta \right) \rightarrow 0$$

by (73) and

$$\begin{aligned}
\frac{1}{n^{1+\alpha}} \sum_{t=2}^n \mathbb{E} \left[\|\zeta_{nt-1}\|^2 \mathbb{E}_{\mathcal{F}_{t-1}} (\|\varepsilon_t\|^2) \right] &= \frac{1}{n^{1+\alpha}} \sum_{t=2}^n \mathbb{E} \left(\|\zeta_{nt-1}\|^2 \|\varepsilon_t\|^2 \right) \\
&\leq \left(\max_{t \leq n} \mathbb{E} \left\| \frac{1}{n^{\alpha/2}} \zeta_{nt-1} \right\|^4 \right)^{1/2} (\mathbb{E} \|\varepsilon_1\|^4)^{1/2} \\
&= O(1)
\end{aligned}$$

by Lemma 3.2(i). For the second term, the same chain of inequalities give

$$\begin{aligned}
\mathbb{E}(\mathcal{L}_{2n}) &\leq \frac{1}{n^{1+\alpha}} \sum_{t=2}^n \left(\mathbb{E} \|\zeta_{nt-1}\|^4 \right)^{1/2} \left\{ \mathbb{E} \left[\mathbb{E}_{\mathcal{F}_{t-1}} (\|\varepsilon_t\|^2 \mathbf{1} \{ \|\varepsilon_t\| > L_n \}) \right]^2 \right\}^{1/2} \\
&\leq \left(\max_{t \leq n} \mathbb{E} \left\| \frac{1}{n^{\alpha/2}} \zeta_{nt-1} \right\|^4 \right)^{1/2} \left\{ \mathbb{E} \|\varepsilon_1\|^4 \mathbf{1} \{ \|\varepsilon_1\| > L_n \} \right\}^{1/2} \rightarrow 0
\end{aligned}$$

since $\mathbb{E} \|\varepsilon_1\|^4 < \infty$ and $L_n \rightarrow \infty$.

Proof of Theorem 3.6. Part (i) is an immediate consequence of established results. By Lemma 3.2(iv), $n^{-(1+\alpha)/2} \sum_{t=1}^n (x_{t-1} \otimes \varepsilon_t)$ and the martingale transform N_n in (19) have the same the limit distribution. By Lemma 3.5(i), the predictable quadratic variation of N_n in (20) is given by

$$\langle N \rangle_n = \frac{1}{n^{1+\alpha}} \sum_{t=1}^n (\zeta_{n,t-1} \zeta'_{n,t-1} \otimes \Sigma_{\varepsilon\varepsilon}) + o_p(1) \rightarrow_p V_C \otimes \Sigma_{\varepsilon\varepsilon}$$

where the last convergence in probability follows by Lemma 3.3. Since the Lindeberg condition of Lemma 3.5(ii) holds, a standard martingale central limit theorem, e.g. Corollary 3.1 of Hall and Heyde (1980), establishes the asymptotic distribution of part (i). Part (ii) follows immediately from part (i) and Lemma 3.3 and (11). For completeness, we provide a proof of (11): Letting $\underline{x}_t = x_t - \bar{x}_{n-1}$ and $\underline{\varepsilon}_t = \varepsilon_t - \bar{\varepsilon}_n$, the fact that $\sum_{t=1}^n x_{t-1} = O_p(n^{1/2+\alpha})$ implies that

$$\begin{aligned} \frac{1}{n^{1+\alpha}} \left\| \sum_{t=1}^n \underline{x}_t \underline{x}'_t - \sum_{t=1}^n x_t x'_t \right\| &= \frac{1}{n^\alpha} \left\| \bar{x}_{n-1} \bar{x}'_{n-1} \right\| \leq \left\| \frac{1}{n^{\alpha/2}} \bar{x}_{n-1} \right\|^2 \\ &= O_p \left[\left(\frac{1}{n^{\alpha/2}} \frac{n^{1/2+\alpha}}{n} \right)^2 \right] = O_p \left(\frac{1}{n^{1-\alpha}} \right). \end{aligned}$$

