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

Tassos Magdalinos
University of Southampton

The paper examines the effect of conditional heteroskedasticity on least squares inference in stochastic regression models of unknown integration order and proposes an inference procedure that is robust to models within the (near) I(0)–(near) I(1) range with GARCH innovations. We show that a regressor signal of exact order $O_p(n\kappa_n)$ for arbitrary $\kappa_n \to \infty$ is sufficient to eliminate stationary GARCH effects from the limit distributions of least squares based estimators and self-normalized test statistics. The above order dominates the $O_p(n)$ signal of stationary regressors but may be dominated by the $O_p(n^2)$ 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 normalized test statistics based on the ordinary least squares estimator when $\kappa_n \to \infty$ and $\kappa_n/n \to 0$ and the IVX estimator (Phillips and Magdalinos (2009a), Econometric Inference in the Vicinity of Unity. Working paper, Singapore Management University; Kostakis, Magdalinos, and Stamatogiannis, 2015a, *Review of Financial Studies* 28(5), 1506–1553.) when $\kappa_n \to \infty$ and the innovation sequence of the system is a covariance stationary vec-GARCH process. An adjusted version of the IVX–Wald test is shown to also accommodate GARCH effects in purely stationary regressors, thereby extending the procedure's validity over the entire (near) I(0)–(near) I(1) range of regressors under conditional heteroskedasticity in the innovations. It is hoped that the wide range of applicability of this adjusted IVX–Wald test, established in Theorem 4.4, presents an advantage for the procedure's suitability as a tool for applied research.

This paper is dedicated to Peter Phillips with gratitude for his mentorship, friendship and generosity over the years.

1. INTRODUCTION

The effect of conditional heteroskedasticity in autoregressive and stochastic regression models has been a topic of intense research activity since the introduction

© The Author(s), 2021. Published by Cambridge University Press.

875

of ARCH and GARCH processes by Engle (1982) and Bollersev (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 nonstationary times series with GARCH innovations: see Ling and Li (1997b), Ling and Li (1998), Ling and McAleer (2003) and references therein. More recently, Andrews and Guggenberger (2012) and Andrews and Guggenberger (2014) have established asymptotic theory for the ordinary least squares (OLS) and (feasible) GLS estimator of autoregressive models with conditionally heteroskedastic errors as well as a conditional-heteroskedasticity-robust confidence interval for the parameter of an AR(1) process.

The 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 (Phillips, 1987b). 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-I(1) regressor has sufficient signal, of order $O_p\left(n^2\right)$, to asymptotically eliminate stationary GARCH effects; such elimination cannot be achieved by the weaker $O_p\left(n\right)$ signal of a stationary regressor, resulting in 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 (Phillips and Magdalinos, 2007a & Phillips and Magdalinos, 2007b) in the case of scalar autoregressions and Magdalinos and Phillips (2009), henceforth MP (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 $\rho_n = 1 + c/k_n$, where $c < 0$, $k_n \to \infty$ and $k_n/n \to 0$. The signal generated by such processes is of order $O_p(nk_n)$ and may approach the stationary $O_p(n)$ signal for sequences k_n diverging to ∞ at sufficiently slow rate.

The present work develops a limit theory for near-stationary predictive regression systems with covariance stationary GARCH innovations. We show that, for arbitrary $k_n \to \infty$, 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 MP (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. Subsequent to the result of Andrews and Guggenberger (2012) for a (scalar) AR(1) process, the current paper provides the first instance of standard Gaussian and chi-squared asymptotics applying respectively to the OLS estimator and the Wald statistic in a vector autoregression or predictive regression model with conditionally heteroskedastic innovations.

The development of least squares limit theory for the case of near stationary regressors is the key step toward extending the validity of the IVX endogenous instrumentation procedure, introduced by Phillips and Magdalinos (2009a) and further developed by Kostakis et al. (2015a) (henceforth PM (2009a) 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, near stationary and near integrated time series regressors. In the current 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; as a result, the IVX estimator inherits the usual GARCH effects present in the asymptotic variance of the least squares estimator. A White (1980) type correction is shown to make the IVX procedure operational for all persistence regimes in the I(0)–I(1) range under conditional heteroskedasticity. The adjusted IVX–Wald test, presented in Theorem 4.4, is the paper's main methodological contribution, accommodating a sufficiently general class of models to provide a tool for applied research.

The paper is organized as follows. Section 2 outlines a general modeling framework for a system of predictive regressions with unknown persistence properties and conditionally heteroskedastic innovations of a covariance stationary vec-GARCH type. Section 3 develops a limit theory for the OLS estimator in the near stationary 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 and concluding remarks and Section 6 includes all proofs.

2. PREDICTIVE REGRESSION AND VECTOR AUTOREGRESSION WITH GARCH INNOVATIONS

We consider a first-order vector autoregression

$$x_t = R_n x_{t-1} + u_t, \quad t \in \{1, ..., n\} \tag{1}$$

with conditionally heteroskedastic innovations u_t and autoregressive matrix R_n that induces persistence characteristics ranging from stability to unit root nonstationarity and includes intermediate integration regimes, as specified by Assumption P below. The VAR process in (1) may be regarded as the statistical model of interest or as a data generating mechanism for a predictive regression

$$y_t = \mu + A x_{t-1} + \epsilon_t, \tag{2}$$

where A is an $m \times r$ coefficient matrix. In the former case, a multivariate GARCH parametrization will be imposed on the innovations in (1) and inference will be conducted on the autoregressive matrix in (6). In the predictive regression case, the coefficient matrix A in (2) is the parameter of interest and a GARCH parametrization will be assumed for the innovations of (2), while no parametric specification of the conditional heteroskedasticity of the innovations of the VAR process in (1) will be required. For clarity of exposition, we provide the form of the vec-GARCH parametric specification for a generic sequence (ε_t), ahead of stating the assumptions on the innovations u_t and ϵ_t of (1) and (2).

Assumption G. Let $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ be a filtration and $(\eta_t)_{t \in \mathbb{Z}}$ be an $\mathcal{F}_t$-adapted sequence of i.i.d. random vectors with $\mathbb{E}(\eta_1) = 0$ and $\mathbb{E}(\eta_1 \eta_1') = I$. Let $(\varepsilon_t)_{t \in \mathbb{Z}}$ be a sequence admitting the following covariance stationary vec-GARCH (p, q) representation:

$$\varepsilon_t = H_t^{1/2} \eta_t, \quad \text{vech}(H_t) = \varphi + \sum_{i=1}^{q} A_i \text{vech}\left(\varepsilon_{t-i} \varepsilon_{t-i}'\right) + \sum_{k=1}^{p} B_k \text{vech}(H_{t-k}) \tag{3}$$

with $E\|\varepsilon_1\|^4 < \infty$, 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 $\rho(\Gamma) < 1$.

Assumption G accounts for conditionally heteroskedastic innovations of a very general parametric form: the vec-GARCH process 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 (3) and the condition on the spectral radius of their sum are related to 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 adaptation property of the GARCH recursion equation in (3) imply that $(\varepsilon_t, \mathcal{F}_t)$

is a martingale difference sequence satisfying

$$\mathbb{E}_{\mathcal{F}_{t-1}}\left(\varepsilon_t\varepsilon_t'\right) = H_t. \tag{4}$$

The most restrictive condition of Assumption G is the requirement of finite fourth moments, which has well-documented implications on GARCH parametrization.

The stochastic properties of the innovation sequences (ϵ_t) and (u_t) are summarized by the following assumption.

Assumption M.

(i) Let the statistical model be given by the predictive regression system (1)–(2) with $v_t = \left[u_t', \epsilon_t'\right]'$ a strictly stationary process satisfying $\mathbb{E}v_1 v_1' = \Sigma_{vv} > 0$. The sequence ϵ_t in (2) satisfies Assumption G with $\epsilon_t = \varepsilon_t$ and ε_t given by (3). The sequence u_t in (1) is a linear process

$$u_t = \sum_{j=0}^{\infty} F_j e_{t-j} \quad \sum_{j=0}^{\infty} j \left\| F_j \right\|^2 < \infty, \tag{5}$$

where $(e_t)_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic $\mathcal{F}_t$-martingale difference sequence satisfying $\mathbb{E}\left\| e_1 \right\|^4 < \infty$, $\left(F_j\right)_{j\geq 0}$ is a sequence of constant matrices such that $F(1) = \sum_{j=0}^{\infty} F_j$ has full rank and $F_0 = I_r$.

(ii) Let the statistical model be given by the VAR process (1). The sequence u_t in (1) satisfies Assumption G with $u_t = \varepsilon_t$ and ε_t given by (3).

Assumption M provides two separate sets of assumptions according to the statistical model under consideration: when the latter is given by the VAR(1) process in (1), the sequence of innovations (u_t) is assumed to be a vec-GARCH process of the form (3) of Assumption G. On the other hand, when considering a predictive regression system (1)–(2) as the statistical model, a parametric GARCH specification is imposed only the innovation sequence (ϵ_t) of the model equation (2); the innovation sequence (u_t) may exhibit conditional heteroskedasticity as a result of the potential conditional heteroskedasticity of the sequence (e_t) in (5), but there is no need to model this parametrically. The stationarity and ergodicity assumption of (e_t) ensures the validity of a strong law of large numbers for $n^{-1}\sum_{t=1}^{n} e_t e_t'$ and is satisfied when (e_t) is a stationary GARCH process of the form (3). We denote the autocovariance matrix of u_t by $\Gamma_u(j) = \mathbb{E}\left(u_1 u_{1-j}'\right)$ and the associated long run covariance $\Omega_{uu} = \sum_{j=-\infty}^{\infty} \Gamma_u(j) > 0$, the positive definiteness of Ω_{uu} ensured by Assumption M.

The persistence properties of the vector autoregression x_t are characterized through assumptions on an adjusted version of the autoregressive matrix R_n, that generalizes the parametrization employed by MP (2009) and PM (2009a),[1] introduced by Magdalinos and Phillips (2020), henceforth referred to as MP

---

[1] The original parametrization of the autoregressive matrix in MP (2009), PM (2009a), and KMS (2015a) took the form $R_n = I_r + C/n^{\alpha}$ with C a diagonal matrix and $\alpha \geq 0$. The current parametrizaton in Assumption P employs the more general parametrization of Magdalinos and Phillips (2020).

(2020), to systems with conditionally homoskedastic innovations. We distinguish between three classes of regressor processes, presented below.

Assumption P. The autoregressive matrix R_n in (1) satisfies

$$C_n := \kappa_n (R_n - I_r) \to C \ as \ n \to \infty \tag{6}$$

for some $r \times r$ matrix C satisfying $\|C\| < \infty$ and some sequence $(\kappa_n)_{n \in \mathbb{N}}$ of positive numbers. The regressor x_t in (1) belongs to one of the following classes:

(i) Near-nonstationary and nonstationary regressors, if (6) holds with $\kappa_n/n \to \kappa \in (0, \infty]$.
(ii) Near-stationary regressors, if (6) holds with $\kappa_n/n \to 0$, $\kappa_n \to \infty$ and C a negative stable[2] matrix.
(iii) Stationary regressors, if (6) holds with $\kappa_n = 1$ and $R = I_r + C$ has spectral radius $\rho(R) < 1$.

The process x_t in (1) is initialized at $x_0 = o_p\left(\kappa_n^{1/2}\right)$ with $(\kappa_n)_{n \in \mathbb{N}}$ satisfying (6) under Assumptions P(i)–(ii) and $x_0 = O_p(1)$ under Assumption P(iii).

Observe that (6) implies that the localizing coefficient matrix C_n of R_n is allowed to depend on the sample size n. In the near nonstationary regressor case P(i), quite general localizing coefficient matrices C_n are permitted with the sole condition $\|C\| < \infty$ that the limit matrix in (6) is bounded; the exact I(1) case is included in P(i) when $C = 0$. The negative stability requirement of Assumption P(ii) on the limiting matrix C in (6) implies restrictions on the sequences (C_n) and (R_n) that give rise to regressors with near-stationary characteristics; see the discussion following Assumption N and Lemma 2.1 of MP (2020). Moreover, the negative-stability property of C is a necessary and sufficient condition for the matrix

$$V_{xx} = \int_0^\infty e^{rC} \Omega_{uu} e^{rC'} dr \tag{7}$$

to be well defined (and positive definite since $\Omega_{uu} > 0$). Since V_{xx} is the probability limit of the sample moment matrix $n^{-1}\kappa_n^{-1} \sum_{t=1}^n x_{t-1} x_{t-1}'$ (Lemma 2.2(ii) of MP (2020) for the conditionally homoskedastic case and Lemma 3.2(ii) below), the requirement of Assumption P(ii) in the definition of a near-stationary process seems minimal. Assumption P(iii) covers the usual stable root regressor case. Assumption P allows for an initial condition $x_0(n)$ that takes the form of a random process (e.g., a linear process of past innovations $(u_t)_{t \leq 0}$) provided that the order of magnitude of $x_0(n)$ is dominated by that of x_n (strictly under P(i)-(ii) and weakly under P(iii)), a restriction that rules out distant and infinite past initializations (see Andrews and Guggenberger (2008) and Phillips and Magdalinos (2009b)) that contribute to least squares limit theory.

---

[2] A square matrix is negative stable if all its eigenvalues have negative real part.

Denoting the demeaned regression matrices in the system (1)–(2) by $\underline{Y} = \left(\underline{y}'_1, ..., \underline{y}'_n\right)'$, $\underline{X} = \left(\underline{x}'_0, ..., \underline{x}'_{n-1}\right)'$, 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 (2) is given, as in KMS (2015a), by

$$\hat{A}_n = \underline{Y}'\underline{X}\left(\underline{X}'\underline{X}\right)^{-1}. \tag{8}$$

The effect of GARCH innovations on the asymptotic theory of the least squares estimator of A is known to differ according to the persistence class of the regressor x_t in (1). For stationary processes in class (iii), $\sqrt{n}\,\text{vec}\left(\hat{A}_n - A\right)$ is asymptotically zero mean Gaussian with nonstandard asymptotic variance that depends on the GARCH parameters and the fourth moment of the innovations. As a result, the usual self-normalized 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 nonstandard limit distributions of $n\,\text{vec}\left(\hat{A}_n - A\right)$ in the unit root and local to unity cases (Phillips 1987, Phillips 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 $\sum_{t=1}^n x_{t-1}x'_{t-1} = O_p\left(n^2\right)$ 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 OLS estimator in the intermediate case of near-stationary regressors of class P(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 centered and scaled least squares regression estimate

$$\sqrt{n\kappa_n}\left(\hat{A}_n - A\right) = \left(\frac{1}{\sqrt{n\kappa_n}}\sum_{t=1}^n \epsilon_t x'_{t-1}\right)\left(\frac{1}{n\kappa_n}\sum_{t=1}^n x_{t-1}x'_{t-1}\right)^{-1} + O_p\left(\sqrt{\frac{\kappa_n}{n}}\right) \tag{9}$$

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 (2) 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 asymptotically.[3] Our approach follows MP (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 (9) 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 (9) when ϵ_t is a vec-GARCH process defined in (3). 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. Stability of the solution permits standard martingale approximation arguments that resolve the asymptotics in (9). The above analysis, summarized in Lemma 3.3, is the main technical contribution of the paper.

