ASYMPTOTICS FOR LS, GLS, AND FEASIBLE GLS STATISTICS IN AN AR(1) MODEL WITH CONDITIONAL HETEROSKEDASTICITY

By

Donald W.K. Andrews, Patrik Guggenberger

COWLES FOUNDATION PAPER NO. 1365

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2012

Journal of Econometrics 169 (2012) 196–210

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Asymptotics for LS, GLS, and feasible GLS statistics in an AR(1) model with conditional heteroskedasticity

Donald W.K. Andrews [a,*], Patrik Guggenberger [b]

[a] Cowles Foundation for Research in Economics, Yale University, United States

[b] Department of Economics, U.C.S.D., United States

ARTICLE INFO

Article history:
Available online 24 January 2012

JEL classification:
C22

Keywords:
Asymptotic distribution
Autoregression
Conditional heteroskedasticity
Generalized least squares
Least squares

ABSTRACT

We consider a first-order autoregressive model with conditionally heteroskedastic innovations. The asymptotic distributions of least squares (LS), infeasible generalized least squares (GLS), and feasible GLS estimators and t statistics are determined. The GLS procedures allow for misspecification of the form of the conditional heteroskedasticity and, hence, are referred to as quasi-GLS procedures. The asymptotic results are established for drifting sequences of the autoregressive parameter ρ_n and the distribution of the time series of innovations. In particular, we consider the full range of cases in which ρ_n satisfies $n(1-\rho_n) \to \infty$ and $n(1-\rho_n) \to h_1 \in [0, \infty)$ as $n \to \infty$, where n is the sample size. Results of this type are needed to establish the uniform asymptotic properties of the LS and quasi-GLS statistics.

1. Introduction

We are very happy to contribute this paper to the Special Issue in Honor of Peter C.B. Phillips. The topic of the paper is the first-order autoregressive AR(1) model with a stationary, unit, or near unit root. This is a topic to which Peter Phillips has made seminal contributions over several decades ranging from Phillips (1977) to Phillips and Magdalinos (2007). The current paper considers an AR(1) model with conditional heteroskedasticity and, hence, is closely related to Guo and Phillips (2001).

This paper establishes the asymptotic distributions of LS and quasi-GLS statistics in an AR(1) model with intercept and conditional heteroskedasticity. The LS and GLS procedures allow for misspecification of the form of the conditional heteroskedasticity and, hence, are referred to as quasi-GLS procedures. The statistics considered include infeasible and feasible quasi-GLS estimators, heteroskedasticity-consistent (HC) standard error estimators, and the t statistics formed from these estimators. The paper considers: (i) the stationary and near stationary case, where the autoregressive parameter ρ_n satisfies $n(1-\rho_n) \to \infty$ as $n \to \infty$ and (ii) the unit-root and near unit-root case, where $n(1-\rho_n) \to h_1 \in [0, \infty)$. Our interest in asymptotics under drifting sequences of parameters is due to the fact that near unit-root asymptotics are well-known to provide better finite-sample approximations than fixed parameter asymptotics for parameter values that are close to, but different from, unity. In addition, uniform asymptotic results rely on asymptotic results under drifting sequences of parameters, see Andrews and Guggenberger (2010a).

In case (i), the quasi-GLS t statistic is shown to have a standard normal asymptotic distribution. In case (ii), its asymptotic distribution is shown to be that of a convex linear combination of a random variable with a "demeaned near unit-root distribution" and an independent standard normal random variable. The weights on the two random variables depend on the correlation between the innovation, say U_i, and the innovation rescaled by the quasi-conditional variance, say U_i/ϕ_i^2. Here ϕ_i^2 is the (possibly misspecified) conditional variance used by the GLS estimator. In the case of LS, we have $\phi_i^2 = 1$, the correlation between U_i and U_i/ϕ_i^2 is one, and the asymptotic distribution is a demeaned near unit-root distribution (based on an Ornstein–Uhlenbeck process).

For an AR(1) model without conditional heteroskedasticity, case (i) is studied by Park (2002), Giraitis and Phillips (2006), and Phillips and Magdalinos (2007). An AR(1) model with conditional heteroskedasticity and $\rho = 1$, which falls within case (ii) above, has been considered by Seo (1999) and Guo and Phillips (2001). The results given here make use of ideas in these two papers. Case (ii) is the "near integrated" case that has been studied in AR models without conditional heteroskedasticity by Bobkowski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987), Elliott (1999), Elliott and Stock (2001), and Müller and Elliott (2003). The latter three papers consider the situation that also is considered here in which the initial condition yields a stationary process. Gonçalves and Kilian (2004, 2007) consider inference in

* Corresponding author.
E-mail address: donald.andrews@yale.edu (D.W.K. Andrews).

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.01.017

autoregressive models with conditional heteroskedasticity but do not allow for unit roots or roots near unity.

As noted above, in this paper, we consider a heteroskedasticity-consistent (HC) standard error estimator. Such an estimator is needed in order for the quasi-GLS t statistic to have a standard normal asymptotic distribution in case (i) when the form of the conditional heteroskedasticity is misspecified.

The paper provides high-level conditions under which infeasible and feasible quasi-GLS estimators are asymptotically equivalent.[1] The high-level conditions are verified for cases in which the GLS estimator employs a parametric model, with some parameter π, for the form of the conditional heteroskedasticity. For technical reasons, we take the estimator of π to be a discretized estimator and we require the parametric form of the conditional heteroskedasticity to be such that the conditional variance depends upon a finite number of lagged squared innovations. Neither of these conditions is particularly restrictive because (a) the grid size for the discretized estimator can be defined such that there is little difference between the discretized and non-discretized versions of the estimator of π, (b) the parametric model for the conditional heteroskedasticity may be misspecified, and (c) any parametric model with stationary conditional heteroskedasticity, such as a GARCH(1, 1) model, can be approximated arbitrarily well by a model with a large finite number of lags.

The results of this paper are used in Andrews and Guggenberger (2009) to show that symmetric two-sided subsampling confidence intervals (based on the quasi-GLS t statistic described above) have correct asymptotic size in an AR(1) model with conditional heteroskedasticity. (Here "asymptotic size" is defined to be the limit as the sample size n goes to infinity of the exact, i.e., finite-sample, size.) This result requires uniformity in the asymptotics and, hence, relies on asymptotic results in which the autoregressive parameter and the innovation distribution may depend on n. (Triangular array asymptotics are needed to establish uniformity in the asymptotics in a wide variety of models, e.g., see Andrews and Guggenberger, 2010a.) In addition, Andrews and Guggenberger (2009) shows that upper and lower one-sided and symmetric and equal-tailed two-sided hybrid-subsampling confidence intervals have correct asymptotic size. No other confidence intervals in the literature, including those in Stock (1991), Andrews (1993), Andrews and Chen (1994), Nankervis and Savin (1996), Hansen (1999), Chen and Deo (2011), and Mikusheva (2007), have correct asymptotic size in an AR(1) model with conditional heteroskedasticity.