Also,

$$\begin{aligned} \frac{1}{n^{\frac{1+\alpha}{2}}} \left\| \sum_{t=1}^n \underline{x}_{t-1} \underline{\varepsilon}'_t - \sum_{t=1}^n x_{t-1} \varepsilon'_t \right\| &= \frac{1}{n^{\frac{1+\alpha}{2}}} \left\| n \bar{x}_{n-1} \bar{\varepsilon}'_n \right\| \\ &\leq \left\| \frac{1}{n^{\alpha/2}} \bar{x}_{n-1} \right\| \left\| \sqrt{n} \bar{\varepsilon}_n \right\| \\ &= O_p \left(\frac{1}{n^{(1-\alpha)/2}} \right). \end{aligned}$$

Combining the two remainder terms proves (11). Note that the same orders of magnitude apply for the purely stationary case, by putting $\alpha = 0$. Part (iii) follows immediately by part (ii), (11) and $n^{-1-\alpha} \|\underline{X}'\underline{X} - X'X\| = o_p(1)$.

Proof of Theorem 4.1 and Corollary 4.2. See the proof of Theorem A and Theorem 1 in KMS(2015b).

Proof of Lemma 4.3. To show part (i) we employ the decompositions

$$\tilde{z}_t = x_t + \frac{C_z}{n^\beta} \psi_{nt}, \quad \alpha \leq \beta \tag{74}$$

$$\tilde{z}_t = z_t + \frac{C}{n^\alpha} \psi_{nt}, \quad \alpha > \beta \tag{75}$$

see equations (13) and (23) of PM (2009), where $\psi_{nt} = \sum_{j=1}^t R_{nz}^{t-j} x_{j-1}$. We first establish the bound

$$\max_{1 \leq t \leq n} \mathbb{E} \|\psi_{nt}\|^4 = O(n^{2\alpha+4\beta}). \quad (76)$$

The Minkowski inequality gives

$$\begin{aligned} \mathbb{E} \|\psi_{nt}\|^4 &= \mathbb{E} \left\| \sum_{j=1}^t R_{nz}^{t-j} x_{j-1} \right\|^4 \leq \left(\sum_{j=1}^t \left\{ \mathbb{E} \|R_{nz}^{t-j} x_{j-1}\|^4 \right\}^{1/4} \right)^4 \\ &\leq \left(\sum_{j=1}^t \|R_{nz}\|^{t-j} \left\{ \max_{1 \leq t \leq n} \mathbb{E} \|x_{j-1}\|^4 \right\}^{1/4} \right)^4 \\ &\leq \left\{ \max_{1 \leq t \leq n} \mathbb{E} \|x_{j-1}\|^4 \right\} \left(\sum_{j=1}^n \|R_{nz}\|^j \right)^4 \\ &= O(n^{2\alpha+4\beta}) \end{aligned}$$

uniformly in $t \leq n$. Employing the decomposition (74) when $\alpha \leq \beta$

$$\max_{1 \leq t \leq n} \mathbb{E} \|\tilde{z}_t\|^4 \leq 8 \max_{1 \leq t \leq n} \mathbb{E} \|x_t\|^4 + \frac{8 \|C_z\|^4}{n^{4\beta}} \max_{1 \leq t \leq n} \mathbb{E} \|\psi_{nt}\|^4 = O(n^{2\alpha})$$

by Lemma 3.2(i) and (76). Employing the decomposition (75) when $\alpha > \beta$

$$\max_{1 \leq t \leq n} \mathbb{E} \|\tilde{z}_t\|^4 \leq 8 \max_{1 \leq t \leq n} \mathbb{E} \|z_t\|^4 + \frac{8 \|C\|^4}{n^{4\alpha}} \max_{1 \leq t \leq n} \mathbb{E} \|\psi_{nt}\|^4 = O(n^{2\beta})$$

by Lemma 3.2(i) (since z_t is n^β -near stationary) and (76).