To fix ideas, we establish some notation for the recursive equations that we employ in the development of the asymptotic theory (see part (ii) of Lemma 3.3). Given the matrices $A_1, ..., A_q, B_1, ..., B_p$ in (3), define

$$
\Phi_i = \begin{cases} A_i + B_i, & \text{if } i \le p \wedge q \\ A_i, & \text{if } p < i \le q \\ B_i, & \text{if } q < i \le p \end{cases} \tag{10}
$$

$$
\Gamma_i = I_{r^2} \otimes \Phi_i, \quad \Gamma_{n,i} = R_n^i \otimes R_n^i \otimes \Phi_i \tag{11}
$$

and consider the stochastic difference equations:

$$
Y(j) = \sum_{l=1}^{\kappa} \Gamma_l Y(j-l) + v(j) \tag{12}
$$

$$
Y_n(j) = \sum_{l=1}^{\kappa} \Gamma_{n,l} Y_n(j-l) + v_n(j) \tag{13}
$$

for $j \ge 1$ and $\kappa := q \vee p$, with $v(j)$ and $v_n(j)$ denoting generic innovations (to be specified by (24) and (25) below). The companion matrix associated with (13) is given by

$$
M_{n,\kappa} = \begin{bmatrix} \Gamma_{n,1} & \Gamma_{n,2} & ... & \Gamma_{n,\kappa-1} & \Gamma_{n,\kappa} \\ I & 0 & ... & 0 & 0 \\ 0 & I & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & ... & 0 & I & 0 \end{bmatrix}, \tag{14}
$$

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 (12), denoted by M_κ, has the same form as the matrix in (14) with $\Gamma_{n,i}$ replaced by Γ_i for all $i \in \{1, ..., \kappa\}$. It is relatively

---

[3] See the proof of Theorem 3.4 in the Appendix for details and the proof of (9).

straightforward to show from first principles that Assumption G on the GARCH process ensures the stability of the solution of (12) and (13).

LEMMA 3.1. *Under Assumption* G, *the spectral radius and norm of the companion matrices* $M_{n,\kappa}$ *and* M_κ *defined in* (14) *satisfy:*

(i) $\rho(M_\kappa) < 1$ *and* $\sum_{j=0}^\infty \left\| M_\kappa^j \right\| < \infty$ *and*

(ii) $\lim_{n\to\infty} \rho(M_{n,\kappa}) < 1$ *and* $\sup_{n\geq 1} \sum_{j=0}^\infty \left\| M_{n,\kappa}^j \right\| < \infty$.

We start by listing some properties of near-stationary processes generated by potentially conditionally heteroskedastic martingale difference innovations (e_t) that have finite fourth order moments. Some results are direct generalizations of the conditionally homoskedastic case of MP (2009). In particular, we show that the sample moment matrix $n^{-1}\kappa_n^{-1}\sum_{t=1}^n x_t x_t'$ has the same probability limit as in the case of a near-stationary regressor generated by a conditionally homoskedastic martingale difference (e_t).

LEMMA 3.2. *Denote by* $x_{0t} = \sum_{j=1}^t R_n^{t-j} u_j$ *the regressor in* (1) *with* $x_0 = 0$ *and define the process*

$$\zeta_{n,t} = \sum_{j=0}^{t-1} R_n^j F(1) e_{t-j}. \tag{15}$$

Under Assumptions[4] *M and P(ii), the following hold:*

(i) $\max_{1\leq t\leq n} \mathbb{E}\left\| \kappa_n^{-1/2} x_{0t} \right\|^4 = O(1)$ *and* $\max_{1\leq t\leq n} \mathbb{E}\left\| \kappa_n^{-1/2} \zeta_{n,t} \right\|^4 = O(1)$.

(ii) *Both* $n^{-1}\kappa_n^{-1}\sum_{t=1}^n x_{t-1} x_{t-1}'$ *and* $n^{-1}\kappa_n^{-1}\sum_{t=1}^n \zeta_{n,t-1}\zeta_{n,t-1}'$ *converge in probability to the matrix* V_{xx} *in* (7).

(iii) $(n\kappa_n)^{-1/2}\left\| \sum_{t=1}^n (x_{t-1}\otimes\varepsilon_t) - \sum_{t=1}^n (\zeta_{n,t-1}\otimes\varepsilon_t) \right\| = o_p(1)$.

Lemma 3.2 shows that stationary GARCH effects are eliminated from the first order asymptotics of the denominator of the matrix quotient (9), for a near stationary regressor x_t of arbitrary order. The asymptotic development of near-stationary regression theory so far was based on unconditional moment bounds and truncation and did not employ any properties of the GARCH specification (3). Obtaining the limit distribution of the martingale transform in the numerator of the matrix quotient (9), asymptotically equivalent to

$$N_n = \frac{1}{\sqrt{n\kappa_n}}\sum_{t=1}^n (\zeta_{n,t-1}\otimes\varepsilon_t) \tag{16}$$

---

[4]Under Assumption M(ii), $u_t = e_t = \varepsilon_t$, so $C(1) = I$ in the definition of $\zeta_{n,t}$

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

$$\langle N \rangle_n = \frac{1}{n\kappa_n} \sum_{t=1}^{n} \left(\zeta_{n,t-1} \zeta'_{n,t-1} \otimes \mathbb{E}_{\mathcal{F}_{t-1}} \varepsilon_t \varepsilon'_t \right) = \frac{1}{n\kappa_n} \sum_{t=1}^{n} \left(\zeta_{n,t-1} \zeta'_{n,t-1} \otimes H_t \right) \tag{17}$$

with H_t defined in (3), can be approximated by

$$V_n = \frac{1}{n\kappa_n} \sum_{t=1}^{n} \left(\zeta_{n,t-1} \zeta'_{n,t-1} \otimes \Sigma_{\varepsilon\varepsilon} \right), \quad \Sigma_{\varepsilon\varepsilon} = \mathbb{E}\varepsilon_1 \varepsilon'_1 \tag{18}$$

with approximation error expressed in terms of the solutions of the stochastic recurrence relations (12) and (13) arising from (3), and that the stability of these solutions implies the asymptotic negligibility of the approximation error. The approximation of $\langle N \rangle_n$ by V_n and the characterization of the approximation error in terms of bounds that depend on the solutions of the stochastic difference equations (12) and (13), summarized by the next result, is the main technical contribution of the paper.

LEMMA 3.3. *Consider the vector-valued processes*

$$\Sigma_t(j) = vec\left(e_t e'_t - \Sigma_{ee}\right) \otimes vech\left(H_{t+j}\right) \tag{19}$$

$$S_{n,t}(j) = R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \otimes vech\left(H_{t+j}\right) \tag{20}$$

with H_t defined in (3) and

$$w_t = vech\left(\varepsilon_t \varepsilon'_t - H_t\right). \tag{21}$$

 (i) *Under Assumptions M and P(ii), $\langle N \rangle_n$ and V_n in (17) and (18) satisfy*

$$\|\langle N \rangle_n - V_n\| \leq b\left(\sigma_n + s_n\right) + o_p(1) \tag{22}$$

as $n \to \infty$, where

$$\sigma_n = \frac{1}{n\kappa_n} \sum_{j=1}^{n-1} \|R_n\|^{2(j-1)} \left\| \sum_{t=1}^{n-j} \Sigma_t(j) \right\|, \quad s_n = \frac{1}{n\kappa_n} \left\| \sum_{j=1}^{n-1} \sum_{t=2}^{n-j} S_{n,t}(j) \right\| \tag{23}$$

and $b \in (0, \infty)$ is a constant independent of n.

 (ii) *For each $j \geq 1$ and fixed t, n: $\Sigma_t(j)$ satisfies (12) with innovations $v(j) \equiv v_t(j)$ given by*

$$v_t(j) = vec\left(e_t e'_t - \Sigma_{ee}\right) \otimes \varphi + \sum_{l=1}^{q} (I \otimes A_l) \left[vec\left(e_t e'_t - \Sigma_{ee}\right) \otimes w_{t+j-l}\right] \tag{24}$$

and $S_{n,t}(j)$ satisfies (13) with innovations $v_n(j) \equiv v_{n,t}(j)$ given by

$$\upsilon_{n,t}(j) = R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \otimes \varphi$$

$$+ \sum_{i=1}^{q} (I \otimes A_i) \left[R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \otimes w_{t+j-i} \right]. \tag{25}$$

(iii) *Under Assumptions M and P(ii), the bounding sequences in (22) satisfy* $\sigma_n \to_p$ 0 *and* $s_n \to_p 0$.

By Lemma 3.3, the processes in (19) and (20) can be expressed as the companion form solutions of the stochastic recurrence relations (12) and (13), see (59) and (60) in the Appendix. The leading terms of these solutions consist of "moving averages" of the martingale difference sequences (24) and (25) 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 (22) are asymptotically negligible.

The above lemma implies that the predictable quadratic variation of the martingale transform in (16) with ε_t following the vec-GARCH process (3) can be approximated by its counterpart when ε_t is conditionally homoskedastic. Combined with a Lindeberg condition established in the Appendix, a standard martingale central limit theorem applies to the numerator of the matrix quotient (9), 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

$$W_n = \left(H \mathrm{vec} \hat{A}_n - h \right)' \left\{ H \left[(\underline{X}'\underline{X})^{-1} \otimes \hat{\Sigma}_{\varepsilon\varepsilon} \right] H' \right\}^{-1} \left(H \mathrm{vec}\, \hat{A}_n - h \right) \tag{26}$$

for testing linear restrictions on the coefficient matrix

$$H_0 : H \mathrm{vec}\,(A) = h, \tag{27}$$

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 (26) 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 (2): $\hat{\varepsilon}_t = y_t - \bar{y}_n - \hat{A}_n (x_{t-1} - \bar{x}_{n-1})$. These results are summarized below.

THEOREM 3.4. *Consider the system of predictive regressions (1), (2), and (6) under Assumption P(ii) with ϵ_t and u_t satisfying Assumption M(i). The following limits apply as $n \to \infty$:*

(i) $(n\kappa_n)^{-1/2} \sum_{t=1}^{n} (x_{t-1} \otimes \epsilon_t) \Rightarrow N(0, V_{xx} \otimes \Sigma_{\epsilon\epsilon})$

(ii) $\sqrt{n\kappa_n}\, vec\left(\hat{A}_n - A \right) \Rightarrow N\left(0, V_{xx}^{-1} \otimes \Sigma_{\epsilon\epsilon} \right)$

(iii) $W_n \Rightarrow \chi^2(q)$, *under (27)*

where q is the rank of H in (27), V_{xx} is defined in (7) and $\Sigma_{\epsilon\epsilon} = \mathbb{E}\epsilon_1 \epsilon_1'$.

Analogous results apply to the OLS estimator $\hat{R}_n$ when the statistical model is given by the vector autoregressive process in (1), and u_t are martingale difference innovations that satisfy the vec-GARCH recursion (3), as specified by Assumption

M(ii). 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

$$W_n^R = \left(H_R \text{vec} \hat{R}_n - h_R\right)' \left\{H_R\left[(X'X)^{-1} \otimes \hat{\Sigma}_{uu}\right]H_R'\right\}^{-1}\left(H_R \text{vec} \hat{R}_n - h_R\right), \qquad \textbf{(28)}$$

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

THEOREM 3.5. *Consider the vector autoregression (1) and (6) under Assumption P(ii) with u_t satisfying Assumption M(ii). The following limits apply as $n \to \infty$:*

(i) $(n\kappa_n)^{-1/2}\sum_{t=1}^{n}(x_{t-1} \otimes u_t) \Rightarrow N(0, V_{xx} \otimes \Sigma_{uu})$
(ii) $\sqrt{n\kappa_n}\, vec\left(\hat{R}_n - R_n\right) \Rightarrow N\left(0, V_{xx}^{-1} \otimes \Sigma_{uu}\right)$
(iii) $W_n^R \Rightarrow \chi^2(q)$,

where q is the rank of H_R and V_{xx} is defined in (7).

Remark 3.6.

(i) Theorems 3.4 and 3.5 provide a full characterization of the effect of GARCH innovations in stochastic regression models by considering regressors with a 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

$$\sum_{t=1}^{n}x_{t-1}x_{t-1}' = O_p(n\kappa_n), \qquad \textbf{(29)}$$

where $\kappa_n \to \infty$ at arbitrary rate is sufficient to asymptotically eliminate stationary GARCH effects from the distribution of the least squares estimator and the associated self-normalized 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.
(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 (29), while the order of the conditionally heteroskedastic innovations remains I(0).
(iii) Andrews and Guggenberger (2012), herafter AG (2012), establish OLS and GLS limit theory in an AR(1) process with a root that admits an equivalent parametrization to (6) for scalar R_n and C. Their results include a univariate version of Theorem 3.5 obtained under a strong mixing assumption instead of a GARCH parametrization as in the current paper. While avoiding a parametric specification of the conditional heteroskedasticity, AG (2012) impose stronger moment conditions on the innovation sequence u_t than

our Assumptions G, M(ii): Assumption INNOV(iii) of AG (2012) requires $\mathbb{E}\|u_1\|^{\zeta} < \infty$ with (at least) $\zeta > 6$ with further moment conditions imposed for the GARCH(1,1) example of equation (5) of that paper. In short, the OLS limit distribution result of AG (2012) coincides with a univariate version of Theorem 3.5 above and is derived by weak dependence arguments without imposing a GARCH parametrization on u_t at the expense of assuming higher order moments on u_t. On the other hand, the current paper makes use of the GARCH parametrization for the asymptotic development based on the recursive techniques of Lemmata 3.1 and 3.3 and requires four finite moments. In this sense, the OLS asymptotic results of AG (2012) and Theorem 3.5 above are complementary. It is worth noting the weak dependence properties of vector-valued conditionally heteroskedastic processes are not as well developed as in the scalar case: to our knowledge, there are no available results on the mixing properties of vec-GARCH models, with geometric ergodicity established by Boussama, Fuchs and Stelzer (2011) for the less general class of BEKK–GARCH models.