The remainder of the paper is organized as follows. Section 2 introduces the model and statistics considered. Section 3 gives the assumptions, normalization constants, and asymptotic results. Section 4 and Andrews and Guggenberger (2010b) provide proofs of the results.

2. Model, estimators, and t statistic

We use the unobserved components representation of the AR(1) model. The observed time series $\{Y_i : i = 0, \ldots, n\}$ is based on a latent no-intercept AR(1) time series $\{Y_i^* : i = 0, \ldots, n\}$:

$$Y_i = \alpha + Y_i^*,$$
$$Y_i^* = \rho Y_{i-1}^* + U_i, \quad \text{for } i = 1, \ldots, n, \tag{1}$$

where $\rho \in [-1+\varepsilon, 1]$ for some $0 < \varepsilon < 2$, $\{U_i : i = \ldots, 0, 1, \ldots\}$ are stationary and ergodic with conditional mean 0 given a σ-field $\mathcal{G}_{i-1}$ defined at the end of this section, conditional variance $\sigma_i^2 = E(U_i^2|\mathcal{G}_{i-1})$, and unconditional variance $\sigma_U^2 \in (0, \infty)$. The distribution of Y_0^* is the distribution that yields strict stationarity for $\{Y_i^* : i \le n\}$ when $\rho < 1$, i.e., $Y_0^* = \sum_{j=0}^{\infty} \rho^j U_{-j}$, and is arbitrary when $\rho = 1$.

The model can be rewritten as

$$Y_i = \widetilde{\alpha} + \rho Y_{i-1} + U_i, \quad \text{where } \widetilde{\alpha} = \alpha(1-\rho), \tag{2}$$

for $i = 1, \ldots, n$.[2]

We consider a feasible quasi-GLS (FQGLS) estimator of ρ and a t statistic based on it. The FQGLS estimator depends on estimators $\{\widehat{\phi}_{n,i}^2 : i \le n\}$ of the conditional variances $\{\sigma_i^2 : i \le n\}$. The estimators $\{\widehat{\phi}_{n,i}^2 : i \le n\}$ may be from a parametric specification of the conditional heteroskedasticity, e.g., a GARCH(1, 1) model, or from a nonparametric estimator, e.g., one based on q lags of the observations. We do not assume that the conditional heteroskedasticity estimator is consistent. For example, we allow for incorrect specification of the parametric model in the former case and conditional heteroskedasticity that depends on more than q lags in the latter case. The estimated conditional variances $\{\widehat{\phi}_{n,i}^2 : i \le n\}$ are defined such that they approximate a stationary $\mathcal{G}_{i-1}$-adapted sequence $\{\phi_i^2 : i \le n\}$ in the sense that certain normalized sums have the same asymptotic distribution whether $\widehat{\phi}_{n,i}^2$ or ϕ_i^2 appears in the sum. This is a typical property of feasible and infeasible GLS estimators.

As an example, the results (i.e., Theorems 1 and 2 below) allow for the case where (i) $\{\widehat{\phi}_{n,i}^2 : i \le n\}$ are from a GARCH(1, 1) parametric model with parameters π estimated using LS residuals with GARCH and LS parameter estimators $\widetilde{\pi}_n$ and $(\widetilde{\alpha}_n, \widetilde{\rho}_n)$, respectively, (ii) $(\widetilde{\alpha}_n, \widetilde{\rho}_n)$ have a probability limit given by the true values $(\widetilde{\alpha}_0, \rho_0)$, (iii) $\widetilde{\pi}_n$ has a probability limit defined as the "pseudo-true" value π_0, (iv) $\widehat{\phi}_{n,i}^2 = \phi_{i,1}^2(\widetilde{\alpha}_n, \widetilde{\rho}_n, \widetilde{\pi}_n)$, where $\phi_{i,1}^2(\widetilde{\alpha}, \rho, \pi)$ is the i-th GARCH conditional variance based on a start-up at time 1 and parameters $(\widetilde{\alpha}, \rho, \pi)$, and (v) $\phi_{i,-\infty}^2(\widetilde{\alpha}, \rho, \pi)$ is the GARCH conditional variance based on a start-up at time $-\infty$ and parameters $(\widetilde{\alpha}, \rho, \pi)$. In this case, $\phi_i^2 = \phi_{i,-\infty}^2(\widetilde{\alpha}_0, \rho_0, \pi_0)$. Thus, ϕ_i^2 is just $\widehat{\phi}_{n,i}^2$ with the estimation error and start-up truncation eliminated.

Under the null hypothesis that $\rho = \rho_n$, the studentized t statistic is

$$T_n^*(\rho_n) = \frac{n^{1/2}(\widehat{\rho}_n - \rho_n)}{\widehat{\sigma}_n}, \tag{3}$$

where $\widehat{\rho}_n$ is the LS estimator from the regression of $Y_i/\widehat{\phi}_{n,i}$ on $Y_{i-1}/\widehat{\phi}_{n,i}$ and $1/\widehat{\phi}_{n,i}$, and $\widehat{\sigma}_n^2$ is the (1, 1) element of the standard heteroskedasticity-robust variance estimator for the LS estimator in the preceding regression.

To define $T_n^*(\rho_n)$ more explicitly, let Y, U, X_1, and X_2 be n-vectors with ith elements given by $Y_i/\widehat{\phi}_{n,i}$, $U_i/\widehat{\phi}_{n,i}$, $Y_{i-1}/\widehat{\phi}_{n,i}$, and $1/\widehat{\phi}_{n,i}$, respectively. Let Δ be the diagonal $n \times n$ matrix with ith diagonal element given by the ith element of the residual vector $M_X Y$, where $X = [X_1 : X_2]$ and $M_X = I_n - X(X'X)^{-1}X'$. That is, $\Delta = \text{Diag}(M_X Y)$. Then, by definition,

$$\widehat{\rho}_n = \left(X_1' M_{X_2} X_1\right)^{-1} X_1' M_{X_2} Y, \quad \text{and} \tag{4}$$
$$\widehat{\sigma}_n^2 = \left(n^{-1} X_1' M_{X_2} X_1\right)^{-1} \left(n^{-1} X_1' M_{X_2} \Delta^2 M_{X_2} X_1\right) \times \left(n^{-1} X_1' M_{X_2} X_1\right)^{-1}.$$

[1] By definition, the feasible quasi-GLS estimator is based on (possibly misspecified) estimators $\{\widehat{\phi}_{n,i}^2 : i \le n\}$ of the conditional variances of the innovations. The corresponding infeasible quasi-GLS estimator is based on the limits $\{\phi_i^2 : i \le n\}$ of the estimators $\{\widehat{\phi}_{n,i}^2 : i \le n\}$ in the sense of Assumption CHE. If the latter are misspecified, then the true conditional variances are different from $\{\phi_i^2 : i \le n\}$.