For part (ii), denoting $\underline{x}_{t-1} = x_{t-1} - \bar{x}_{n-1}$ and $\underline{\varepsilon}_t = \varepsilon_t - \bar{\varepsilon}_n$, using the identity

$$\hat{\underline{\varepsilon}}_t = \underline{y}_t - \hat{A}_n \underline{x}_{t-1} = \underline{\varepsilon}_t - (\hat{A}_n - A) \underline{x}_{t-1}$$

and the fact that the OLS estimator satisfies $\|\hat{A}_n - A\| = O_p(n^{-(1+\alpha)/2})$ we obtain that

$$\frac{1}{n^{1+\alpha\wedge\beta}} \hat{\Upsilon}_n = \frac{1}{n^{1+\alpha\wedge\beta}} \sum_{t=1}^n (\tilde{z}_{t-1} \tilde{z}'_{t-1} \otimes \underline{\varepsilon}_t \underline{\varepsilon}'_t) + o_p(1) \quad (77)$$

provided that both

$$\begin{aligned} r_{1n} &= \frac{1}{n^{(1+\alpha)/2}} \frac{1}{n^{1+\alpha\wedge\beta}} \sum_{t=1}^n (\tilde{z}_{t-1} \tilde{z}'_{t-1} \otimes \underline{x}_{t-1} \underline{\varepsilon}'_t) \\ r_{2n} &= \frac{1}{n^{1+\alpha}} \frac{1}{n^{1+\alpha\wedge\beta}} \sum_{t=1}^n (\tilde{z}_{t-1} \tilde{z}'_{t-1} \otimes \underline{x}_{t-1} \underline{x}'_{t-1}) \end{aligned}$$

are $o_p(1)$. First note that,

$$\begin{aligned}
\mathbb{E} \|\underline{x}_t\|^4 &\leq \mathbb{E} (\|x_t\| + \|\bar{x}_n\|)^4 \\
&\leq 8 (\mathbb{E} \|x_t\|^4 + \mathbb{E} \|\bar{x}_n\|^4) \\
&\leq 8 \left\{ \mathbb{E} \|x_t\|^4 + \frac{1}{n^4} \sum_{i,j,k,l=1}^n \mathbb{E} (\|x_i\| \|x_j\| \|x_k\| \|x_l\|) \right\} \\
&\leq 8\mathbb{E} \|x_t\|^4 + \frac{8}{n^4} \sum_{i,j,k,l=1}^n \left\{ \mathbb{E} (\|x_i\|^2 \|x_j\|^2) \mathbb{E} (\|x_k\|^2 \|x_l\|^2) \right\}^{1/2} \\
&\leq 8\mathbb{E} \|x_t\|^4 + 8 \max_{1 \leq i \leq n} \mathbb{E} \|x_i\|^4
\end{aligned}$$

so that

$$\max_{1 \leq t \leq n} \mathbb{E} \|\underline{x}_t\|^4 \leq 16 \max_{1 \leq t \leq n} \mathbb{E} \|x_t\|^4.$$