4. IVX LIMIT THEORY WITH GARCH INNOVATIONS

Having characterized the asymptotic behavior 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)–(2) 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 (2) across all persistence regimes P(i)–(iii) based on an endogenous instrumentation procedure, termed IVX, has been proposed by PM (2009a) 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. The main result of the section is Theorem 4.4 which establishes a testing procedure that is valid over the entire range of persistence regimes P(i)–(iii).

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

$$\tilde{z}_t = R_{nz}\tilde{z}_{t-1} + \Delta x_t, \quad \tilde{z}_0 = 0 \tag{30}$$

is generated according to an artificial autoregressive matrix

$$R_{nz} = I_r + \frac{C_z}{\kappa_{nz}}, \quad \frac{n^{1/2}}{\kappa_{nz}} + \frac{\kappa_{nz}}{n} \to 0 \tag{31}$$

with specified persistence degree κ_{nz}, where C_z is a negative stable matrix. The matrix A of coefficients in (2) is then estimated by a standard instrumental variable estimator that employs the instruments in (30):

$$\tilde{A}_{IVX} = \underline{Y}'\tilde{Z}\left(\underline{X}'\tilde{Z}\right)^{-1}. \tag{32}$$

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 behavior of the numerator of the matrix quotient in (32) is driven by the martingale transform

$$\tilde{N}_n = \frac{1}{\sqrt{n\left(\kappa_{nz} \wedge \kappa_{nz}\right)}} \sum_{t=1}^{n} \left(\tilde{z}_{t-1} \otimes \varepsilon_t\right) \tag{33}$$

with instrument process $\tilde{z}_t$ behaving asymptotically like a near stationary process of the type P(ii): when $\kappa_{nz} \ll \kappa_n$ (shorthand notation for $\kappa_{nz}/\kappa_n \to 0$) in which case the instruments are less persistent than the regressors, $\tilde{z}_{t-1}$ can be replaced asymptotically in (33) by z_{t-1}, where

$$z_t = R_{nz} z_{t-1} + u_t \tag{34}$$

a κ_{nz}-near-stationary process satisfying Assumption P(ii); when $\kappa_n \ll \kappa_{nz}$, employing more persistent instruments than the regressor in (1) leads to $\tilde{z}_t$ behaving asymptotically as the regressor x_t, a necessarily near-stationary process by the choice of κ_{nz} in (31), in which case $\tilde{N}_n$ in (33) is asymptotically equivalent to N_n in (16):

$$\left\| \tilde{N}_n - \frac{1}{\sqrt{n\kappa_{nz}}} \sum_{t=1}^{n} \left(z_{t-1} \otimes \varepsilon_t\right) \right\| = o_p\left(1\right) \text{ if } \kappa_{nz}/\kappa_n \to 0,$$

$$\left\| \tilde{N}_n - \frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^{n} \left(x_{t-1} \otimes \varepsilon_t\right) \right\| = o_p\left(1\right) \text{ if } \kappa_n/\kappa_{nz} \to 0,$$

see Lemma 3.2 MP (2020). In both cases, Lemma 3.3 above applies to the predictable quadratic variation $\left\langle \tilde{N}_n \right\rangle$ of the martingale transform in (33): denoting the moment matrices of z_{t-1} and x_{t-1} by $V_{zz}^{(n)} = n^{-1} \kappa_{nz}^{-1} \sum_{t=1}^{n} z_{t-1} z_{t-1}'$ and $V_{xx}^{(n)} = n^{-1} \kappa_n^{-1} \sum_{t=1}^{n} x_{t-1} x_{t-1}'$ Lemma 3.3 yields

$$\left\| \left\langle \tilde{N}_n \right\rangle - V_{zz}^{(n)} \otimes \Sigma_{\varepsilon\varepsilon} \right\| \to_p 0 \text{ if } \frac{\kappa_{nz}}{\kappa_n} \to 0 \text{ and } \left\| \left\langle \tilde{N}_n \right\rangle - V_{xx}^{(n)} \otimes \Sigma_{\varepsilon\varepsilon} \right\| \to_p 0 \text{ if } \frac{\kappa_n}{\kappa_{nz}} \to 0$$

with Lemma 3.2(ii) applying to $V_{xx}^{(n)}$ and $V_{zz}^{(n)}$. Moreover, the asymptotic approximation of $n^{-1} \left(\kappa_{nz} \wedge \kappa_{nz}\right)^{-1} \sum_{t=1}^{n} \tilde{z}_{t-1} \tilde{z}_{t-1}'$ by $V_{zz}^{(n)}$ when $\kappa_{nz}/\kappa_n \to 0$ and by $V_{xx}^{(n)}$ when $\kappa_n/\kappa_{nz} \to 0$ (Lemma 3.1(iv) of MP (2020)) as well as the limit distribution of $n^{-1} \left(\kappa_{nz} \wedge \kappa_{nz}\right)^{-1} \sum_{t=1}^{n} x_{t-1} \tilde{z}_{t-1}'$ (Lemma 3.1(iii) of MP (2020)) are derived through arguments that do not depend on the conditional homoskedasticity assumption maintained by MP (2020),[5] and hence remain valid under the current conditionally heteroskedastic framework of Assumption G. We conclude that the limit distribution of $\tilde{N}_n$ in (33) can be derived directly from Theorem 3.5(i) and the limit

---

[5]The proof of Lemma 3.2(ii) in the Appendix shows that Proposition A2 of MP (2020) continues to hold under Assumptions G and M, so parts (ii)–(iv) of Lemma 3.1 of MP (2020) continue to hold in the current context.

distribution of the remaining IVX sample moments can be deduced by Lemma 3.1 (iii) and (iv) of MP (2020).

Having characterized the asymptotic behavior of IVX sample moments, the asymptotic distribution of the normalized and centred IVX estimator follows as in PM (2009a) and MP (2020): we collect the results across the different persistence regimes of Assumption P in Theorem 4.1 below. Corresponding to the matrix V_{xx} in (7), we define the matrices

$$V_{zz} = \int_0^\infty e^{rC_z} \Omega_{uu} e^{rC_z'} dr \text{ and } V_{xz} = \int_0^\infty e^{rC} V_{xx} e^{rC_z'} dr \tag{35}$$

and, under Assumption P(iii), we denote a strictly stationary ergodic version of x_t by

$$\check{x}_t = \sum_{j=0}^\infty R^j u_{t-j}. \tag{36}$$

Finally, we denote by B_u an r-variate Brownian motion with covariance matrix Ω_{uu}, by $J_C(t) = \int_0^t e^{C(t-s)} dB_u(s)$ an Ornstein–Uhlenbeck process and by $\underline{J}_C(t) = J_C(t) - \int_0^1 J_C(t) dt$ the demeaned version of J_C.

THEOREM 4.1. *Consider the model (1)–(6) with instruments $\tilde{z}_t$ defined by (30) and (31). The following limit theory as $n \to \infty$ applies for the estimator $\tilde{A}_{IVX}$ in (32) under Assumptions P and M(i).*

(i) *Under Assumptions P(i)–(ii),*

$$\sqrt{n(\kappa_n \wedge \kappa_{nz})} vec\left(\tilde{A}_n - A\right) \Rightarrow MN\left(0, \left(\Psi_C^{-1}\right)' V_{\bar{z}\bar{z}} \Psi_C^{-1} \otimes \Sigma_{\varepsilon\varepsilon}\right),$$

where Ψ_C and $V_{\bar{z}\bar{z}}$ take the following form: under Assumption P(i),

$$\Psi_C = -\left[\Omega_{xx} + \int_0^1 \underline{J}_C(s) dB_x'(s) + \int_0^1 \underline{J}_C(s) \underline{J}_C'(s) ds C'\right] \left(C_z^{-1}\right)'$$

and $V_{\bar{z}\bar{z}} = V_{zz}$; under Assumption P(ii): $\Psi_C = CV_{xx}\left(C_z^{-1}\right)'$ and $V_{\bar{z}\bar{z}} = V_{zz}$ when $\kappa_{nz}/\kappa_n \to 0$; $\Psi_C = V_{\bar{z}\bar{z}} = V_{xx}$ when $\kappa_n/\kappa_{nz} \to 0$; $\Psi_C = -CV_{xz}$ and

$$V_{\bar{z}\bar{z}} = \int_0^\infty e^{s\kappa_z C_z} \left(CV_{xz}C_z' + C_z V_{xz}'C'\right) e^{sC_z'} ds$$

when $\kappa_n/\kappa_{nz} \to \kappa_z > 0$, where the matrices V_{xx}, V_{zz}, and V_{xz} are defined in (7) and (35),

(ii) *Under Assumption P(iii), $\sqrt{n} vec\left(\tilde{A}_{IVX} - A\right) \Rightarrow N(0, V_0)$, where*

$$V_0 = \left(\left[\mathbb{E}\check{x}_1 \check{x}_1'\right]^{-1} \otimes I_m\right) \mathbb{E}\left(\check{x}_1 \check{x}_1' \otimes \varepsilon_2 \varepsilon_2'\right) \left(\left[\mathbb{E}\check{x}_1 \check{x}_1'\right]^{-1} \otimes I_m\right) \tag{37}$$

and $\check{x}_t$ is defined in (36).

COROLLARY 4.2. *Under Assumption M(i), the IVX–Wald statistic*

$$\tilde{W}_n = \left(Hvec\tilde{A}_{IVX} - h\right)' \left\{H\left[\left(\underline{X}'P_{\tilde{Z}}\underline{X}\right)^{-1} \otimes \hat{\Sigma}_{\varepsilon\varepsilon}\right]H'\right\}^{-1} \left(Hvec\tilde{A}_{IVX} - h\right) \qquad \textbf{(38)}$$

for testing the hypothesis (27) has a $\chi^2(q)$ asymptotic distribution when the regressor in (1)–(6) satisfies Assumptions P(i)–(ii).

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 nonstandard limit distribution under conditionally heteroskedastic innovations. When x_t is a stationary process and the innovation sequence ε_t in (2) is conditionally heteroskedastic, the asymptotic variance of $n^{-1/2}\sum_{t=1}^{n}(x_{t-1} \otimes \varepsilon_t)$ is given by $\Upsilon = \mathbb{E}\left(x_{t-1}x_{t-1}' \otimes \varepsilon_t\varepsilon_t'\right)$ and does not factorize to $\mathbb{E}\left(x_{t-1}x_{t-1}'\right) \otimes \Sigma_{\varepsilon\varepsilon}$ as in the case when ε_t are conditionally homoskedastic; consequently, the matrix $n\left(\underline{X}'\underline{X}\right)^{-1} \otimes \hat{\Sigma}_{\varepsilon\varepsilon}$ 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

$$\hat{\Upsilon}_n = \sum_{t=1}^{n} \left(\tilde{z}_{t-1}\tilde{z}_{t-1}' \otimes \hat{\varepsilon}_t\hat{\varepsilon}_t'\right), \qquad \textbf{(39)}$$

where $\tilde{z}_t$ are the IVX instruments in (30) and $\hat{\varepsilon}_t$ are the OLS residuals from (2). The corrected IVX–Wald statistic takes the form

$$\hat{W}_n = \left(Hvec\tilde{A}_{IVX} - h\right)' \left(H\hat{Q}_nH'\right)^{-1} \left(Hvec\tilde{A}_{IVX} - h\right), \qquad \textbf{(40)}$$

$$\hat{Q}_n = \left[\left(\tilde{Z}'\underline{X}\right)^{-1} \otimes I_m\right]\hat{\Upsilon}_n\left[\left(\underline{X}'\tilde{Z}\right)^{-1} \otimes I_m\right]. \qquad \textbf{(41)}$$

The next results characterize the asymptotic behavior of $\hat{\Upsilon}_n$ in (39) and confirm the validity of the White-type correction for all persistence regimes of Assumption P.

LEMMA 4.3. *Denote by $x_{0t} = \sum_{j=1}^{t} R_n^{t-j}u_j$ the regressor in (1) with $x_0 = 0$ and by $\tilde{z}_{0t}$ the IVX instrument in (30) generated by x_{0t}. Under Assumptions P and M(i), the following hold as $n \to \infty$:*

(i) $\max_{1 \le t \le n} \mathbb{E}\left\|(\kappa_n \wedge \kappa_{nz})^{-1/2}\tilde{z}_{0t}\right\|^4 = O(1)$.

(ii) $n^{-1}(\kappa_n \wedge \kappa_{nz})^{-1}\hat{\Upsilon}_n \to_p \Phi$, *where: $\Phi = V_{zz} \otimes \Sigma_{\varepsilon\varepsilon}$ when $\kappa_{nz}/\kappa_n \to 0$; $\Phi = V_{xx} \otimes \Sigma_{\varepsilon\varepsilon}$ when $\kappa_n/\kappa_{nz} \to 0$ and $\kappa_n \to \infty$; $\Phi = \mathbb{E}\left(\check{x}_1\check{x}_1' \otimes \varepsilon_2\varepsilon_2'\right)$ under Assumption P(iii) with $\check{x}_t$, V_{xx} and V_{zz} defined in (36), (7) and (35).*

THEOREM 4.4. *Under Assumptions M(i) and P, the corrected IVX–Wald statistic $\hat{W}_n$ in (40) has a $\chi^2(q)$ asymptotic distribution under (27).*

Remark 4.5.

(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 in that $\kappa_n \to \infty$. GARCH effects are present in the limit distribution only in the case P(iii) where the regressor x_t is a stable autoregressive process (and $\kappa_n = 1$). These results are a direct consequence of the asymptotic development in Section 3 and the fact that any degree of persistence $\kappa_n \to \infty$ 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 $\kappa_{nz} \ll \kappa_n$ and x_t if $\infty \uparrow \kappa_n \ll \kappa_{nz}$), the martingale transform $\tilde{N}_n$ in (33) will behave asymptotically as its near stationary counterpart (16), 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 $\kappa_n = 1$ 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 of data generating mechanisms considered in classes P(i)–(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 in (39) by using the IVX instruments instead of the regressors, in order to ensure the robustness of the corrected IVX–Wald statistic in (40) to regressors with degree of persistence κ_n dominating the instrument rate κ_{nz}. The robustness of the corrected IVX–Wald test statistic $\hat{W}_n$ in (40) to conditional heteroskedasticity under all persistence regimes makes it practically relevant and suitable for general application.

(iii) KMS (2015a) have proposed a finite sample correction to the IVX–Wald test statistic $\tilde{W}_n$ in (38) 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), leading to the adjusted version having a $\chi^2(q)$ limit distribution for models covering the entire range of Assumption P. Extensive Monte Carlo experiments in KMS (2015a) suggest that choosing $R_{nz} = \rho_{nz}I_r$ with $\rho_{nz} = 1 - 1/n^{0.95}$ in (31) gives rise to an IVX-Wald test statistic with very good finite sample size and power properties.