[2] By writing the model as in (1), the case $\rho = 1$ and $\widetilde{\alpha} \ne 0$ is automatically ruled out. Doing so is desirable because when $\rho = 1$ and $\widetilde{\alpha} \ne 0$, Y_i is dominated by a deterministic trend and the LS estimator of ρ converges at rate $n^{3/2}$.

We assume $\{(U_i, \phi_i^2) : i \geq 1\}$ are stationary and strong mixing. We define $\mathcal{G}_i$ to be some non-decreasing sequence of σ-fields for $i \geq 1$ for which $(U_j, \phi_{j+1}^2) \in \mathcal{G}_i$ for all $j \leq i$.

3. Asymptotic results

3.1. Assumptions

We let F denote the distribution of $\{(U_i, \phi_i^2) : i = \ldots, 0, 1, \ldots\}$. Our asymptotic results below are established under drifting sequences $\{(\rho_n, F_n) : n \geq 1\}$ of autoregressive parameters ρ_n and distributions F_n. In particular, we provide results for the cases $n(1-\rho_n) \to \infty$ and $n(1-\rho_n) \to h_1 \in [0, \infty)$. When F_n depends on n, $\{(U_i, \phi_i^2) : i \leq n\}$ for $n \geq 1$ form a triangular array of random variables and $(U_i, \phi_i^2) = (U_{n,i}, \phi_{n,i}^2)$. We now specify assumptions on $(U_{n,i}, \phi_{n,i}^2)$. The assumptions place restrictions on the drifting sequence of distributions $\{F_n : n \geq 1\}$ that are considered.

The statistics $\widehat{\rho}_n$, $\widehat{\sigma}_n$, and $T_n^*(\rho_n)$ are invariant to the value of α. Hence, without loss of generality, from now on we take $\alpha = 0$ and $Y_{n,i} = Y_{n,i}^*$.

Let $\lambda_{\min}(A)$ denote the smallest eigenvalue of the matrix A.

Assumption INNOV. (i) For each $n \geq 1$, $\{(U_{n,i}, \phi_{n,i}^2, \sigma_{n,i}^2) : i = \ldots, 0, 1, \ldots\}$ are stationary and strong mixing, where $\sigma_{n,i}^2 = E_{F_n}(U_{n,i}^2 | \mathcal{G}_{n,i-1})$, $E_{F_n}(U_{n,i} \mid \mathcal{G}_{n,i-1}) = 0$ a.s., and $\mathcal{G}_{n,i}$ is some non-decreasing sequence of σ-fields for $i = \ldots, 1, 2, \ldots$ for $n \geq 1$ for which $(U_{n,j}, \phi_{n,j+1}^2) \in \mathcal{G}_{n,i}$ for all $j \leq i$,[3]

(ii) the strong-mixing numbers $\{\alpha_n(m) : m \geq 1\}$ satisfy $\alpha(m) = \sup_{n \geq 1} \alpha_n(m) = O(m^{-3\zeta/(\zeta-3)})$ as $m \to \infty$ for some $\zeta > 3$,

(iii) $\sup_{n,i,s,t,u,v,A} E_{F_n} |\prod_{a \in A} a|^{\zeta} < \infty$, where $0 \leq i, s, t, u, v < \infty$, $n \geq 1$, and A is a non-empty subset of $\{U_{n,i-s}, U_{n,i-t}, U_{n,i+1}^2/\phi_{n,i+1}^4, U_{n,-u}, U_{n,-v}, (U_{n,1}^2 + \sigma_{n,1}^2)/\phi_{n,1}^4\}$ or a subset of $\{U_{n,i-s}, U_{n,i-t}, \phi_{n,i+1}^{-k}, U_{n,-u}, U_{n,-v}, \phi_{n,1}^{-k}\}$ for $k = 2, 3, 4$, $\sup_n E_{F_n}(\sigma_{n,i}^2)^{\zeta} < \infty$, $\inf_n E_{F_n} U_{n,i}^2 \geq \delta > 0$.

(iv) $\phi_{n,i}^2 \geq \delta > 0$ a.s.,

(v) $\lambda_{\min}(E_{F_n}(X^1 X^{1\prime} U_{n,1}^2/\phi_{n,1}^2)) \geq \delta > 0$, where $X^1 = (Y_{n,0}^*/\phi_{n,1}, \phi_{n,1}^{-1})'$, and

(vi) the following limits exist and are positive: $h_{2,1} = \lim_{n\to\infty} E_{F_n} U_{n,i}^2$, $h_{2,2} = \lim_{n\to\infty} E_{F_n}(U_{n,i}^2/\phi_{n,i}^4)$, $h_{2,3} = \lim_{n\to\infty} E_{F_n}(U_{n,i}^2/\phi_{n,i}^2)$, $h_{2,4} = \lim_{n\to\infty} E_{F_n}\phi_{n,i}^{-1}$, $h_{2,5} = \lim_{n\to\infty} E_{F_n}\phi_{n,i}^{-2}$, and $h_{2,6} = \lim_{n\to\infty} E_{F_n}\phi_{n,i}^{-4}$.

Assumptions INNOV(i) and (ii) specify the dependence structure of the innovations. These conditions rule out long-memory innovations, but otherwise are not very restrictive. Assumption INNOV(iii) is a moment condition on the innovations. This assumption can be restrictive because it restricts the thickness of the tails of the innovations and financial time series often have thick tails. It would be desirable to relax this assumption but the current methods of proof, namely the proofs of Lemmas 6–9, require the assumption as stated. Note that the use of the heteroskedasticity-robust variance estimator $\widehat{\sigma}_n^2$ requires stronger moment conditions than would a variance estimator that is designed for homoskedasticity, but the latter would not yield a standard normal asymptotic distribution under stationarity and heteroskedasticity. Assumption INNOV(iv) bounds $\phi_{n,i}^2$ away from zero. This is not restrictive because most conditional variance estimators $\widehat{\phi}_{n,i}^2$ are defined so that they are bounded away from zero. The terms $\phi_{n,i}^2$ then inherit the same property, see Assumption CHE below. Assumption INNOV(v) is a nonsingularity condition that is not very restrictive because $Y_{n,0}^*$ is not equal to a constant. For example, in the trivial case in which $\{U_{n,i} : i \leq n\}$ are i.i.d. and $\phi_{n,i}^2 = 1$, it reduces to $E_{F_n} U_{n,i}^2$ being bounded away from zero. Assumption INNOV(vi) requires that the limits of certain moments exist. This assumption is not very restrictive. For example, it still allows one to establish uniform asymptotic results for tests and confidence intervals, see Andrews and Guggenberger (2009).