For r_{1n} , part (i) and Lemma 3.2(i) give

$$\begin{aligned}
\mathbb{E} \|r_{1n}\| &\leq \frac{1}{n^{(1+\alpha)/2}} \frac{1}{n^{\alpha\wedge\beta}} \max_{t \leq n} \mathbb{E} (\|\tilde{z}_{t-1}\|^2 \|\underline{x}_{t-1}\| \|\underline{\varepsilon}_t\|) \\
&\leq \frac{1}{n^{(1+\alpha)/2}} \frac{1}{n^{\alpha\wedge\beta}} \max_{t \leq n} (\mathbb{E} \|\tilde{z}_{t-1}\|^4)^{1/2} \left(\mathbb{E} \|\underline{x}_{t-1}\|^2 \|\underline{\varepsilon}_t\|^2 \right)^{1/2} \\
&\leq 4 \|\varepsilon_1\|_{L_4} \frac{1}{n^{(1+\alpha)/2} n^{\alpha\wedge\beta}} \left(\max_{t \leq n} \mathbb{E} \|\tilde{z}_{t-1}\|^4 \right)^{1/2} \left(\max_{t \leq n} \mathbb{E} \|\underline{x}_{t-1}\|^4 \right)^{1/4} \\
&= O\left(\frac{n^{\alpha\wedge\beta} n^{\alpha/2}}{n^{(1+\alpha)/2} n^{\alpha\wedge\beta}} \right) = O\left(\frac{1}{\sqrt{n}} \right)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E} \|r_{2n}\| &\leq \frac{1}{n^{1+\alpha}} \frac{1}{n^{\alpha\wedge\beta}} \max_{t \leq n} \mathbb{E} (\|\tilde{z}_{t-1}\|^2 \|\underline{x}_{t-1}\|^2) \\
&\leq \frac{1}{n^{1+\alpha}} \frac{1}{n^{\alpha\wedge\beta}} \left\{ \max_{t \leq n} \mathbb{E} \|\tilde{z}_{t-1}\|^4 \right\}^{1/2} \left\{ \max_{t \leq n} \mathbb{E} \|\underline{x}_{t-1}\|^4 \right\}^{1/2} \\
&= \frac{1}{n^{1+\alpha}} \frac{1}{n^{\alpha\wedge\beta}} O(n^{\alpha\wedge\beta}) O(n^\alpha) = O\left(\frac{1}{n} \right).
\end{aligned}$$

Returning to (77), we can write

$$\begin{aligned}
\frac{1}{n^{1+\alpha\wedge\beta}} \hat{\Upsilon}_n &= \frac{1}{n^{1+\alpha\wedge\beta}} \sum_{t=1}^n (\tilde{z}_{t-1} \tilde{z}'_{t-1} \otimes \varepsilon_t \varepsilon'_t) + O_p\left(\frac{1}{\sqrt{n}} \right) \\
&= \begin{cases} n^{-1-\beta} \sum_{t=1}^n (z_{t-1} z'_{t-1} \otimes \varepsilon_t \varepsilon'_t) + O_p\left(n^{-\frac{\alpha-\beta}{2}} \right), & \beta < \alpha \\ n^{-1-\alpha} \sum_{t=1}^n (x_{t-1} x'_{t-1} \otimes \varepsilon_t \varepsilon'_t) + O_p\left(n^{-\frac{\beta-\alpha}{2}} \right), & \alpha < \beta \end{cases} \quad (78)
\end{aligned}$$

from which the result for $\alpha > 0$ follows by Lemma 3.3(i) and Lemma 3.5(i), since z_t is a n^β -near stationary process. For $\alpha = 0$, denoting by $x_{0t} = \sum_{j=0}^{\infty} R^j u_{t-j}$ a strictly stationary version of the stable autoregression $x_t = R^t x_0 + \sum_{j=0}^{t-1} R^j u_{t-j}$,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \|x_{t-1} - x_{0t-1}\| \|x_{t-1}\| \|\varepsilon_t\|^2 &\leq \frac{1}{n} \|x_0\| \sum_{t=1}^n \|R\|^t \|x_{t-1}\| \|\varepsilon_t\|^2 \\ &\quad + \frac{1}{n} \sum_{t=1}^n \sum_{j=t}^{\infty} \|R\|^j \|u_{t-j}\| \|x_{t-1}\| \|\varepsilon_t\|^2 \\ &= O_p\left(\frac{1}{n}\right) \end{aligned}$$

the second part of (78) gives

$$\frac{1}{n} \hat{\Upsilon}_n = \frac{1}{n} \sum_{t=1}^n (x_{0t-1} x'_{0t-1} \otimes \varepsilon_t \varepsilon'_t) + o_p(1) \rightarrow_{a.s.} \mathbb{E} (x_{0t-1} x'_{0t-1} \otimes \varepsilon_t \varepsilon'_t)$$

by the ergodic theorem.