(iv) A version of Theorems 4.1 and 4.4 were stated under more restrictive assumptions in KMS (2015a) but no theory was provided. The current paper establishes the theory of IVX estimation and inference in predictive regression systems with conditionally heteroskedastic innovations under the more general vector autoregressive framework of Assumption P (see footnote 1 above) and less restrictive assumptions on the innovations compared to those imposed by KMS (2015a): while the GARCH parametric specification for ϵ_t is common to both papers, KMS (2015a) imposes an additional weak dependence condition (equation (8) of that paper) and the stricter condition $\sum_{j=1}^{\infty} j \left\| F_j \right\| < \infty$ of Phillips and Solo (1992) on the linear process for u_t compared to the summability requirement (5) of the current paper.

5. DISCUSSION

The paper provides a complete characterization 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-normalized 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 $O_p(n)$ 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 the same limit distribution that applies under conditionally homoskedastic innovations, given in MP (2009) and Phillips (1988) respectively. A similar result has been established by Andrews and Guggenberger (2012) for a scalar AR(1) process using weak dependence techniques rather than a GARCH parametrization.

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 (2009a) 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 to 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 $\|M\| = \max\left\{\sqrt{\lambda} : \lambda \in \sigma\left(M'M\right)\right\}$ and $\|M\|_F = \left(\text{tr}M'M\right)^{1/2}$ the spectral and Frobenius matrix norms and by $\sigma\left(A\right)$ and $\rho\left(A\right)$ the spectrum and the spectral radius of a square matrix A.

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

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

with $s = r^2 m\left(m+1\right)/2$, 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

$$\det\left(M_\kappa - \lambda I_{\kappa s}\right) = \left(-\lambda\right)^{\kappa s} \det G_\kappa\left(\lambda\right). \tag{42}$$

The identity (42) implies that any nonzero eigenvalue λ of M_κ satisfies

$$\det G_\kappa\left(\lambda\right) = 0. \tag{43}$$

Denoting by M^H the conjugate transpose of a square complex matrix M, the Hermitian part of $G_\kappa\left(\lambda\right)$ is given by

$$\begin{aligned}
\mathcal{H}[G_\kappa\left(\lambda\right)] &= \frac{1}{2}\left[G_\kappa\left(\lambda\right) + G_\kappa^H\left(\lambda\right)\right] \\
&= \frac{1}{2}\left[2I_s - \left(\frac{1}{\lambda} + \frac{1}{\bar{\lambda}}\right)\Gamma_1 - \cdots - \left(\frac{1}{\lambda^\kappa} + \frac{1}{\bar{\lambda}^\kappa}\right)\Gamma_\kappa\right] \\
&= \frac{1}{2}\left[2I_s - \frac{2\,\text{Re}\left(\lambda\right)}{|\lambda|^2}\Gamma_1 - \frac{2\,\text{Re}\left(\lambda^2\right)}{|\lambda|^4}\Gamma_2 - \cdots - \frac{2\,\text{Re}\left(\lambda^\kappa\right)}{|\lambda|^{2\kappa}}\Gamma_\kappa\right] \\
&= I_m - \sum_{i=1}^\kappa \Gamma_i + \sum_{j=1}^\kappa \left[1 - \frac{\text{Re}\left(\lambda^j\right)}{|\lambda|^{2j}}\right]\Gamma_j.
\end{aligned} \tag{44}$$

The conditions $\Gamma_i \geq 0$ and $\rho\left(\Gamma\right) < 1$ of Assumption G imply that the matrix $I_m - \sum_{i=1}^\kappa \Gamma_i = I_m - \Gamma$ is positive definite: $\lambda_{\min}\left(I_m - \Gamma\right) = 1 - \lambda_{\max}\left(\Gamma\right) > 0$. Moreover, for arbitrary $\lambda \in \mathbb{C}$ with $\lambda \neq 0$ and $j \in \mathbb{N}$,

$$|\lambda| \geq 1 \implies |\lambda|^{2j} \geq |\lambda|^j = |\lambda^j| \geq \left|\text{Re}\left(\lambda^j\right)\right| \implies 1 - \frac{\text{Re}\left(\lambda^j\right)}{|\lambda|^{2j}} \geq 0,$$

which implies that the second sum on the right of (44) is a positive semidefinite matrix. Since $I_m - \sum_{i=1}^\kappa \Gamma_i > 0$, (44) implies that $\mathcal{H}[G_\kappa\left(\lambda\right)]$ is a positive definite matrix for all $\lambda \in \mathbb{C}$ satisfying $|\lambda| \geq 1$. Positivity of $\mathcal{H}[G_\kappa\left(\lambda\right)]$ implies the

nonsingularity of $G_\kappa(\lambda)$, so (43) is violated when $|\lambda| \geq 1$. We conclude that M_κ cannot have an eigenvalue with $|\lambda| \geq 1$, so $\rho(M_\kappa) < 1$.

To show the second assertion, Householder's theorem (Lemma 5.6.10 in Horn and Johnson, 2013) ensures that for any $\delta > 0$ there exists $n_0(\delta) \in \mathbb{N}$ and a matrix norm $\|\cdot\|_{n_0(\delta)}$ such that $\|M_\kappa\|_{n_0(\delta)} \leq \rho(M_\kappa) + \delta$. By equivalence of norms in finite dimensional spaces, for each $\delta > 0$ there exists a constant $c(\delta) \in (0, \infty)$ such that

$$\left\| M_\kappa^j \right\| \leq c(\delta) \left\| M_\kappa^j \right\|_{n_0(\delta)} \leq c(\delta) \|M_\kappa\|_{n_0(\delta)}^j \leq c(\delta) [\rho(M_\kappa) + \delta]^j$$

for each $j \in \mathbb{N}$. Since $\rho(M_\kappa) < 1$, we may choose $\delta \in (0, 1 - \rho(M_\kappa))$, which implies that $\lambda_\delta := \rho(M_\kappa) + \delta \in (0, 1)$ and

$$\sum_{j=0}^{\infty} \left\| M_\kappa^j \right\| \leq c(\delta) \sum_{j=0}^{\infty} \lambda_\delta^j = \frac{c(\delta)}{1 - \lambda_\delta} < \infty.$$

This shows part (i). For part (ii), $M_{n,\kappa} \to M_\kappa$ as $n \to \infty$ so, by continuity of the eigenvalues of a matrix as a function of the matrix elements, $\rho(M_{n,\kappa}) \to \rho(M_\kappa)$ and $\|M_{n,\kappa}\| \to \|M_\kappa\|$ as $n \to \infty$. Since $\rho(M_\kappa) < 1$, convergence of $\rho(M_{n,\kappa})$ implies that $\lim_{n\to\infty} \rho(M_{n,\kappa}) < 1$. Also, since $\sum_{j=0}^{\infty} \left\| M_\kappa^j \right\| < \infty$, $\|M_{n,\kappa}\| \to \|M_\kappa\|$ and dominated convergence yield $m_n := \sum_{j=0}^{\infty} \left\| M_{n,\kappa}^j \right\| \to \sum_{j=0}^{\infty} \left\| M_\kappa^j \right\| < \infty$. Convergence of (m_n) implies that $\sup_{n\geq 1} m_n < \infty$, completing the proof of the lemma. ∎

Proof of Lemma 3.2. We employ a corollary of the Rosenthal and Minkowski inequalities for a martingale difference sequence (Y_j) satisfying $\mathbb{E}|Y_j|^p < \infty$:

$$\mathbb{E}\left| \sum_{j=1}^{n} Y_j \right|^p \leq C_p \left(\sum_{j=1}^{n} \left(\mathbb{E}|Y_j|^p \right)^{2/p} \right)^{p/2} \quad \text{for } p > 2, \tag{45}$$

where $C_p > 0$ is a constant depending on p only, see Lemma 2.5.2 in Giraitis et al. (2012). Letting $I, J_{t,i} \subseteq \mathbb{N} \cup \{0\}$, the Minkowski inequality followed by the inequality (45) with $p = 4$ imply that

$$\left\| \sum_{i \in I} \sum_{j \in J_{t,i}} R_n^j F_i e_{t-j-i} \right\|_{L_4}^4$$

$$\leq \left(\sum_{i \in I} \left\| \sum_{j \in J_{t,i}} R_n^j F_i e_{t-j-i} \right\|_{L_4} \right)^4 = \left(\sum_{i \in I} \left\{ \mathbb{E} \left\| \sum_{j \in J_{t,i}} R_n^j F_i e_{t-j-i} \right\|^4 \right\}^{1/4} \right)^4$$

$$\leq c \left(\sum_{i \in I} \left\{ \sum_{j \in J_{t,i}} \left(\mathbb{E} \left\| R_n^j F_i e_{t-j-i} \right\|^4 \right)^{1/2} \right\}^{1/2} \right)^4$$

for a uniform constant $c > 0$, giving

$$\left\| \sum_{i \in I} \sum_{j \in J_{t,i}} R_n^j F_i e_{t-j-i} \right\|_{L_4}^4 \leq c \mathbb{E} \|e_1\|^4 \left(\sum_{i \in I} \|F_i\| \left\{ \sum_{j \in J_{t,i}} \left\| R_n^j \right\|^2 \right\}^{1/2} \right)^4. \tag{46}$$

For part (i), (46) yields for each $t \in \{1, ..., n\}$

$$\mathbb{E} \left\| \frac{x_{0t}}{\kappa_n^{1/2}} \right\|^4 = \frac{1}{\kappa_n^2} \left\| \sum_{j=0}^{t-1} R_n^j u_{t-j} \right\|_{L_4}^4 = \frac{1}{\kappa_n^2} \left\| \sum_{i=0}^{\infty} \sum_{j=0}^{t-1} R_n^j F_i e_{t-j-i} \right\|_{L_4}^4$$

$$\leq \frac{c \mathbb{E} \|e_1\|^4}{\kappa_n^2} \left(\sum_{j=0}^{n-1} \left\| R_n^j \right\|^2 \right)^2 \left(\sum_{i=0}^{\infty} \|F_i\| \right)^4.$$

Since the last bound is independent of t and $\sum_{j=0}^{n-1} \left\| R_n^j \right\|^2 = O(\kappa_n)$, part (i) for x_{0t} follows. A simpler version of the above argument applies to $\mathbb{E} \left\| \kappa_n^{-1/2} \zeta_{n,t-1} \right\|^4$ since (45) with $p = 4$ applies directly to the sum defined by $\zeta_{n,t-1}$.

For part (ii), we start by noting that Proposition A2 of Magdalinos and Phillips (2020) continues to hold in the current conditional heteroskedastic setup with the following modification to the proof: when a second moment bound is employed, we replace the role of conditional homoskedasticity in the computation of products of second moments by an L_2 Cauchy–Schwarz inequality followed by the inequality (46) when necessary; the resulting bounds depend on fourth moments (instead of second moments) of (e_t) but the orders of magnitude of the resulting L_2-bounds remain the same. Specifically, referring to the proof of Proposition A2 of MP (2020): the bounds for $\max_{0 \leq m \leq n} \left\| \tilde{\Sigma}_{1n}^{(m)} - \Sigma_{1n}^{(m)} \right\|_{L_1}$ in (39) and for $\max_{0 \leq m \leq n} \left\| \Sigma_{3n}^{(m)} \right\|_{L_1}$ are the same with $\|e_1\|_{L_2} \max_{1 \leq t \leq n} \|x_{0t}\|_{L_2}$ replaced by $\|e_1\|_{L_4} \max_{1 \leq t \leq n} \|x_{0t}\|_{L_4}$; the bounds for $\max_{0 \leq m \leq n} \left\| S_{2n}^{(m)} \right\|_{L_1}$ and $\max_{0 \leq m \leq n} \left\| S_{3n}^{(m)} \right\|_{L_1}$ in equations (42) and (43) are the same with $\|e_1\|_{L_2}^2$ replaced by $c^{1/4} \|e_1\|_{L_4}^2$ where c is the constant in (46); the bound for $\max_{0 \leq m \leq n} \left\| \Sigma_{2n}^{(m)} - S_{1n}^{(m)} \right\|_{L_1}$ is the same. Having established the validity of Proposition A2 of MP (2020) under the assumptions of the current paper, $n^{-1} \sum_{t=1}^{n} x_{t-1} u_{xt}' \to_p \Lambda_{xx}'$ and $n^{-1} \kappa_n^{-1} \sum_{t=1}^{n} x_{t-1} x_{t-1}' \to_p V_{xx}$ from the argument of Lemma 2.2 of MP (2020) with $n^{-1} \sum_{t=1}^{n} e_t e_t' \to_{L_1} \Sigma$ following by the mean ergodic theorem instead of the law of large numbers employed in MP (2020). For the last claim of part (ii),

note that

$$\max_{1 \le m \le n} \left\| n^{-1} \sum_{t=1}^{m} \zeta_{n,t-1} e_t' \right\|_{L_2} \le \|e_1\|_{L_4} \max_{t \le n} \left\| \frac{\zeta_{n,t}}{\sqrt{n}} \right\|_{L_4} = O\left(\sqrt{\frac{\kappa_n}{n}}\right) \tag{47}$$

by part (i) of the lemma. Since $\zeta_{n,t}$ satisfies the recursion $\zeta_{n,t} = R_n \zeta_{n,t-1} + F(1) e_t$, we know that

$$n^{-1} \kappa_n^{-1} \sum_{t=1}^{n} \zeta_{n,t-1} \zeta_{n,t-1}' \to_p \int_0^{\infty} e^{rC} G e^{rC'} dr,$$

where, since $n^{-1} \sum_{t=1}^{n} \zeta_{n,t-1} e_t' \to_p 0$ by (47), $G = \mathrm{Re}_{n \to \infty} n^{-1} \sum_{t=1}^{n} F(1) e_t e_t' F(1)'$ $= \Omega_{uu}$. Since $G = \Omega_{uu}$, $\int_0^{\infty} e^{rC} G e^{rC'} dr = V_{xx}$ and the proof of part (ii) is complete.