We now discuss Assumption INNOV for the example of a correctly-specified GARCH(1, 1) model

$$U_{n,i} = \sigma_{n,i}\varepsilon_{n,i}, \quad \text{for } \{\varepsilon_{n,i}\} \text{ i.i.d. in } i = \ldots, 0, 1, \ldots,$$
$$E_{F_n}\varepsilon_{n,i} = 0, \qquad E_{F_n}\varepsilon_{n,i}^2 = 1, \quad \text{and}$$
$$\sigma_{n,i}^2 = \phi_{n,i}^2 = c_n + \alpha_n U_{n,i}^2 + \beta_n \sigma_{n,i-1}^2 \tag{5}$$

with GARCH innovations $\{\varepsilon_{n,i}\}$ that satisfy $\sup_{n\geq 1} E_{F_n}|\varepsilon_{n,i}|^{6\zeta} < \infty$ with $\zeta = 3 + \varepsilon$ for any small $\varepsilon > 0$ and with GARCH parameters (c_n, α_n, β_n) restricted by $\inf_{n\geq 1} c_n > 0$, $\sup_{n\geq 1} c_n < \infty$, $\sup_{n\geq 1}(\alpha_n + \beta_n) < 1$, $\alpha_n > 0$, $\beta_n \geq 0$ for all n, and the additional restriction $\sup_{n\geq 1} E_{F_n}(\beta_n + \alpha_n \varepsilon_{n,1}^2)^{3\zeta} < 1$.[4] We show in Section 4.1 how these conditions imply the stationarity part of Assumptions INNOV(i), (iii), and (iv). To do so, we use results about GARCH(1, 1) processes given in Bollerslev (1986) and Lindner (2009). Lindner (2009, Theorem 8) states that for given n, $\{(U_{n,i}, \sigma_{n,i}^2) : i = \ldots, 0, 1, \ldots\}$ is strongly mixing with geometric decay rate of the mixing numbers, i.e. $\alpha_n(m) = O(\lambda_n^m)$ as $m \to \infty$ for a $\lambda_n \in (0, 1)$, if in addition $\varepsilon_{n,1}$ is absolutely continuous with Lebesgue density $f_n(x) \geq f(x) > 0$ for all $|x| < \delta$ for some $\delta > 0$ and some function f. For example, this requirement is satisfied if $\varepsilon_{n,1}$ is normally distributed. Therefore, the mixing part of Assumptions INNOV(i) and (ii) holds provided $\sup_{n\geq 1} \lambda_n < 1$. (The latter obviously holds when the GARCH parameters and the distribution of $\varepsilon_{n,1}$ do not depend on n and should hold when they do depend on n given the restrictions that $\sup_{n\geq 1}(\alpha_n + \beta_n) < 1$ and the innovation densities are bounded away from zero in a neighborhood of the origin.) Regarding Assumption INNOV(v), we refer to the discussion above. Assumption INNOV(vi) just requires the existence of certain limits and is innocuous.

We now return to the general case. If $\rho_n = 1$, the initial condition $Y_{n,0}^*$ is arbitrary. If $\rho_n < 1$, then the initial condition satisfies the following assumption:

Assumption STAT. $Y_{n,0}^* = \sum_{j=0}^{\infty} \rho_n^j U_{n,-j}$.

Assumption STAT states that a stationary initial condition is employed when $\rho_n < 1$. If a different initial condition is employed, such as $Y_{n,0}^* = 0$, then the asymptotic distributions in Theorems 1 and 2 below are different in the near unit-root case (which corresponds to $h_1 \in (0, \infty)$ in those theorems). In particular, in (15), the second summand in the definition of $I_h^*(r)$ is attributable to the stationary initial condition.

We determine the asymptotic distributions $\widehat{\rho}_n$, $\widehat{\sigma}_n^2$, and $T_n^*(\rho_n)$ under sequences $\{(\rho_n, F_n) : n \geq 1\}$ such that (a) Assumption INNOV holds and if $\rho_n < 1$ Assumption STAT also holds, and

$$\text{(b) } n(1-\rho_n) \to h_1 \quad \text{for (i) } h_1 = \infty \text{ and (ii) } 0 \leq h_1 < \infty. \tag{6}$$

[3] By "$(U_{n,j}, \phi_{n,j+1}^2) \in \mathcal{G}_{n,i}$ for all $j \leq i$" we mean that the σ-field generated by $\{(U_{n,j}, \phi_{n,j+1}^2) : j \leq i\}$ is a sub-σ-field of $\mathcal{G}_{n,i}$.

[4] E.g., for the case where $\varepsilon_{n,1}$ is $N(0, 1)$ and $\varepsilon = 1/30$, the latter restriction implies that for given α_n, β_n is restricted to the interval $[0, \beta_{\alpha_n}]$, where some values of $(\alpha_n, \beta_{\alpha_n})$ are given as (0.01, 0.98), (0.02, 0.97), (0.03, 0.96), (0.04, 0.94), (0.05, 0.91), (0.06, 0.88), (0.07, 0.83), (0.08, 0.78), (0.09, 0.71), (0.1, 0.62), (0.11, 0.51), (0.12, 0.39), (0.13, 0.25), and (0.14, 0.1). For $\alpha_n \geq 0.15$, the set of possible β_n values is empty.

The asymptotic distributions of $\widehat{\rho}_n$ and $\widehat{\sigma}_n^2$ are shown to depend on the parameters h_1, $h_{2,1}$, and $h_{2,2}$ (where $h_{2,1}$ and $h_{2,2}$ are defined in Assumption INNOV(vi)) and the parameter $h_{2,7}$, which is defined by

$$h_{2,7} = \frac{h_{2,3}}{(h_{2,1}h_{2,2})^{1/2}} = \lim_{n\to\infty} \mathrm{Corr}_{F_n}(U_{n,i}, U_{n,i}/\phi_{n,i}^2). \tag{7}$$

The asymptotic distribution of $T_n^*(\rho_n)$ is shown to depend only on h_1 and $h_{2,7}$.

Define

$$h_2 = (h_{2,1}, \ldots, h_{2,7})' \quad \text{and}$$
$$h = (h_1, h_2')' \in H = R_{+,\infty} \times H_2, \tag{8}$$

where $R_+ = \{x \in R : x \geq 0\}$, $R_{+,\infty} = R_+ \cup \{\infty\}$, and $H_2 \subset (0,\infty)^6 \times (0,1]$.

For notational simplicity, we index the asymptotic distributions of $\widehat{\rho}_n$, $\widehat{\sigma}_n^2$, and $T_n^*(\rho_n)$ by h below (even though they only depend on a subvector of h).