It remains to show (78). For the $\alpha > \beta$ part, the decomposition (75) implies that

$$\begin{aligned} r_{1n} &= \frac{1}{n^{1+\beta+\alpha}} \sum_{t=1}^n \|z_{t-1}\| \|\psi_{nt-1}\| \|\varepsilon_t\|^2 \rightarrow_p 0 \\ r_{2n} &= \frac{1}{n^{1+\beta+2\alpha}} \sum_{t=1}^n \|\psi_{nt-1}\|^2 \|\varepsilon_t\|^2 \rightarrow_p 0 \end{aligned}$$

are sufficient for (78). Using (76) and Lemma 3.2(i), we obtain

$$\begin{aligned} \mathbb{E} r_{1n} &\leq \frac{1}{n^{\beta+\alpha}} \left(\max_{t \leq n} \mathbb{E} \|z_t\|^4 \right)^{1/4} \left(\max_{t \leq n} \mathbb{E} \|\psi_{nt}\|^4 \right)^{1/4} (\mathbb{E} \|\varepsilon_1\|^4)^{1/2} \\ &= O\left(\frac{n^{\beta/2} n^{\alpha/2+\beta}}{n^{\beta+\alpha}}\right) = O\left(\frac{1}{n^{(\alpha-\beta)/2}}\right) \end{aligned}$$

$$\begin{aligned} \mathbb{E} r_{2n} &\leq \frac{1}{n^{\beta+2\alpha}} \left(\max_{t \leq n} \mathbb{E} \|\psi_{nt}\|^4 \right)^{1/2} (\mathbb{E} \|\varepsilon_1\|^4)^{1/2} \\ &= O\left(\frac{n^{\alpha+2\beta}}{n^{\beta+2\alpha}}\right) = O\left(\frac{1}{n^{\alpha-\beta}}\right) \end{aligned}$$

which proves (78) for $\alpha > \beta$. For $\alpha < \beta$, the same argument can be applied to the decomposition (74).

Proof of Theorem 4.4. When $\alpha > 0$, Lemma 4.3 shows that

$$\frac{1}{n^{1+\alpha\wedge\beta}} \hat{\Upsilon}_n = \left(\frac{1}{n^{1+\alpha\wedge\beta}} \sum_{t=1}^n \tilde{z}_{t-1} \tilde{z}'_{t-1} \right) \otimes \Sigma_{\varepsilon\varepsilon} + o_p(1)$$

which implies that

$$n^{1+\alpha\wedge\beta} \left\| \hat{Q}_n - (\underline{X}' P_{\tilde{Z}} \underline{X})^{-1} \otimes \Sigma_{\varepsilon\varepsilon} \right\| = o_p(1)$$

and $\left\| \hat{W}_n - \tilde{W}_n \right\| = o_p(1)$ under the null hypothesis (30). Corollary 4.2 then gives $\hat{W}_n \Rightarrow \chi^2(q)$.

When $\alpha = 0$, Lemma B2 of KMS (2015b) implies that $n^{-1} \left\| \underline{X}' \tilde{Z} - X'X \right\| = o_p(1)$. Combined with Lemma 4.3, this yields

$$\begin{aligned} n\hat{Q}_n &= \left[\left(\frac{1}{n} \tilde{Z}' \underline{X} \right)^{-1} \otimes I_m \right] \frac{1}{n} \hat{\Upsilon}_n \left[\left(\frac{1}{n} \underline{X}' \tilde{Z} \right)^{-1} \otimes I_m \right] \\ &= \left[\left(\frac{1}{n} X'X \right)^{-1} \otimes I_m \right] \frac{1}{n} \hat{\Upsilon}_n \left[\left(\frac{1}{n} X'X \right)^{-1} \otimes I_m \right] + o_p(1) \\ &= V_0 + o_p(1) \end{aligned}$$