For part (iii), $x_{t-1} = x_{0t-1} + R_n^{t-1} x_0$, so since $\|x_0\| = o_p\left(\kappa_n^{1/2}\right)$, we may write $\left\| \sum_{t=1}^{n} R_n^{t-1} x_0 \otimes \varepsilon_t \right\| \le \|x_0\| \left\| \sum_{t=1}^{n} R_n^{t-1} \otimes \varepsilon_t \right\| = o_p(\kappa_n)$ and

$$\frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^{n} x_{t-1} \otimes \varepsilon_t = \frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^{n} x_{0t-1} \otimes \varepsilon_t + o_p\left(\sqrt{\frac{\kappa_n}{n}}\right). \tag{48}$$

Writing out x_{0t-1} and letting $(m_n)_{n \in \mathbb{N}} \subset \mathbb{N}$ be a sequence satisfying $m_n \to \infty, m_n/\kappa_n \to 0$, we obtain

$$
\begin{aligned}
x_{0t-1} &= \sum_{j=0}^{t-2} R_n^j u_{t-j-1} = \sum_{j=0}^{t-2} R_n^j \sum_{i=0}^{\infty} F_i e_{t-j-i-1} \\
&= \sum_{i=0}^{m_n} \sum_{j=i}^{t+i-2} R_n^{j-i} F_i e_{t-j-1} + \sum_{i>m_n} \sum_{j=0}^{t-2} R_n^j F_i e_{t-j-i-1} \\
&= \sum_{j=0}^{t-2} R_n^j \left(\sum_{i=0}^{\infty} F_i\right) e_{t-j-1} - \sum_{j=0}^{t-2} R_n^j \left(\sum_{i>m_n} F_i\right) e_{t-j-1} + \sum_{i=0}^{m_n} \sum_{j=t-1}^{t+i-2} R_n^j F_i e_{t-j-1} \\
&\quad - \sum_{i=0}^{m_n} \sum_{j=0}^{i-1} R_n^j F_i e_{t-j-1} + \sum_{i=0}^{m_n} \sum_{j=i}^{t+i-2} R_n^j \left(R_n^{-i} - I_r\right) F_i e_{t-j-1} + \sum_{i>m_n} \sum_{j=0}^{t-2} R_n^j F_i e_{t-j-i-1} \\
&= \zeta_{n,t-1} - \eta_{n,t-1}^{(1)} + \eta_{n,t-1}^{(2)} - \eta_{n,t-1}^{(3)} + \eta_{n,t-1}^{(4)} + \eta_{n,t-1}^{(5)}
\end{aligned} \tag{49}
$$

in order of appearance. Noting that $\eta_{n,t-1}^{(k)} \otimes \varepsilon_t$ is an $\mathcal{F}_t$-martingale difference sequence, we will show that

$$\mathbb{E} \left\| \frac{1}{\sqrt{n\kappa_n}} \sum_{t=1}^{n} \eta_{n,t-1}^{(k)} \otimes \varepsilon_t \right\|^2 \to 0, \quad k \in \{1, ..., 5\} \tag{50}$$

and the result will follow from (48) and (49). For $\eta_{n,t}^{(5)}$ and $\eta_{n,t}^{(1)}$, the Cauchy–Schwarz inequality followed by (46) give

$$
\mathbb{E}\left\|\frac{1}{\sqrt{n\kappa_n}}\sum_{t=1}^{n}\eta_{n,t}^{(5)}\otimes\varepsilon_t\right\|^2 = \frac{1}{n\kappa_n}\sum_{t=1}^{n}\mathbb{E}\left(\left\|\eta_{n,t}^{(5)}\right\|^2\|\varepsilon_t\|^2\right)
$$

$$
\leq \frac{1}{n\kappa_n}\left(\mathbb{E}\|\varepsilon_1\|^4\right)^{1/2}\sum_{t=1}^{n}\left(\mathbb{E}\left\|\eta_{n,t}^{(5)}\right\|^4\right)^{1/2}
$$

$$
\leq \left(c\mathbb{E}\|\varepsilon_1\|^4\,\mathbb{E}\|e_1\|^4\right)^{1/2}\left(\sum_{i>m_n}\|F_i\|\right)^2\frac{1}{\kappa_n}\sum_{j=0}^{n-2}\left\|R_n^j\right\|^2
$$

$$
= O(1)\left(\sum_{i>m_n}\|F_i\|\right)^2 \to 0
$$

since $m_n \to \infty$. The same bound as above applies to the left side of (50) with $k = 1$. For $\eta_{n,t-1}^{(2)}$ and $\eta_{n,t-1}^{(3)}$, a similar argument based on (46) and using the fact that $\left\|R_n^j\right\| < 1$ gives

$$
\mathbb{E}\left\|\frac{1}{\sqrt{n\kappa_n}}\sum_{t=1}^{n}\eta_{n,t}^{(2)}\otimes\varepsilon_t\right\|^2
$$

$$
\leq \frac{1}{n\kappa_n}\left(\mathbb{E}\|\varepsilon_1\|^4\right)^{1/2}\sum_{t=1}^{n}\left(\mathbb{E}\left\|\eta_{n,t}^{(2)}\right\|^4\right)^{1/2}
$$

$$
\leq \frac{\left(c\mathbb{E}\|e_1\|^4\,\mathbb{E}\|\varepsilon_1\|^4\right)^{1/2}}{n\kappa_n}\sum_{t=1}^{n}\left(\sum_{i=0}^{m_n}\|F_i\|\left\{\sum_{j=t-1}^{t+i-2}\left\|R_n^j\right\|^2\right\}^{1/2}\right)^2
$$

$$
\leq \frac{\left(c\mathbb{E}\|e_1\|^4\,\mathbb{E}\|\varepsilon_1\|^4\right)^{1/2}}{\kappa_n}\left(\sup_{l\geq 0}\|F_l\|\right)^2\left(\sum_{i=0}^{m_n}i^{1/2}\right)^2
$$

$$
\leq O\left(\frac{m_n}{\kappa_n}\right) \to 0
$$

from the choice of $(m_n)_{n\in\mathbb{N}}$. The same bound as above applies to the (50) with $k = 3$. Finally, applying (46) with F_i replaced by $\left(R_n^{-i} - I_r\right)F_i$ gives

$$
\mathbb{E}\left\|\frac{1}{\sqrt{n\kappa_n}}\sum_{t=1}^{n}\eta_{n,t}^{(4)}\otimes\varepsilon_t\right\|^2
$$

$$
\leq \frac{1}{n\kappa_n}\left(\mathbb{E}\|\varepsilon_1\|^4\right)^{1/2}\sum_{t=1}^{n}\left(\mathbb{E}\left\|\eta_{n,t}^{(4)}\right\|^4\right)^{1/2}
$$

$$\leq \left(c\mathbb{E}\left\|e_1\right\|^4\mathbb{E}\left\|\varepsilon_1\right\|^4\right)^{1/2}\frac{1}{\kappa_n}\sum_{j=0}^{n-2}\left\|R_n^j\right\|^2\left(\sum_{i=0}^{m_n}\left\|R_n^{-i}-I_r\right\|\left\|F_i\right\|\right)^2$$

$$\leq O\left(1\right)\left(\max_{0\leq i\leq m_n}\left\|R_n^{-i}-I_r\right\|\right)^2\left(\sum_{i=0}^{m_n}\left\|F_i\right\|\right)^2$$

$$\leq O\left(1\right)\left\|R_n\right\|^{-2m_n}\left(\max_{0\leq i\leq m_n}\left\|I_r-R_n^i\right\|\right)^2=O\left[\left(\frac{m_n}{\kappa_n}\right)^2\right]=o\left(1\right)$$

from the choice of $(m_n)_{n\in\mathbb{N}}$, where the last order of magnitude is obtained since $\|R_n\|^{-2m_n}\to 1$ when $m_n/\kappa_n\to 0$ and $\max_{0\leq i\leq m_n}\left\|I_r-R_n^i\right\|=O(m_n/\kappa_n)$ by Lemma 2.1(ii) of MP (2020). This shows (50) and completes the proof. ∎

Proof of Lemma 3.3. By definition of the process ζ_{nt-1}, we can write

$$\langle M\rangle_n-V_n=\frac{1}{n^{1+\alpha}}\sum_{t=1}^n\left[\zeta_{nt-1}\zeta_{nt-1}'\otimes(H_t-\Sigma_{\varepsilon\varepsilon})\right]$$

$$=A_n+B_n+B_n',\tag{51}$$

where

$$A_n=\frac{1}{n\kappa_n}\sum_{t=1}^n\left[\left\{\sum_{j=1}^{t-1}R_n^{j-1}F\left(1\right)e_{t-j}e_{t-j}'F\left(1\right)'R_n^{j-1}\right\}\otimes(H_t-\Sigma_{\varepsilon\varepsilon})\right]\tag{52}$$

$$B_n=\frac{1}{n\kappa_n}\sum_{t=1}^n\left[\left\{\sum_{j=1}^{t-2}\sum_{i=j+1}^{t-1}R_n^{j-1}F\left(1\right)e_{t-j}e_{t-i}'F\left(1\right)'R_n^{i-1}\right\}\otimes(H_t-\Sigma_{\varepsilon\varepsilon})\right].\tag{53}$$

We first expand the term in (52) by adding and subtracting Σ_{ee} from as follows:

$$A_n=\frac{1}{n\kappa_n}\sum_{j=1}^{n-1}\left(R_n^{j-1}F\left(1\right)\otimes I\right)\sum_{t=1}^{n-j}\left[e_te_t'\otimes\left(H_{t+j}-\Sigma_{\varepsilon\varepsilon}\right)\right]\left(F\left(1\right)'R_n^{j-1}\otimes I\right)$$

$$=\frac{1}{n\kappa_n}\sum_{j=1}^{n-1}\left(R_n^{j-1}F\left(1\right)\otimes I\right)\sum_{t=1}^{n-j}\left[\left(e_te_t'-\Sigma_{ee}\right)\otimes H_{t+j}\right]\left(F\left(1\right)'R_n^{j-1}\otimes I\right)$$

$$-\frac{1}{n\kappa_n}\sum_{j=1}^{n-1}\left(R_n^{n-j-1}F\left(1\right)\otimes I\right)\left[\left\{\sum_{t=1}^j\left(e_te_t'-\Sigma_{ee}\right)\right\}\otimes\Sigma_{\varepsilon\varepsilon}\right]F\left(1\right)'\left(R_n^{n-j-1}\otimes I\right)$$

$$+\frac{1}{n\kappa_n}\sum_{j=1}^{n-1}R_n^{j-1}\Omega_uR_n^{j-1}\otimes\sum_{t=j+1}^n(H_t-\Sigma_{\varepsilon\varepsilon})$$

$$=A_{1n}-A_{2n}+A_{3n}$$

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

$$\frac{1}{n}\sum_{t=1}^{n}\left(e_t e_t' - \Sigma_{ee}\right) \to_{a.s.} 0 \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n}\left(H_t - \Sigma_{\varepsilon\varepsilon}\right) \to_{a.s.} 0 \tag{54}$$

by the ergodic theorem on (H_t) and (e_t), we can write

$$\|A_{3n}\| \le \|\Omega_u\| \max_{1\le j\le n}\left\|\frac{1}{n}\sum_{t=j+1}^{n}\left(H_t - \Sigma_{\varepsilon\varepsilon}\right)\right\| \frac{1}{\kappa_n}\sum_{j=0}^{n-2}\left\|R_n^j\right\|^2 \to_{a.s.} 0$$

by (54) since $\sum_{j=0}^{n-1}\left\|R_n^j\right\|^2 = O(\kappa_n)$. The same argument works for A_{2n}:

$$\|A_{2n}\| \le \|\Sigma_{\varepsilon\varepsilon}\| \, \|F(1)\|^2 \max_{1\le j\le n}\left\|\frac{1}{n}\sum_{t=1}^{j}\left(e_t e_t' - \Sigma_{ee}\right)\right\| \frac{1}{\kappa_n}\sum_{j=0}^{n-2}\left\|R_n^j\right\|^2 \to_{a.s.} 0.$$

We conclude that $\|A_n - A_{1n}\| = o_{a.s.}(1)$ and

$$\text{vec}(A_{1n}) = \frac{1}{n\kappa_n}\sum_{j=1}^{n-1}\left(R_n^{j-1}\otimes R_n^{j-1}\right)\left(F(1)\otimes F(1)\right)\sum_{t=1}^{n-j}\text{vec}\left[\left(e_t e_t' - \Sigma_{ee}\right)\otimes H_{t+j}\right].$$

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 $\{K_{ij}L_{kl} : 1\le i,j\le r, 1\le k,l\le m\}$ but appear in different order in the two vectors. Therefore, there exists an $m^2 r^2 \times m^2 r^2$ permutation matrix Π such that $\text{vec}(K\otimes L) = \Pi\left[\text{vec}(K)\otimes\text{vec}(L)\right]$. 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

$$\text{vec}(A_{1n}) = \Pi\left(I_{r^2}\otimes D_m\right)\frac{1}{n\kappa_n}\sum_{j=1}^{n-1}\left(R_n^{j-1}\otimes R_n^{j-1}\right)\left(F(1)\otimes F(1)\right)\sum_{t=1}^{n-j}\Sigma_t(j), \tag{55}$$

with $\Sigma_t(j) = \text{vec}\left(e_t e_t' - \Sigma_{ee}\right)\otimes\text{vech}\left(H_{t+j}\right)$ as in Lemma 3.5. Since $\|\Pi\| = 1$ and $\|D_m\| = \sqrt{2}$ for a permutation matrix Π, (55) yields the asymptotic bound for (52):

$$\|\text{vec}(A_n)\| \le \sqrt{2}\,\|F(1)\|^2\frac{1}{n\kappa_n}\sum_{j=1}^{n-1}\left\|R_n^{j-1}\right\|^2\left\|\sum_{t=1}^{n-j}\Sigma_t(j)\right\| + o_{a.s.}(1). \tag{56}$$

The term in (53) can be written as:

$$B_n = \frac{1}{n\kappa_n}\sum_{t=1}^{n}\sum_{j=1}^{t-2}\sum_{i=j+1}^{t-1}\left[\left\{R_n^{j-1}F(1)\,e_{t-j}e_{t-i}'F(1)'R_n^{i-1}\right\}\otimes\left(H_t - \Sigma_{\varepsilon\varepsilon}\right)\right]$$

$$= \frac{1}{n\kappa_n}\sum_{j=1}^{n-2}\sum_{t=2}^{n-j}\sum_{i=1}^{t-1}\left[\left\{R_n^{j-1}F(1)\,e_t e_{t-i}'F(1)'R_n^{i+j-1}\right\}\otimes\left(H_{t+j} - \Sigma_{\varepsilon\varepsilon}\right)\right]$$

$$= \frac{1}{n\kappa_n} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} \left[\left\{ R_n^{j-1} F(1) e_t \left(\sum_{i=1}^{t-1} R_n^{i-1} F(1) e_{t-i} \right)' R_n^j \right\} \otimes \left(H_{t+j} - \Sigma_{\varepsilon\varepsilon} \right) \right]$$

$$= \frac{1}{n\kappa_n} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} \left[\left\{ R_n^{j-1} F(1) e_t \zeta_{n,t-1}' R_n^j \right\} \otimes \left(H_{t+j} - \Sigma_{\varepsilon\varepsilon} \right) \right]$$

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

$$\text{vec}\,(B_n)$$

$$= \Pi \frac{1}{n\kappa_n} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} \left[\text{vec}\left\{ R_n^{j-1} F(1) e_t \zeta_{n,t-1}' R_n^j \right\} \otimes \text{vec}\left(H_{t+j} - \Sigma_{\varepsilon\varepsilon} \right) \right]$$

$$= \Pi \left(I_r \otimes R_n^{-1} \otimes D_m \right) \frac{1}{n\kappa_n} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} \left[R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \otimes \text{vech}\left(H_{t+j} \right) \right] + o_p(1)$$

$$= \Pi \left(I_r \otimes R_n^{-1} \otimes D_m \right) \frac{1}{n\kappa_n} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} S_{n,t}(j) + o_p(1),$$

where $S_{n,t}(j)$ is defined in (20) and the term involving $\Sigma_{\varepsilon\varepsilon}$ is $o_p(1)$ because

$$\left\| \frac{1}{n\kappa_n} \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \right\|_{L_1}$$

$$\leq \frac{\|F(1)\|}{\kappa_n} \sum_{j=1}^{n-2} \|R_n^j\|^2 \left\| \frac{1}{n} \sum_{t=2}^{n-j} \zeta_{n,t-1} \otimes e_t \right\|_{L_2} = O\left(\sqrt{\frac{\kappa_n}{n}} \right)$$

by (47). The last expression for $\text{vec}\,(B_n)$ yields

$$\|\text{vec}\,(B_n)\| \leq b \frac{1}{n\kappa_n} \left\| \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} S_{n,t}(j) \right\| + o_p(1) \tag{57}$$

for some uniform bounding constant $b > 0$. Combining (51), (56), and (57) shows (22) and part (i).