3.2. Normalization constants

The normalization constants a_n and d_n used to obtain the asymptotic distributions of $\widehat{\rho}_n$ and $\widehat{\sigma}_n$, respectively, depend on (ρ_n, F_n) and are denoted $a_n(\rho_n, F_n)$ and $d_n(\rho_n, F_n)$. They are defined as follows. Let $\{\rho_n : n \geq 1\}$ be a sequence for which $n(1-\rho_n) \to \infty$ or $n(1-\rho_n) \to h_1 \in [0,\infty)$. Define the 2-vectors

$$X^1 = (Y_{n,0}^*/\phi_{n,1}, \phi_{n,1}^{-1})' \quad \text{and}$$
$$Z = (1, -E_{F_n}(Y_{n,0}^*/\phi_{n,1}^2)/E_{F_n}(\phi_{n,1}^{-2}))'. \tag{9}$$

Define

$$a_n = a_n(\rho_n, F_n) = n^{1/2} d_n(\rho_n, F_n) \quad \text{and} \tag{10}$$
$$d_n = d_n(\rho_n, F_n) = \begin{cases} \dfrac{E_{F_n}(Y_{n,0}^{*2}/\phi_{n,1}^2) - (E_{F_n}(Y_{n,0}^*/\phi_{n,1}^2))^2/E_{F_n}(\phi_{n,1}^{-2})}{(Z'E_{F_n}(X^1X^{1\prime}U_{n,1}^2/\phi_{n,1}^2)Z)^{1/2}} \\ \quad \text{if } n(1-\rho_n) \to \infty \\ n^{1/2} \quad \text{if } n(1-\rho_n) \to h_1 \in [0,\infty). \end{cases}$$

Note that the normalization constant for the t statistic $T_n^*(\rho_n)$ is $a_n(\rho_n, F_n)/d_n(\rho_n, F_n) = n^{1/2}$.

In certain cases, the normalization constants simplify. In the case where $n(1-\rho_n) \to \infty$ and $\rho_n \to 1$, the constants a_n and d_n in (10) simplify to

$$a_n = n^{1/2} \frac{E_{F_n}(Y_{n,0}^{*2}/\phi_{n,1}^2)}{(E_{F_n}(Y_{n,0}^{*2}U_{n,1}^2/\phi_{n,1}^4))^{1/2}} \quad \text{and}$$
$$d_n = \frac{E_{F_n}(Y_{n,0}^{*2}/\phi_{n,1}^2)}{(E_{F_n}(Y_{n,0}^{*2}U_{n,1}^2/\phi_{n,1}^4))^{1/2}} \tag{11}$$

up to lower order terms. This holds because by Lemma 6 below

$$\begin{aligned} &Z'E_{F_n}(X^1X^{1\prime}U_{n,1}^2/\phi_{n,1}^2)Z \\ &= E_{F_n}(Y_{n,0}^{*2}U_{n,1}^2/\phi_{n,1}^4) - 2E_{F_n}(Y_{n,0}^*U_{n,1}^2/\phi_{n,1}^4) \\ &\quad \times E_{F_n}(Y_{n,0}^*/\phi_{n,1}^2)/E_{F_n}(\phi_{n,1}^{-2}) \\ &\quad + (E_{F_n}(Y_{n,0}^*/\phi_{n,1}^2))^2 E_{F_n}(U_{n,1}^2/\phi_{n,1}^4)/(E_{F_n}(\phi_{n,1}^{-2}))^2 \\ &= E_{F_n}(Y_{n,0}^{*2}U_{n,1}^2/\phi_{n,1}^4)(1 + O(1-\rho_n)) \end{aligned} \tag{12}$$

and

$$\begin{aligned} &E_{F_n}(Y_{n,0}^{*2}/\phi_{n,1}^2) - (E_{F_n}(Y_{n,0}^*/\phi_{n,1}^2))^2/E_{F_n}(\phi_{n,1}^{-2}) \\ &= E_{F_n}(Y_{n,0}^{*2}/\phi_{n,1}^2)(1 + O(1-\rho_n)). \end{aligned} \tag{13}$$

If, in addition, $\{U_{n,i} : i = \ldots, 0, 1, \ldots\}$ are i.i.d. with mean 0, variance $\sigma_{U,n}^2 \in (0,\infty)$, and distribution F_n and $\phi_{n,i}^2 = 1$, then the constants a_n and d_n simplify to

$$a_n = n^{1/2}(1-\rho_n^2)^{-1/2} \quad \text{and} \quad d_n = (1-\rho_n^2)^{-1/2}. \tag{14}$$

This follows because in the present case $\phi_{n,i}^2 = 1$, $E_{F_n} Y_{n,0}^{*2} = \sum_{j=0}^{\infty} \rho_n^{2j} E_{F_n} U_{n,-j}^2 = (1-\rho_n^2)^{-1}\sigma_{U,n}^2$, and $E_{F_n}(Y_{n,0}^{*2}U_{n,1}^2/\phi_{n,1}^2) = (1-\rho_n^2)^{-1}\sigma_{n,U}^4$. The expression for a_n in (14) is as in Giraitis and Phillips (2006).

The form of d_n in (11) is explained as follows. For the infeasible QGLS estimator, one can write $n^{1/2}(\widehat{\rho}_n - \rho_n) = (n^{-1}X_1'M_{X_2}X_1)^{-1}n^{-1/2}X_1'M_{X_2}U$ as in (4) with X_1, X_2, and U defined with $\phi_{n,i}$ in place of $\widehat{\phi}_{n,i}$. The numerator of d_n in (11) is the rate of growth of $n^{-1}X_1'M_{X_2}X_1$, see (37) and (40), and the denominator of d_n in (11) is the rate of growth of $n^{-1/2}X_1'M_{X_2}U$, see (37)–(39).

3.3. Results for LS and infeasible QGLS

In this section, we provide results for the infeasible QGLS estimator which is based on $\{\phi_{n,i}^2 : i \leq n\}$ rather than $\{\widehat{\phi}_{n,i}^2 : i \leq n\}$ (i.e., the estimator $\widehat{\rho}_n$ in (4) with $\phi_{n,i}$ in place of $\widehat{\phi}_{n,i}$). Conditions under which feasible and infeasible QGLS estimators are asymptotically equivalent are given in Section 3.4. The LS estimator is covered by the results of this section by taking $\phi_{n,i}^2 = 1$ for all n, i (i.e., the estimator $\widehat{\rho}_n$ in (4) with $\widehat{\phi}_{n,i} = 1$ for all n, i).

Let $W(\cdot)$ and $W_2(\cdot)$ be independent standard Brownian motions on $[0,1]$. Let Z_1 be a standard normal random variable that is independent of $W(\cdot)$ and $W_2(\cdot)$. We define

$$I_h(r) = \int_0^r \exp(-(r-s)h_1)dW(s),$$
$$I_h^*(r) = I_h(r) + \frac{1}{\sqrt{2h_1}}\exp(-h_1 r)Z_1 \quad \text{for } h_1 > 0 \quad \text{and}$$
$$I_h^*(r) = W(r) \quad \text{for } h_1 = 0,$$
$$I_{D,h}^*(r) = I_h^*(r) - \int_0^1 I_h^*(s)ds, \quad \text{and}$$
$$Z_2 = \left(\int_0^1 I_{D,h}^*(r)^2 dr\right)^{-1/2} \int_0^1 I_{D,h}^*(r)dW_2(r). \tag{15}$$

As defined, $I_h(r)$ is an Ornstein–Uhlenbeck process. Note that the conditional distribution of Z_2 given $W(\cdot)$ and Z_1 is standard normal. Hence, its unconditional distribution is standard normal and it is independent of $W(\cdot)$ and Z_1.