where the matrix V_0 is defined in (40). We can then write $\hat{W}_n = w'_n w_n + o_p(1)$ where, under (30),

$$\begin{aligned} w_n &= \left(Hn\hat{Q}_nH' \right)^{-1/2} H \text{vec} \left[\sqrt{n} \left(\tilde{A}_{IVX} - A \right) \right] \\ &\Rightarrow N(0, I_q) \end{aligned}$$

by Theorem 4.1(iv).

References

- Abadir, K.M. and J. R. Magnus (2005). *Matrix Algebra*. Econometric Exercises, vol.1. Cambridge University Press.
- Billingsley, P. (1968). *Convergence of probability measures*. Wiley.
- Bollerslev (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31, 307-327.
- Boussama, F. (2006). Ergodicité des chaînes de Markov à valeurs dans une variété algébrique: application aux modèles GARCH multivariés. *Comptes Rendus de l'Académie des Sciences Paris*, 343, 275-278.

- Chan, N. H. and C. Z. Wei (1987). “Asymptotic inference for nearly nonstationary AR(1) processes, *Annals of Statistics* 15, 1050–1063.
- R. F. Engle (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, 50, 987-1008.
- C. Francq and J-M. Zakoian (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley
- A. Kostakis, T. Magdalinos and M.P. Stamatogiannis (2015a). Robust Econometric Inference for Stock Return Predictability. *Review of Financial Studies* 28 (5): 1506-1553.
- A. Kostakis, T. Magdalinos and M.P. Stamatogiannis (2015b). Online Appendix to: Robust Econometric Inference for Stock Return Predictability.
- Ling, S. and W.K. Li (1997a). Estimating and testing for unit root processes with GARCH(1,1) errors. Technical report, Department of Statistics, University of Hong Kong.
- Ling, S. and W.K. Li (1997b). On fractionally integrated autoregressive moving average time series models with conditional heteroskedasticity. *Journal of the American Statistical Association* 92, 1184-1194.
- Ling, S. and M. McAleer (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 280-310.
- Magdalinos, T. and P.C.B. Phillips (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory* 25, 482-526.
- Phillips, P.C.B. and T. Magdalinos (2009). Econometric inference in the vicinity of unity. Working paper.
- D.L. McLeish (1974). Dependent central limit theorems and invariance principles, *Annals of Probability* 2, 620–628.
- Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and its Application*. Academic Press.
- R.A. Horn and C.R. Johnson (2013). *Matrix Analysis*. Cambridge University Press.
- Pantula, S.G. (1989) Estimation of autoregressive models with ARCH errors. *Sankhya* B 50, 119-38.
- Phillips, P. C. B. (1987a). “Towards a unified asymptotic theory for autoregression,” *Biometrika* 74, 535–547.

- Phillips, P. C. B. (1987b). “Time series regression with a unit root,” *Econometrica*, 55, 277–302.
- Phillips, P. C. B. (1988). Regression Theory for Near-Integrated Time Series, *Econometrica*, 56, 1021-1044.
- Phillips, P. C. B. and T. Magdalinos (2007a), “Limit theory for Moderate deviations from a unit root.” *Journal of Econometrics*, 136, 115-130.
- Phillips, P. C. B. and T. Magdalinos (2007b), “Limit theory for Moderate deviations from a unit root under weak dependence.” in G. D. A. Phillips and E. Tzavalis (Eds.) *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*. Cambridge University Press.
- Phillips, P. C. B. and V. Solo (1992), “Asymptotics for Linear Processes,” *Annals of Statistics* 20, 971–1001.
- D. Serre (2010). *Matrices*. Graduate Texts in Mathematics 216, Springer.
- H. White (1980). A Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817-838.