Next, we show that $\Sigma_t(j)$ and $S_{n,t}(j)$ satisfy (12) with innovations given by (24) and (25) respectively. For $S_{n,t}(j)$, applying (3) to $\text{vech}\,(H_{t+j})$ yields

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

$$= R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \otimes \left\{ \sum_{i=1}^{q} A_i \text{vech}\,(H_{t+j-i}) + \sum_{k=1}^{p} B_k \text{vech}\,(H_{t+j-k}) \right\}$$

$$+ R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \otimes \left\{ \sum_{i=1}^q A_i \text{vech} \left(\varepsilon_{t+j-i} \varepsilon'_{t+j-i} - H_{t+j-i} \right) \right\}$$

$$+ R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \otimes \varphi$$

$$= \sum_{i=1}^{\kappa} \left(R_n^i \otimes R_n^i \otimes \Gamma_i \right) \left[R_n^{j-i} \zeta_{n,t-1} \otimes R_n^{j-i} F(1) e_t \otimes \text{vech} \left(H_{t+j-i} \right) \right] + \upsilon_{n,t}(j)$$

$$= \sum_{i=1}^{\kappa} \left(R_n^i \otimes R_n^i \otimes \Gamma_i \right) S_{n,t}(j-i) + \upsilon_{n,t}(j), \tag{58}$$

where

$$\upsilon_{n,t}(j) = R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \otimes \varphi + \sum_{i=1}^q \left(I_{r^2} \otimes A_i \right) \left[R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \otimes w_{t+j-i} \right]$$

and $w_{t+j-i} = \text{vech} \left(\varepsilon_{t+j-i} \varepsilon'_{t+j-i} - H_{t+j-i} \right)$. Since the above expression for $\upsilon_{n,t}(j)$ coincides with (25) and $\Gamma_{n,i} = R_n^i \otimes R_n^i \otimes \Gamma_i$ by (11), (58) shows part (ii) for $S_{n,t}(j)$. Applying (3) to vech $\left(H_{t+j} \right)$ in

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

and proceeding as in (58) shows part (ii) for $\Sigma_t(j)$.

For part (iii), the processes in (19) and (20) have companion form solutions

$$\tilde{\Sigma}_t(j) = M_\kappa^j \tilde{\Sigma}_t(0) + \sum_{l=1}^j M_\kappa^{j-l} \tilde{v}_t(l), \quad j \geq 1 \tag{59}$$

$$\tilde{S}_{n,t}(j) = M_{n,\kappa}^j \tilde{S}_{n,t}(0) + \sum_{l=1}^j M_{n,\kappa}^{j-l} \tilde{\upsilon}_{n,t}(l), \quad j \geq 1 \tag{60}$$

for the "stacked" processes

$$\tilde{\Sigma}_t(j) = \left[\Sigma_t(j)', \Sigma_t(j-1)', ..., \Sigma_t(j-\kappa+1)' \right]' \tag{61}$$

$$\tilde{S}_{n,t}(j) = \left[S_{n,t}(j)', S_{n,t}(j-1)', ..., S_{n,t}(j-\kappa+1)' \right]' \tag{62}$$

and

$$\tilde{v}_t(j) = \left[v_t(j)', 0, ..., 0 \right]', \quad \tilde{\upsilon}_{n,t}(j) = \left[\upsilon_{n,t}(j)', 0, ..., 0 \right]'. \tag{63}$$

By (61), (63) 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 (59) to the first term of the bound (22) of Lemma 3.3(i) to obtain

$$
\sigma_n = \frac{1}{n\kappa_n} \sum_{j=1}^{n-1} \left\| R_n^{j-1} \right\|^2 \left\| \sum_{t=1}^{n-j} \Sigma_t(j) \right\| \leq \frac{1}{n\kappa_n} \sum_{j=1}^{n-1} \left\| R_n^{j-1} \right\|^2 \left\| \sum_{t=1}^{n-j} \tilde{\Sigma}_t(j) \right\|
$$

$$
\leq \frac{1}{\kappa_n} \sum_{j=1}^{\infty} \left\| M_\kappa^j \right\| \frac{1}{n} \sum_{t=1}^{n} \left\| \tilde{\Sigma}_t(0) \right\| + \frac{1}{n\kappa_n} \sum_{j=1}^{n-1} \left\| R_n^{j-1} \right\|^2 \sum_{l=1}^{j} \left\| M_\kappa^{j-l} \right\| \left\| \sum_{t=1}^{n-j} v_t(l) \right\|
$$

$$
= \sigma_{1n} + \sigma_{2n} \tag{64}
$$

in order of appearance. The first term of (64) satisfies $\mathbb{E} \|\sigma_{1n}\| = O\left(\kappa_n^{-1}\right)$ because $\sum_{j=1}^{\infty} \left\| M_\kappa^j \right\| < \infty$ by Lemma 3.1 and

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

$$
\leq \kappa \max_{i<\kappa} \left\{ \mathbb{E} \|e_1\|^4 \mathbb{E} \|H_{t-i}\|^2 \right\}^{1/2} \leq \kappa \left\{ \mathbb{E} \|e_1\|^4 \mathbb{E} \|\varepsilon_1\|^4 \right\}^{1/2}
$$

by the Jensen inequality for conditional expectations. For the second term of (64), letting

$$
\omega_t(k) = \text{vec}\left(e_t e_t' - \Sigma_{ee} \right) \otimes w_{t+k} \tag{65}
$$

and using the expression in (24) we can write

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

$$
+ \|\varphi\| \max_{j \leq n} \left\| \frac{1}{n} \sum_{t=1}^{j} \text{vec}\left(e_t e_t' - \Sigma_{ee} \right) \right\| \sum_{l=1}^{\infty} \left\| M_\kappa^l \right\| \frac{1}{\kappa_n} \sum_{j=1}^{n-1} \left\| R_n^{n-j-1} \right\|^2
$$

$$
= \frac{\|\varphi\|}{n\kappa_n} \sum_{i=1}^{q} \|A_i\| \sum_{l=0}^{n-2} \left\| M_\kappa^l \right\| \sum_{j=l+1}^{n-1} \left\| R_n^{j-1} \right\|^2 \left\| \sum_{t=1}^{n-j} \omega_t(j-l-i) \right\| + o_{a.s.}(1)
$$

$$
\leq \frac{\|\varphi\|}{n\kappa_n} \sum_{i=1}^{q} \|A_i\| \sum_{l=0}^{n-2} \left\| M_\kappa^l \right\| \sum_{j=1}^{n-1} \left\| R_n^{j-1} \right\|^2 \left\| \sum_{t=1}^{n-j} \omega_t(j-i) \right\| + o_{a.s.}(1)
$$

where the term on the second line is $o_{a.s.}(1)$ by the ergodic theorem. The remaining term is $O_p\left(\kappa_n^{-1}\right)$ when $j \leq i$ because, in this case, it is bounded in L_1 norm by

$$
\frac{b}{n\kappa_n} \max_{1 \leq l, i \leq q} \sum_{t=1}^{n} \mathbb{E} \|\omega_t(l-i)\| \leq \frac{2b}{n\kappa_n} \max_{1 \leq l, i \leq q} \sum_{t=1}^{n} \mathbb{E} \|e_t\|^2 \|w_{t+l-i}\|
$$

$$
\leq \frac{4b}{\kappa_n} \left\{ \mathbb{E} \|e_1\|^4 \mathbb{E} \|\varepsilon_1\|^4 \right\}^{1/2},
$$

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

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

so the condition

$$\max_{1 \leq j, l \leq n} \left\| \frac{1}{n} \sum_{t=1}^{n-j} \omega_t (l) \right\|_{L_1} \to 0. \tag{66}$$

is sufficient to show that $\sigma_{2n} \to_p 0$. The definition of $\omega_t (l)$ in (65) implies that

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

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 $L_n \to \infty$ and $L_n/n \to 0$. By the Lyapounov inequality and the martingale difference property of $\tilde{\omega}_t (l)$,

$$\left\| \frac{1}{n} \sum_{t=1}^{n-j} \tilde{\omega}_t (l) \right\|_{L_1} \leq \frac{1}{n} \left\| \sum_{t=1}^{n-j} \tilde{\omega}_t (l) \right\|_{L_2} = \frac{1}{n} \left(\sum_{t=1}^{n-j} \mathbb{E} \|\tilde{\omega}_t (l)\|^2 \right)^{1/2}$$

$$\leq b \frac{L_n^{1/2}}{n} \left(\sum_{t=1}^{n-j} \mathbb{E} \|w_{t+l}\|^2 \right)^{1/2} \leq \frac{2b L_n^{1/2}}{\sqrt{n}} \left(\mathbb{E} \|\varepsilon_1\|^4 \right)^{1/2} \to 0$$

uniformly in j, l. We conclude that

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

$$\leq \max_{1 \leq j, l \leq n} \frac{1}{n} \sum_{t=1}^{n-j} \mathbb{E} \left\| \omega_t (l) \mathbf{1} \left\{ \|e_t\|^4 > L_n \right\} \right\| + o(1)$$

$$\leq \max_{1 \leq j, l \leq n} \frac{1}{n} \sum_{t=1}^{n-j} \mathbb{E} \left[\left(\|e_t\|^2 + \|\Sigma_{ee}\| \right) \mathbf{1} \left\{ \|e_t\|^4 > L_n \right\} \|w_{t+l}\| \right] + o(1)$$

$$\leq \left\{ \mathbb{E} \left[\left(\|e_1\|^4 + \|\Sigma_{ee}\|^4 \right) \mathbf{1} \left\{ \|e_1\|^4 > L_n \right\} \right] E \|w_1\|^2 \right\}^{1/2} + o(1)$$

$$= o(1)$$

since $E \|w_1\|^2 \leq 4\mathbb{E} \|\varepsilon_1\|^4$, $\mathbb{E} \|e_1\|^4 < \infty$ and $L_n \to \infty$, the last inequality following by the Cauchy–Schwarz inequality. This proves (66) and $\sigma_n \to_p 0$.

We turn to the second term of the bound (22) of Lemma 3.3(i), to prove that $s_n \to_p 0$. Using the solution (60) and the same argument leading to (64), we obtain

$$
s_n \leq \frac{1}{n\kappa_n} \left\| \sum_{j=1}^{n-1} \sum_{t=2}^{n-j} \tilde{S}_{n,t}(j) \right\|
$$

$$
\leq \frac{1}{n\kappa_n} \sum_{j=1}^{n-1} \left\| M_{n,\kappa}^j \right\| \sum_{t=2}^{n} \left\| \tilde{S}_{n,t}(0) \right\| + \frac{1}{n\kappa_n} \left\| \sum_{j=1}^{n-1} \sum_{t=2}^{n-j} \sum_{l=1}^{j} M_{n,\kappa}^{j-l} \tilde{\upsilon}_{n,t}(l) \right\|
$$

$$
= s_{1n} + s_{2n}.
$$

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

$$
\mathbb{E} \left\| s_{1n} \right\| \leq b \frac{1}{\kappa_n} \max_{t \leq n} \mathbb{E} \left\| \tilde{S}_{n,t}(0) \right\| \leq b \max_{t \leq n} \max_{0 \leq i \leq \kappa-1} \frac{1}{\kappa_n} \mathbb{E} \left(\left\| e_t \right\| \left\| \zeta_{n,t-1} \right\| \left\| H_{t-i} \right\| \right)
$$

$$
\leq \frac{1}{\kappa_n^{1/2}} \left\| H_1 \right\|_{L_2} \left\| e_1 \right\|_{L_4} \max_{t \leq n} \left\| \frac{\zeta_{n,t-1}}{\kappa_n^{1/2}} \right\|_{L_4} = O \left(\frac{1}{\kappa_n^{1/2}} \right)
$$

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

$$
s_{2n} = \frac{1}{n\kappa_n} \left\| \sum_{j=1}^{n-2} \sum_{t=2}^{n-j} \sum_{l=0}^{j-1} M_{n,\kappa}^l \tilde{\upsilon}_{n,t}(j-l) \right\| = \frac{1}{n\kappa_n} \left\| \sum_{l=0}^{n-3} M_{n,\kappa}^l \sum_{j=l+1}^{n-2} \sum_{t=2}^{n-j} \tilde{\upsilon}_{n,t}(j-l) \right\|
$$

$$
\leq \sum_{l=0}^{n-3} \left\| M_{n,\kappa}^l \right\| \frac{1}{n\kappa_n} \left\| \sum_{j=1}^{n-l-2} \sum_{t=2}^{n-l-j} \tilde{\upsilon}_{n,t}(j) \right\| = \sum_{l=0}^{n-3} \left\| M_{n,\kappa}^l \right\| \frac{1}{n\kappa_n} \left\| \sum_{j=1}^{n-l-2} \sum_{t=2}^{n-l-j} \upsilon_{n,t}(j) \right\|.
$$