The asymptotic distribution of the infeasible QGLS estimator and t statistic are given in the following theorem.

Theorem 1. *Suppose that* (i) *Assumption INNOV holds,* (ii) *Assumption STAT holds when $\rho_n < 1$,* (iii) *$\rho_n \in [-1+\varepsilon, 1]$ for some $0 < \varepsilon < 2$, and* (iv) *$\rho_n = 1 - h_{n,1}/n$ and $h_{n,1} \to h_1 \in [0,\infty]$. Then, the infeasible QGLS estimator $\widehat{\rho}_n$ and t statistic $T_n^*(\rho_n)$ (defined in* (3) *and* (4) *with $\phi_{n,i}$ in place of $\widehat{\phi}_{n,i}$) satisfy*

$$a_n(\widehat{\rho}_n - \rho_n) \to_d V_h, \; d_n\widehat{\sigma}_n \to_d Q_h, \quad \text{and}$$
$$T_n^*(\rho_n) = \frac{n^{1/2}(\widehat{\rho}_n - \rho_n)}{\widehat{\sigma}_n} \to_d J_h,$$

where a_n, d_n, V_h, Q_h, and J_h are defined as follows.[5]

[5] For simplicity, in Theorem 1 and Theorem 2 below, for a sequence of random variables $\{W_n : n \geq 1\}$ and a distribution V, we write $W_n \to_d V$ as $n \to \infty$, rather than $W_n \to_d W$ as $n \to \infty$ for a random variable W with distribution V.

(a) *For* $h_1 \in [0,\infty)$, $a_n = n$, $d_n = n^{1/2}$, V_h *is the distribution of*

$$h_{2,1}^{-1/2}h_{2,2}^{1/2}h_{2,5}^{-1}\left(h_{2,7}\frac{\int_0^1 I_{D,h}^*(r)dW(r)}{\int_0^1 I_{D,h}^*(r)^2dr} + (1-h_{2,7}^2)^{1/2}\frac{\int_0^1 I_{D,h}^*(r)dW_2(r)}{\int_0^1 I_{D,h}^*(r)^2dr}\right) \quad (16)$$

Q_h *is the distribution of*

$$h_{2,1}^{-1/2}h_{2,2}^{1/2}h_{2,5}^{-1}\left[\int_0^1 I_{D,h}^*(r)^2dr\right]^{-1/2}, \quad (17)$$

and J_h *is the distribution of*

$$h_{2,7}\frac{\int_0^1 I_{D,h}^*(r)dW(r)}{\left(\int_0^1 I_{D,h}^*(r)^2dr\right)^{1/2}} + (1-h_{2,7}^2)^{1/2}Z_2. \quad (18)$$

(b) *For* $h_1 = \infty$, a_n *and* d_n *are defined as in (10),* V_h *is the* $N(0,1)$ *distribution,* Q_h *is the distribution of the constant one, and* J_h *is the* $N(0,1)$ *distribution.*

Comments.

1. Theorem 1 shows that the asymptotic distribution of the QGLS t statistic is a standard normal distribution when $n(1-\rho_n) \to \infty$ and a mixture of a standard normal distribution and a "demeaned near unit-root distribution" when $n(1-\rho_n) \to h_1 \in [0,\infty)$. In the latter case, the mixture depends on $h_{2,7}$, which is the asymptotic correlation between the innovation $U_{n,i}$ and the rescaled innovation $U_{n,i}/\phi_{n,i}^2$. When the LS estimator is considered (which corresponds to $\phi_{n,i}^2 = 1$), we have $h_{2,7} = 1$ and the asymptotic distribution is a "demeaned near unit-root distribution."
2. It is important to note that the t statistic considered in Theorem 1 employs a heteroskedasticity-robust standard error estimator $\widehat{\sigma}_n$, see its definition in (4). This differs from other papers in the literature, such as Stock (1991), Hansen (1999), Giraitis and Phillips (2006), Mikusheva (2007), and Phillips and Magdalinos (2007), which consider the LS estimator and the usual LS standard error estimator that is designed for homoskedasticity. In consequence, the results of Theorem 1 with $\phi_{n,i} = 1$ (which corresponds to the LS estimator of ρ_n) do not imply that the t statistics considered in the latter papers have a standard normal distribution when $n(1-\rho_n) \to \infty$ in the presence of conditional heteroskedasticity. The standard error estimator designed for homoskedasticity is not consistent under conditional heteroskedasticity.
3. The asymptotic results of Theorem 1 apply to a first-order AR model. They should extend without essential change to a p-th order autoregressive model in which ρ equals the "sum of the AR coefficients". Of course, the proofs will be more complex. We do not provide them here.
4. Theorem 1 is used in the AR(1) example of Andrews and Guggenberger (2009) to verify their Assumptions BB(i) and (iii) for the (infeasible) QGLS estimator (with Q_h playing the role of W_h in Assumption BB). In turn, the results of Andrews and Guggenberger (2009) show that whether or not conditional heteroskedasticity is present: (i) the symmetric two-sided subsampling confidence interval for ρ has correct asymptotic size (defined to be the limit as $n \to \infty$ of exact size) and (ii) upper and lower one-sided and symmetric and equal-tailed two-sided hybrid-subsampling confidence intervals for ρ have correct asymptotic size. These results hold even if the form of the conditional heteroskedasticity is misspecified.

3.4. Asymptotic equivalence of feasible and infeasible QGLS

Here we provide sufficient conditions for the feasible and infeasible QGLS statistics to be asymptotically equivalent. In particular, we give conditions under which Theorem 1 holds when $\widehat{\rho}_n$ is defined using the feasible conditional heteroskedasticity estimators $\{\widehat{\phi}_{n,i} : i \le n\}$.

We assume that the conditional heteroskedasticity estimators (CHE) $\{\widehat{\phi}_{n,i}^2 : i \le n\}$ satisfy the following assumption.

Assumption CHE. (i) For some $\varepsilon > 0$, $\widehat{\phi}_{n,i}^2 \ge \varepsilon$ a.s. for all $i \le n$, $n \ge 1$.
(ii) For random variables $\{(U_{n,i}, \phi_{n,i}^2) : i = \dots, 0, 1, \dots\}$ for $n \ge 1$ that satisfy Assumption INNOV and for $Y_{n,i} = \alpha + Y_{n,i}^*$, $Y_{n,i}^* = \rho_n Y_{n,i-1}^* + U_{n,i}$, with $\alpha = 0$, that satisfies Assumption STAT when $\rho_n < 1$ and $n(1-\rho_n) \to h_1 \in [0,\infty]$, we have
(a) when $h_1 \in [0,\infty)$, $n^{-1/2}\sum_{i=1}^n (n^{-1/2}Y_{n,i-1}^*)^j U_{n,i}(\widehat{\phi}_{n,i}^{-2} - \phi_{n,i}^{-2}) = o_p(1)$ for $j = 0, 1$,
(b) when $h_1 \in [0,\infty)$, $n^{-1}\sum_{i=1}^n |U_{n,i}|^d |\widehat{\phi}_{n,i}^{-j} - \phi_{n,i}^{-j}| = o_p(1)$ for $(d,j) = (0,1), (1,2)$, and $(2,2)$,
(c) when $h_1 = \infty$, $n^{-1/2}\sum_{i=1}^n ((1-\rho_n)^{1/2}Y_{n,i-1}^*)^j U_{n,i}(\widehat{\phi}_{n,i}^{-2} - \phi_{n,i}^{-2}) = o_p(1)$ for $j = 0, 1$, and
(d) when $h_1 = \infty$, $n^{-1}\sum_{i=1}^n |U_{n,i}|^k |\widehat{\phi}_{n,i}^{-j} - \phi_{n,i}^{-j}|^d = o_p(1)$ for $(d,j,k) = (1,2,0), (2,2,0)$, and $(2,4,k)$ for $k = 0, 2, 4$.