Substituting the expression for $\upsilon_{n,t}(j)$ in (25) we obtain

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

$$
+ \left\| \varphi \right\| \left\| F(1) \right\| \sum_{l=0}^{n-3} \left\| M_{n,\kappa}^l \right\| \frac{1}{\kappa_n} \sum_{j=1}^{n-l-2} \left\| R_n^j \right\|^2 \left\| \frac{1}{n} \sum_{t=2}^{n-l-j} \zeta_{n,t-1} \otimes e_t \right\|.
$$

The last term on the right converges to 0 in L_1 by (47). For the first term, partitioning the third sum into $j \leq i$ and $j > i$ we obtain that $s_{2n} \leq s_{3n} + s_{4n} + o_p(1)$ where

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

and, using the Cauchy–Schwarz inequality twice,

$$
\mathbb{E}s_{4n} \leq \|F(1)\| \sum_{l=0}^{n-3} \left\|M_{n,\kappa}^l\right\| \sum_{i=1}^{q} \|A_i\| \frac{1}{n\kappa_n} \sum_{j=1}^{q} \sum_{t=2}^{n-l-j} \mathbb{E}\left\|\zeta_{n,t-1}\right\| \|e_t\| \|w_{t+j-i}\|
$$

$$
\leq q\|F(1)\| \sum_{l=0}^{n-3} \left\|M_{n,\kappa}^l\right\| \sum_{i=1}^{q} \|A_i\| \max_{t\leq n} \left\|\frac{\zeta_{n,t}}{\kappa_n}\right\|_{L_4} \|e_1\|_{L_4} \|w_1\|_{L_2} = O\left(\frac{1}{\kappa_n^{1/2}}\right).
$$

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

$$
\sum_{j=i+1}^{n-l-2} \sum_{t=2}^{n-l-j} R_n^j \zeta_{n,t-1} \otimes R_n^j F(1) e_t \otimes w_{t+j-i}
$$

$$
= \sum_{j=1}^{n-l-i-2} \sum_{t=2}^{n-l-j-i} R_n^{j+i} \zeta_{n,t-1} \otimes R_n^{j+i} F(1) e_t \otimes w_{t+j} = \left(R_n^i \otimes R_n^i\right) \sum_{t=3}^{n-l-i} \xi_{n,t-1} \otimes w_t,
$$

where $\xi_{n,t-1} = \sum_{j=1}^{t-2} R_n^j \zeta_{n,t-j-1} \otimes R_n^j F(1) e_{t-j}$ is a $\mathcal{F}_{t-1}$-martingale array satisfying

$$
\max_{t\leq n} \mathbb{E}\left\|\xi_{n,t-1}\right\|^2 = \|F(1)\|^2 \max_{t\leq n} \sum_{j=1}^{t-2} \left\|R_n^j\right\|^4 \mathbb{E}\left(\left\|\zeta_{n,t-j-1}\right\|^2 \left\|e_{t-j}\right\|^2\right)
$$

$$
\leq \|F(1)\|^2 \sum_{j=1}^{n-2} \left\|R_n^j\right\|^4 \|e_1\|_{L_4}^2 \max_{t\leq n} \left\|\zeta_{n,t}\right\|_{L_4}^2 = O\left(\kappa_n^2\right). \tag{67}
$$

In the above notation, Lemma 3.1(ii) implies that, for some uniform constant B, s_{3n} satisfies

$$
\mathbb{E}s_{3n} \leq B\frac{1}{n\kappa_n} \max_{i,l} \mathbb{E}\left\|\sum_{t=3}^{n-l-i} \xi_{n,t-1} \otimes w_t\right\|
$$

$$
\leq B\frac{1}{n\kappa_n} \max_{i,l} \mathbb{E}\left\|\sum_{t=3}^{n-l-i} \xi_{n,t-1} \otimes w_t \mathbf{1}\{\|H_t\| \leq L_n\}\right\|
$$

$$
+ B\frac{1}{n\kappa_n} \max_{i,l} \mathbb{E}\left\|\sum_{t=3}^{n-l-i} \xi_{n,t-1} \otimes w_t \mathbf{1}\{\|H_t\| > L_n\}\right\|
$$

$$
= B\left(\epsilon_{1n} + \epsilon_{2n}\right).
$$

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

$$
w_t = \text{vech}\left(\varepsilon_t \varepsilon_t' - H_t\right) = \text{vech}H_t^{1/2}\left(\eta_t \eta_t' - I_m\right) H_t^{1/2}
$$

$$
= D_m^+\left(H_t^{1/2} \otimes H_t^{1/2}\right) \text{vec}\left(\eta_t \eta_t' - I_m\right),
$$

where D_m^+ is the Moore–Penrose inverse of the duplication matrix D_m, satisfying $\left\|D_m^+\right\| = 1$ (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 $\mathbf{1}\{\|H_t\| \le L_n\}\xi_{n,t-1} \otimes w_t$ giving

$$\epsilon_{1n} \le \frac{1}{n\kappa_n} \left\{ \max_{i,l} \sum_{t=3}^{n-l-i} \mathbb{E}\left\|\xi_{n,t-1}\right\|^2 \|w_t\|^2 \mathbf{1}\{\|H_t\| \le L_n\} \right\}^{1/2}$$

$$\le \frac{b}{n\kappa_n} \left\{ \max_{i,l} \sum_{t=3}^{n-l-i} \mathbb{E}\left\|\xi_{n,t-1}\right\|^2 \|H_t\|^2 \mathbf{1}\{\|H_t\| \le L_n\} \right\}^{1/2}$$

$$\le \frac{bL_n}{\sqrt{n}\kappa_n} \left\{ \max_{t \le n} \mathbb{E}\left\|\xi_{n,t-1}\right\|^2 \right\}^{1/2} = O\left(\frac{L_n}{n^{1/2}}\right)$$

by (67), where $b = \left\{2\mathbb{E}\left(\|\eta_1\|^4 + 1\right)\right\}^{1/2}$. Taking $L_n \to \infty$ with $L_n/n^{1/2} \to 0$ we can write

$$\epsilon_{2n} \le \frac{1}{n\kappa_n} \sum_{t=3}^{n} \mathbb{E}\left(\left\|\xi_{n,t-1}\right\| \|w_t\| \mathbf{1}\{\|H_t\| > L_n\}\right)$$

$$\le \frac{2}{n\kappa_n} \sum_{t=3}^{n} \mathbb{E}\left(\left\|\xi_{n,t-1}\right\| \|H_t\| \mathbf{1}\{\|H_t\| > L_n\}\right)$$

$$\le \frac{2}{n\kappa_n} \sum_{t=3}^{n} \left\{ \mathbb{E}\left(\left\|\xi_{n,t-1}\right\|^2\right) \mathbb{E}\left(\|H_t\|^2 \mathbf{1}\{\|H_t\| > L_n\}\right) \right\}^{1/2}$$

$$\le \frac{2}{\kappa_n} \left\{ \max_{t \le n} \mathbb{E}\left\|\xi_{n,t-1}\right\|^2 \right\}^{1/2} \left\{ \mathbb{E}\left(\|H_1\|^2 \mathbf{1}\{\|H_1\| > L_n\}\right) \right\}^{1/2}$$

$$= O(1) \left\{ \mathbb{E}\left(\|H_1\|^2 \mathbf{1}\{\|H_1\| > L_n\}\right) \right\}^{1/2} = o(1)$$

by (67) and integrability of $\|H_1\|^2$. This completes the proof of $s_n \to_p 0$. $\blacksquare$

Proof of Theorem 3.4. The Lindeberg condition $\mathcal{L}_n(\delta) = \sum_{t=2}^{n} \mathbb{E}\left(\xi_{nt}^2 \mathbf{1}\{|\xi_{nt}| > \delta\}\right) \to_p 0$ for all $\delta > 0$ with $\xi_{nt} = n^{-1/2}\kappa_n^{-1/2} \|\varepsilon_t\| \|\zeta_{nt-1}\|$ is established as follows: by truncating $\|\varepsilon_t\|^2$ using a sequence $(L_n)_{n\in\mathbb{N}}$ satisfying $L_n \to \infty$ and $L_n/\sqrt{n} \to 0$

$$\mathcal{L}_n(\delta) = \frac{1}{n} \sum_{t=2}^{n} \mathbb{E}\left(\frac{\|\zeta_{nt-1}\|^2}{\kappa_n} \|\varepsilon_t\|^2 \mathbf{1}\left\{\frac{\|\zeta_{nt-1}\|^2}{\kappa_n} \|\varepsilon_t\|^2 > n\delta^2\right\}\right)$$

$$\le \frac{L_n}{n} \sum_{t=2}^{n} \mathbb{E}\left(\frac{\|\zeta_{nt-1}\|^2}{\kappa_n} \mathbf{1}\left\{\frac{\|\zeta_{nt-1}\|^2}{\kappa_n} > \frac{n\delta^2}{L_n}\right\}\right)$$

$$+ \frac{1}{n} \sum_{t=2}^{n} \mathbb{E}\left(\frac{\|\zeta_{nt-1}\|^2}{\kappa_n} \|\varepsilon_t\|^2 \mathbf{1}\left\{\|\varepsilon_t\|^2 > L_n\right\}\right)$$

$$\leq \max_{t \leq n} \left\| \frac{\zeta_{nt}}{\kappa_n^{1/2}} \right\|_{L_4}^2 \left[L_n \max_{t \leq n} \mathbb{P} \left(\frac{\|\zeta_{nt-1}\|^2}{\kappa_n} > \frac{n\delta^2}{L_n} \right) + \mathbb{E} \left(\|\varepsilon_1\|^4 \mathbf{1} \{ \|\varepsilon_1\|^4 > L_n \} \right) \right]$$

$$\leq O(1) \left[\frac{L_n^2}{n} \frac{1}{\delta^2} \max_{t \leq n} \mathbb{E} \left(\left\| \frac{\zeta_{nt}}{\kappa_n^{1/2}} \right\|^2 \right) + o(1) \right] \to 0$$

from the choice of $(L_n)_{n \in \mathbb{N}}$, since $\max_{t \leq n} \left\| \zeta_{nt}/\kappa_n^{1/2} \right\|_{L_4}^2 = O(1)$ by Lemma 3.2(i) and (ε_t) is a strictly stationary sequence satisfying $\mathbb{E} \|\varepsilon_1\|^4 < \infty$, where the second inequality follows by the Cauchy–Schwarz inequality and the third inequality follows by the Markov inequality.

Part (i) is now an immediate consequence of established results. By Lemma 3.2(iii), the martingale transform N_n in (16) and $(n\kappa_n)^{-1/2} \sum_{t=1}^n x_{t-1} \otimes \varepsilon_t$ have the same limit distribution. By Lemma 3.3 (i) and (iii), the predictable quadratic variation of N_n in (17) has the same the limit distribution as $n^{-1}\kappa_n^{-1} \sum_{t=1}^n \zeta_{n,t-1}\zeta'_{n,t-1} \otimes \Sigma_{\varepsilon\varepsilon}$, so that

$$\langle N \rangle_n = \frac{1}{n\kappa_n} \sum_{t=1}^n \zeta_{n,t-1}\zeta'_{n,t-1} \otimes \Sigma_{\varepsilon\varepsilon} + o_p(1) \to_p V_{xx} \otimes \Sigma_{\varepsilon\varepsilon},$$

where the last convergence in probability follows by Lemma 3.2(ii). Having verified the Lindeberg condition, we may apply a standard martingale central limit theorem, e.g., Corollary 3.1 of Hall and Heyde (1980), to establish the asymptotic distribution of part (i). Part (ii) follows immediately from part (i), Lemma 3.2(ii) and (9). For completeness, we provide a proof of (9): Letting $\underline{x}_t = x_t - \bar{x}_{n-1}$ and $\underline{\varepsilon}_t = \varepsilon_t - \bar{\varepsilon}_n$, the fact that $\sum_{t=1}^n x_{t-1} = O_p(\sqrt{n\kappa_n})$ implies that

$$\frac{1}{n\kappa_n} \left\| \sum_{t=1}^n \underline{x}_t \underline{x}'_t - \sum_{t=1}^n x_t x'_t \right\| \leq \frac{1}{\kappa_n} \|\bar{x}_{n-1}\|^2 = O_p \left(\frac{\kappa_n}{n} \right)$$

and $(n\kappa_n)^{-1/2} \left\| \sum_{t=1}^n \underline{x}_{t-1} \underline{\varepsilon}'_t - \sum_{t=1}^n x_{t-1} \varepsilon'_t \right\| = (n\kappa_n)^{-1/2} \left\| n\bar{x}_{n-1} \bar{\varepsilon}'_n \right\| = O_p \left(\sqrt{\kappa_n/n} \right)$. Combining the two remainder terms proves (9). Note that the same orders of magnitude apply for the purely stationary case, by putting $\kappa_n = 1$. Part (iii) follows immediately by part (ii), (9), and $n^{-1}\kappa_n^{-1} \left\| \underline{X}'\underline{X} - X'X \right\| = o_p(1)$. ∎

Proof of Lemma 4.3. We first note that the effect of the initial condition x_0 satisfying Assumption IC is asymptotically negligible for OLS/IVX sample moments as shown in MP (2020) via arguments that do not depend on the conditional homoskedasticity assumption maintained in that paper. We therefore proceed with setting $x_0 = 0$ without loss of generality. For part (i), we employ the decompositions

$$\tilde{z}_{0t} = x_{0t} + \frac{C_z}{\kappa_{nz}} \psi_{nt}^{(2)}, \quad \frac{\kappa_n}{\kappa_{nz}} \to 0, \tag{68}$$

$$\tilde{z}_{0t} = z_t + \frac{1}{\kappa_n} \psi_{nt}^{(1)}, \quad \frac{\kappa_{nz}}{\kappa_n} \to 0, \tag{69}$$

see equations (17) and (19) of MP (2020) with $x_0 = 0$, where

$$\psi_{nt}^{(1)} = \sum_{j=1}^{t-1} R_{nz}^{t-1-j} C_n x_{0j}, \quad \psi_{nt}^{(2)} = \sum_{j=1}^{t-1} R_{nz}^{t-1-j} x_{0j}$$

and $x_{0t} = \sum_{j=0}^{t-1} R_n^j u_{xt-j}$ is the regressor x_t in (1) with zero initialization $x_0 = 0$. We first establish the bound

$$\max_{1 \le t \le n} \mathbb{E} \left\| \psi_{nt}^{(1)} \right\|^4 \bigvee \max_{1 \le t \le n} \mathbb{E} \left\| \psi_{nt}^{(2)} \right\|^4 = O\left(\kappa_n^2 \kappa_{nz}^4\right). \tag{70}$$