Assumption CHE(i) is not restrictive. For example, if $\widehat{\phi}_{n,i}$ is obtained by specifying a parametric model for the conditional heteroskedasticity, then Assumption CHE(i) holds provided the specified parametric model (which is user chosen) consists of an intercept that is bounded away from zero plus a non-negative random component (as in (19)). Most parametric models in the literature have this form and it is always possible to use one that does. Assumption CHE(ii) specifies the sense in which $\widehat{\phi}_{n,i}$ must converge to $\phi_{n,i}$ for $i \le n$, $n \ge 1$ in order for the feasible and infeasible QGLS estimators to be asymptotically equivalent. Typically, Assumptions CHE(ii)(a) and (c) are more difficult to verify than Assumptions CHE(ii)(b) and (d) because they have the scale factor $n^{-1/2}$ rather than n^{-1}.

Theorem 2. *Suppose* (i) *Assumptions CHE and INNOV hold,* (ii) *Assumption STAT holds when* $\rho_n < 1$, (iii) $\rho_n \in [-1+\varepsilon, 1]$ *for some* $0 < \varepsilon < 2$, *and* (iv) $\rho_n = 1 - h_{n,1}/n$ *and* $h_{n,1} \to h_1 \in [0,\infty]$. *Then, the feasible QGLS estimator* $\widehat{\rho}_n$ *and t statistic* $T_n^*(\rho_n)$ *(defined in (3) and (4) using* $\widehat{\phi}_{n,i}$*) satisfy*

$$a_n(\widehat{\rho}_n - \rho_n) \to_d V_h, \ d_n\widehat{\sigma}_n \to_d Q_h, \quad \text{and}$$

$$T_n^*(\rho_n) = \frac{n^{1/2}(\widehat{\rho}_n - \rho_n)}{\widehat{\sigma}_n} \to_d J_h,$$

where a_n, d_n, V_h, Q_h, *and* J_h *are defined as in Theorem 1 (that is, with* a_n *and* d_n *defined using* $\phi_{n,i}$, *not* $\widehat{\phi}_{n,i}$*).*

Comment. Theorem 2 shows that the infeasible and feasible QGLS statistics have the same asymptotic distributions under Assumption CHE.

We now provide sufficient conditions for Assumption CHE. Suppose $\{\widehat{\phi}_{n,i}^2 : i \le n\}$ are based on a parametric model with conditional heteroskedasticity parameter π estimated using residuals. Let $\widetilde{\pi}_n$ be the estimator of π and let $(\widetilde{\alpha}_n, \widetilde{\rho}_n)$ be the estimators of $(\widetilde{\alpha}, \rho)$ used to construct the residuals, where $\widetilde{\alpha}$ is the intercept when the model is written in regression form, see (2). For example, $\widetilde{\pi}_n$ may be an estimator of π based on residuals in place of the true errors and $(\widetilde{\alpha}_n, \widetilde{\rho}_n)$ may be the LS estimators (whose properties are covered by the asymptotic results given in Theorem 1 by taking $\phi_{n,i} = 1$). In particular, suppose that

$$\widehat{\phi}^2_{n,i} = \phi^2_{n,i}(\widetilde{\alpha}_n, \widetilde{\rho}_n, \widetilde{\pi}_n), \quad \text{where}$$
$$\phi^2_{n,i}(\widetilde{\alpha}, \rho, \pi) = \omega + \sum_{j=1}^{L_i} \mu_j(\pi)\widehat{U}^2_{n,i-j}(\widetilde{\alpha}, \rho),$$
$$\widehat{U}_{n,i}(\widetilde{\alpha}, \rho) = Y_{n,i} - \widetilde{\alpha} - \rho Y_{n,i-1}, \tag{19}$$

$L_i = \min\{i-1, L\}$, and ω is an element of π. Here $L < \infty$ is a bound on the maximum number of lags allowed. Any model with stationary conditional heteroskedasticity (bounded away from the nonstationary region), such as a GARCH(1, 1) model, can be approximated arbitrarily well by taking L sufficiently large. Hence, the restriction to finite lags is not overly restrictive. The upper bound L_i, rather than L, on the number of lags in the sum in (19) takes into account the truncation at 1 that naturally occurs because one does not observe residuals for $i < 1$.

The parameter space for π is Π, which is a bounded subset of R^{d_π}, for some $d_\pi > 0$. Let $\widehat{\pi}_n \in \Pi$ be an n^{δ_1}-consistent estimator of π for some $\delta_1 > 0$. For technical reasons, we base $\widehat{\phi}^2_{n,i}$ on an estimator $\widetilde{\pi}_n$ that is a discretized version of $\widehat{\pi}_n$ that takes values in a finite set Π_n ($\subset \Pi$) for $n \geq 1$, where Π_n consists of points on a uniform grid with grid size that goes to zero as $n \to \infty$ and hence the number of elements of Π_n diverges to infinity as $n \to \infty$. The reason for considering a discretized estimator is that when the grid size goes to zero more slowly than $n^{-\delta_1}$, then wp $\to$ 1 the estimators $\{\widetilde{\pi}_n : n \geq 1\}$ take values in a sequence of finite sets $\{\Pi_{n,0} : n \geq 1\}$ whose numbers of elements is bounded as $n \to \infty$. The latter property makes it easier to verify Assumption CHE(ii). The set Π_n can be defined such that there is very little difference between $\widehat{\pi}_n$ and $\widetilde{\pi}_n$ in a finite sample of size n.

We employ the following sufficient condition for the FQGLS estimator to be asymptotically equivalent to the (infeasible) QGLS estimator.