The Minkowski inequality gives

$$\mathbb{E} \left\| \psi_{nt}^{(1)} \right\|^4 = \mathbb{E} \left\| \sum_{j=1}^{t-1} R_{nz}^{t-1-j} C_n x_{0j} \right\|^4 \le \left(\sum_{j=1}^{t-1} \left\{ \mathbb{E} \left\| R_{nz}^{t-1-j} C_n x_{0j} \right\|^4 \right\}^{1/4} \right)^4$$

$$\le \|C_n\|^4 \left\{ \max_{1 \le t \le n} \mathbb{E} \left\| x_{0j} \right\|^4 \right\} \left(\sum_{j=1}^{n} \left\| R_{nz}^j \right\| \right)^4 = O\left(\kappa_n^2 \kappa_{nz}^4\right)$$

uniformly in $t \le n$. The above argument with C_n replaced by I_r yields the same bound applies to $\mathbb{E} \left\| \psi_{nt}^{(2)} \right\|^4$. Employing the decomposition (68) with $x_0 = 0$ we obtain

$$\max_{1 \le t \le n} \mathbb{E} \|\tilde{z}_{0t}\|^4 \le 8 \max_{1 \le t \le n} \mathbb{E} \|x_{0t}\|^4 + \frac{8 \|C_z\|^4}{\kappa_{nz}^4} \max_{1 \le t \le n} \mathbb{E} \left\| \psi_{nt}^{(2)} \right\|^4 = O\left(\kappa_n^2\right)$$

by Lemma 3.2(i) and (70). Employing the decomposition (69) with $x_0 = 0$ we obtain

$$\max_{1 \le t \le n} \mathbb{E} \|\tilde{z}_{0t}\|^4 \le 8 \max_{1 \le t \le n} \mathbb{E} \|z_t\|^4 + \frac{8}{\kappa_n^4} \max_{1 \le t \le n} \mathbb{E} \left\| \psi_{nt}^{(1)} \right\|^4$$

$$= O\left(\kappa_{nz}^2\right) + O\left(\frac{\kappa_{nz}^4}{\kappa_n^2} \right) = O\left(\kappa_{nz}^2\right)$$

when $\kappa_{nz}/\kappa_n \to 0$ by (70) and Lemma 3.2(i) (since z_t is κ_{nz}-near stationary). This proves part (i).

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

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

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

$$\frac{1}{n\left(\kappa_n \wedge \kappa_{nz}\right)}\hat{\Upsilon}_n = \frac{1}{n\left(\kappa_n \wedge \kappa_{nz}\right)}\sum_{t=1}^{n}\left(\tilde{z}_{0t-1}\tilde{z}'_{0t-1} \otimes \underline{\varepsilon}_t\underline{\varepsilon}'_t\right) + o_p(1) \tag{71}$$

provided that both

$$r_{1n} = \frac{1}{\sqrt{n\kappa_n}}\frac{1}{n\left(\kappa_n \wedge \kappa_{nz}\right)}\sum_{t=1}^{n}\left(\tilde{z}_{0t-1}\tilde{z}'_{0t-1} \otimes \underline{x}_{0t-1}\underline{\varepsilon}'_t\right)$$

$$r_{2n} = \frac{1}{n\kappa_n}\frac{1}{n\left(\kappa_n \wedge \kappa_{nz}\right)}\sum_{t=1}^{n}\left(\tilde{z}_{0t-1}\tilde{z}'_{0t-1} \otimes \underline{x}_{0t-1}\underline{x}'_{0t-1}\right)$$

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

$$\mathbb{E}\left\|\bar{x}_{0n}\right\|^4 \leq \frac{1}{n^4}\mathbb{E}\left(\sum_{t=1}^{n}\|x_{0t}\|\right)^4 \leq \frac{1}{n^4}\left(\sum_{t=1}^{n}\|x_{0t}\|_{L_4}\right)^4 \leq \max_{1\leq t\leq n}\mathbb{E}\|x_{0t}\|^4$$

by the Minkowski inequality, so we may write

$$\max_{1\leq t\leq n}\mathbb{E}\left\|\underline{x}_{0t}\right\|^4 \leq \max_{1\leq t\leq n}\mathbb{E}\left(\|x_{0t}\| + \|\bar{x}_{0n}\|\right)^4 \leq 16\max_{1\leq t\leq n}\mathbb{E}\|x_{0t}\|^4.$$

For r_{1n}, the Cauchy–Schwarz inequality gives

$$\begin{aligned}
\mathbb{E}\|r_{1n}\| &\leq \frac{1}{\sqrt{n\kappa_n}}\frac{1}{\kappa_n \wedge \kappa_{nz}}\max_{t\leq n}\mathbb{E}\left(\|\tilde{z}_{0t-1}\|^2\|\underline{x}_{0t-1}\|\|\underline{\varepsilon}_t\|\right) \\
&\leq \frac{1}{\sqrt{n\kappa_n}}\frac{1}{\kappa_n \wedge \kappa_{nz}}\|\varepsilon_1\|_{L_4}\left(\max_{t\leq n}\mathbb{E}\|\tilde{z}_{0t-1}\|^4\right)^{1/2}\left(\max_{t\leq n}\mathbb{E}\|\underline{x}_{0t-1}\|^4\right)^{1/4} \\
&\leq 4\|\varepsilon_1\|_{L_4}\frac{1}{\sqrt{n\kappa_n}}\frac{1}{\kappa_n \wedge \kappa_{nz}}\left(\max_{t\leq n}\mathbb{E}\|\tilde{z}_{0t-1}\|^4\right)^{1/2}\left(\max_{t\leq n}\mathbb{E}\|x_{0t-1}\|^4\right)^{1/4} \\
&= O\left(\frac{1}{\sqrt{n\kappa_n}}\frac{1}{\kappa_n \wedge \kappa_{nz}}\left(\kappa_n \wedge \kappa_{nz}\right)\sqrt{\kappa_n}\right) = O\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

by part (i) and Lemma 3.2(i). Similarly,

$$\begin{aligned}
\mathbb{E}\|r_{2n}\| &\leq \frac{1}{n\kappa_n}\frac{1}{\kappa_n \wedge \kappa_{nz}}\max_{t\leq n}\mathbb{E}\left(\|\tilde{z}_{0t-1}\|^2\|\underline{x}_{0t-1}\|^2\right) \\
&\leq \frac{1}{n\kappa_n}\frac{1}{\kappa_n \wedge \kappa_{nz}}\left\{\max_{t\leq n}\mathbb{E}\|\tilde{z}_{t-1}\|^4\right\}^{1/2}\left\{\max_{t\leq n}\mathbb{E}\|\underline{x}_{t-1}\|^4\right\}^{1/2} = O\left(\frac{1}{n}\right).
\end{aligned}$$

Returning to (71), we can write

$$
\frac{1}{n\left(\kappa_n \wedge \kappa_{nz}\right)} \hat{\Upsilon}_n = \frac{1}{n\left(\kappa_n \wedge \kappa_{nz}\right)} \sum_{t=1}^{n} \left(\tilde{z}_{0t-1}\tilde{z}'_{0t-1} \otimes \varepsilon_t \varepsilon'_t\right) + O_p\left(\frac{1}{\sqrt{n}}\right)
$$

$$
= \begin{cases} \frac{1}{n\kappa_{nz}} \sum_{t=1}^{n} \left(z_{t-1}z'_{t-1} \otimes \varepsilon_t \varepsilon'_t\right) + O_p\left(\sqrt{\frac{\kappa_{nz}}{\kappa_n}}\right), & \frac{\kappa_{nz}}{\kappa_n} \to 0 \\ \frac{1}{n\kappa_n} \sum_{t=1}^{n} \left(x_{t-1}x'_{t-1} \otimes \varepsilon_t \varepsilon'_t\right) + O_p\left(\sqrt{\frac{\kappa_n}{\kappa_{nz}}}\right), & \frac{\kappa_n}{\kappa_{nz}} \to 0 \end{cases} \tag{72}
$$

from which the result when $\kappa_n \to \infty$ follows by Lemma 3.3, since z_t is a κ_{nz}-near stationary process. For $\kappa_n = 1$, letting $\check{x}_t$ in (36) be a strictly stationary version of the stable autoregression $x_t = R^t x_0 + \sum_{j=0}^{t-1} R^j u_{t-j}$,

$$
\frac{1}{n} \sum_{t=1}^{n} \left\| x_{t-1} - \check{x}_{t-1} \right\| \left\| x_{t-1} \right\| \left\| \varepsilon_t \right\|^2 \leq \frac{1}{n} \left\| x_0 \right\| \sum_{t=1}^{n} \left\| R^t \right\| \left\| x_{t-1} \right\| \left\| \varepsilon_t \right\|^2
$$

$$
+ \frac{1}{n} \sum_{t=1}^{n} \left\| R^t \right\| \left\| x_{t-1} \right\| \left\| \varepsilon_t \right\|^2 \sum_{j=0}^{\infty} \left\| R^j \right\| \left\| u_{-j} \right\|
$$

$$
= O_p\left(\frac{1}{n}\right)
$$

and then the second part of (72) gives

$$
\frac{1}{n} \hat{\Upsilon}_n = \frac{1}{n} \sum_{t=1}^{n} \left(\check{x}_{t-1}\check{x}'_{t-1} \otimes \varepsilon_t \varepsilon'_t\right) + o_p(1) \to_p \mathbb{E}\left(\check{x}_1 \check{x}'_1 \otimes \varepsilon_2 \varepsilon'_2\right)
$$

by the ergodic theorem.

It remains to show (72). For the $\kappa_{nz}/\kappa_n \to 0$ part, the decomposition (69) implies that

$$
\eta_{1n} = \frac{1}{n\kappa_{nz}\kappa_n} \sum_{t=1}^{n} \left\| z_{t-1} \right\| \left\| \psi_{nt-1}^{(1)} \right\| \left\| \varepsilon_t \right\|^2 = o_p(1)
$$

$$
\eta_{2n} = \frac{1}{n\kappa_{nz}\kappa_n^2} \sum_{t=1}^{n} \left\| \psi_{nt-1}^{(1)} \right\|^2 \left\| \varepsilon_t \right\|^2 = o_p(1)
$$

are sufficient for (72). Using the Cauchy–Schwarz inequality we obtain

$$
\mathbb{E}\eta_{1n} \leq \frac{\left\| \varepsilon_1 \right\|_{L_4}^2}{\kappa_{nz}\kappa_n} \max_{t \leq n} \left\| z_t \right\|_{L_4} \max_{t \leq n} \left\| \psi_{nt}^{(1)} \right\|_{L_4} = O\left(\sqrt{\frac{\kappa_{nz}}{\kappa_n}}\right)
$$

by (70) and Lemma 3.2(i), and, similarly,

$$
\mathbb{E}\eta_{2n} \leq \left\| \varepsilon_1 \right\|_{L_4}^2 \frac{1}{\kappa_{nz}\kappa_n^2} \max_{t \leq n} \left\| \psi_{nt}^{(1)} \right\|_{L_4}^2 = O\left(\frac{\kappa_{nz}}{\kappa_n}\right)
$$

which proves (72) when $\kappa_{nz}/\kappa_n \to 0$. When $\kappa_n/\kappa_{nz} \to 0$, the same argument can be applied to the decomposition (68). ∎

Proof of Theorem 4.4. Under Assumptions P(i)–(ii), $\kappa_n \to \infty$, Lemma 4.3 shows that

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

which implies that

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

and hence that $\left\|\hat{W}_n - \tilde{W}_n\right\| = o_p\left(1\right)$ under the null hypothesis (27). Corollary 4.2 then gives $\hat{W}_n \Rightarrow \chi^2\left(q\right)$.

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

$$n\hat{Q}_n = \left[\left(\frac{1}{n}X'X\right)^{-1} \otimes I_m\right]\frac{1}{n}\hat{\Upsilon}_n\left[\left(\frac{1}{n}X'X\right)^{-1} \otimes I_m\right] + o_p\left(I\right) \to_p V_0,$$

where the matrix V_0 is defined in (37). We can then write $\hat{W}_n = w_n'w_n + o_p\left(1\right)$ where, under (27),

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

by Theorem 4.1(ii). ∎

REFERENCES

Abadir, K.M. & J.R. Magnus (2005) *Matrix Algebra*. Econometric Exercises, vol.1. Cambridge University Press.

Andrews, D.W.K. & P. Guggenberger (2008) Asymptotics for stationary very nearly unit root processes. *Journal of Time Series Analysis* 29(1), 203–212.

Andrews, D.W.K. & P. Guggenberger (2012) Asymptotics for LS, GLS, and feasible GLS statistics in an AR(1) model with conditional heteroskedasticity. *Journal of Econometrics* 169(2), 196–210.

Andrews, D.W.K. & P. Guggenberger (2014) A conditional-heteroskedasticity-robust confidence interval for the autoregressive parameter. *Review of Economics and Statistics* 96(2), 376–381.

Bollersev, T. (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'Academie des Sciences Paris* 343, 275–278.

Boussama, F., F. Fuchs, & R. Stelzer (2011) Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and Applications* 121, 2331–2360.

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987–1008.

Francq, C. & J.-M. Zakoian (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Giraitis, L., H.L. Koul & D. Surgailis (2012) *Large Sample Theory for Long Memory Processes*. Imperial College Press.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and its Application*. Academic Press.

Horn, R.A. & C.R. Johnson (2013) *Matrix Analysis*. Cambridge University Press.

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

Kostakis, A., T. Magdalinos, & M.P. Stamatogiannis (2015b). Online Appendix to: Robust Econometric Inference for Stock Return Predictability.

Ling, S. & 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. & 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. & M. McAleer (2003) Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 280–310.

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

Magdalinos, T. & P.C.B. Phillips (2020) Econometric Inference in Matrix Vicinities of Unity and Stationarity. Working paper.

Pantula, S.G. (1989) Estimation of autoregressive models with ARCH errors. *Sankhya B* 50, 119–138.

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. & T. Magdalinos (2007a) Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130.

Phillips, P.C.B. & 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, 123–162.

Phillips, P.C.B. & T. Magdalinos (2009a) Econometric Inference in the Vicinity of Unity. Working paper, Singapore Management University.

Phillips, P.C.B. & T. Magdalinos (2009b) Unit root and cointegrating limit theory when the initialization is in the infinite past. *Econometric Theory* 25, 1682–1715.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Weiss, A.A. (1986) Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory* 2, 107–131.

White, H. (1980) A Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.