Assumption CHE2. (i) $\widehat{\phi}^2_{n,i}$ satisfies (19) with $L < \infty$ and $\mu_j(\cdot) \geq 0$ for all $j = 1, \ldots, L$,
(ii) $\phi^2_{n,i} = \omega_n + \sum_{j=1}^{L} \mu_j(\pi_n) U^2_{n,i-j}$ and $\pi_n \to \pi_0$ for some $\pi_0 \in \Pi$ (and π_0 may depend on the sequence), where ω_n is an element of π_n,
(iii) $a_n(\widetilde{\rho}_n - \rho_n) = O_p(1)$, $n^{1/2}\widetilde{\alpha}_n = O_p(1)$, and $n^{\delta_1}(\widehat{\pi}_n - \pi_n) = o_p(1)$ for some $\delta_1 > 0$ under any sequence $(U_{n,i}, \phi^2_{n,i})$ that satisfies Assumption INNOV and for $Y_{n,i}$ defined as in Assumption CHE with $\alpha = \beta = 0$ satisfying Assumption STAT when $\rho_n < 1$, and with $\rho = \rho_n$ that satisfies $n(1-\rho_n) \to h_1 \in [0, \infty]$, where a_n is defined in (10),
(iv) $\widetilde{\pi}_n$ minimizes $\|\pi - \widehat{\pi}_n\|$ over $\pi \in \Pi_n$ for $n \geq 1$, where Π_n ($\subset \Pi$) consists of points on a uniform grid with grid size $Cn^{-\delta_2}$ for some $0 < \delta_2 < \delta_1$ and $0 < C < \infty$,
(v) Π bounds the intercept ω away from zero, and
(vi) $\mu_j(\pi)$ is continuous on Π for $j = 1, \ldots, L$.

The part of Assumption CHE2(iii) concerning $\widetilde{\rho}_n$ holds for the LS estimator by Theorem 1(a) (by taking $\phi_{n,i} = 1$), the part concerning $\widetilde{\alpha}_n$ holds for the LS estimator by similar, but simpler, arguments, and typically the part concerning $\widehat{\pi}_n$ holds for all $\delta_1 < 1/2$. Assumptions CHE2(iv)–(vi) can always be made to hold by choice of $\widehat{\pi}_n$, Π, and $\mu_j(\pi)$.

Lemma 1. *Assumption CHE2 implies Assumption CHE.*

Comment. The use of a discretized estimator $\widetilde{\pi}_n$ and a finite bound L on the number of lags in Assumption CHE2 are made for technical convenience. Undoubtedly, they are not necessary for the lemma to hold (although other conditions may be needed in their place).

4. Proofs

This section verifies parts of Assumption INNOV for a GARCH(1, 1) model and provides proofs of Theorems 1 and 2, and Lemma 1. Section 4.1 is concerned with verification of parts of Assumption INNOV for a GARCH(1, 1) model. Section 4.2.1 states Lemmas 2–9, which are used in the proof of Theorem 1. Section 4.2.2 proves Theorem 1. For brevity, the proofs of Lemmas 2–9 are left out. They are given in Andrews and Guggenberger (2010b). Section 4.3 proves Theorem 2. Section 4.4 proves Lemma 1.

4.1. Verification of Assumption INNOV for GARCH(1, 1)

To verify the stationarity part of Assumption INNOV(i) for the model in (5), we use Lindner (2009, Theorem 1(a)) for the case $\beta_n > 0$ and Lindner (2009, Theorem 1(b)(i)–(ii)) for the case $\beta_n = 0$. These results imply that $\{(U_{n,i}, \sigma^2_{n,i}) : i = \ldots, 0, 1, \ldots\}$ are strictly stationary if for all $n \geq 1$ we have $c_n > 0$, $\alpha_n > 0$, $\beta_n \geq 0$, $E_{F_n} \log(\beta_n + \alpha_n\varepsilon^2_{n,1}) > -\infty$, and $E_{F_n} \log(\beta_n + \alpha_n\varepsilon^2_{n,1}) < 0$. When $\beta_n = 0$, the fourth and fifth conditions can be replaced by $P(\varepsilon_{n,1} = 0) > 0$. The first three restrictions hold by assumption. The fourth requirement clearly holds when $\beta_n > 0$. When $\beta_n = 0$ and $P(\varepsilon_{n,1} = 0) = 0$, it also follows that $E_{F_n} \log(\alpha_n\varepsilon^2_{n,1}) > -\infty$. By Jensen's inequality, a sufficient condition for the fifth requirement is that $\alpha_n E_{F_n}\varepsilon^2_{n,1} + \beta_n = \alpha_n + \beta_n < 1$, which is assumed.

To verify Assumption INNOV(iii), we use Bollerslev (1986, Theorem 2) and Lindner (2009, Theorem 5). First, for all n, $E_{F_n}U^2_{n,i} = E_{F_n}\sigma^2_{n,i} \geq \inf_{n\geq 1} c_n > 0$. Next, it is enough to establish that $\sup_{n\geq 1} E_{F_n}(\sigma^2_{n,i})^{\zeta} < \infty$ and

$$\sup_{n\geq 1} E_{F_n}|U_{n,i-s}U_{n,i-t}(U^2_{n,i+1}/\phi^4_{n,i+1}) \times U_{n,-u}U_{n,-v}((U^2_{n,1} + \sigma^2_{n,1})/\phi^4_{n,1})|^{\zeta} < \infty. \tag{20}$$

For notational simplicity, we now often leave out the subscript n on random variables and F_n on expectations. We first establish that $\sup_{n\geq 1} E|\varepsilon_1|^{6\zeta} < \infty$ and $\sup_{n\geq 1} E|\sigma_1|^{6\zeta} < \infty$ imply $\sup_{n\geq 1} E|U_{i-s}U_{i-t}(U^2_{i+1}/\phi^4_{i+1})U_{-u}U_{-v}((U^2_1 + \sigma^2_1)/\phi^4_1)|^{\zeta} < \infty$. We then specify conditions on (c_n, α_n, β_n) that imply $\sup_{n\geq 1} E|\sigma_1|^{6\zeta} < \infty$.

To deal with the first task, we consider only the case where $i - t < 1 < i - s$. All other cases can be handled analogously (or more easily). Note that because $s \geq 0$ it follows that $i - s < i + 1$. Therefore, using the law of iterated expectations (LIE),

$$\begin{aligned} &E|U_{-u}U_{-v}U_{i-t}((U^2_1 + \sigma^2_1)/\phi^4_1)U_{i-s}(U^2_{i+1}/\phi^4_{i+1})|^{\zeta} \\ &= EE(|U_{-u}U_{-v}U_{i-t}((U^2_1 + \sigma^2_1)/\sigma^4_1)U_{i-s}(U^2_{i+1}/\sigma^4_{i+1})|^{\zeta}|\mathcal{G}_i) \\ &= E[(|U_{-u}U_{-v}U_{i-t}((U^2_1 + \sigma^2_1)/\sigma^4_1)U_{i-s}\sigma^{-2}_{i+1}|^{\zeta})E(|\varepsilon_{i+1}|^{2\zeta}\ |\mathcal{G}_i)]. \end{aligned} \tag{21}$$

Because $(\sigma^{-2}_{i+1})^{\zeta}$ and $(\sigma^{-2}_1)^{\zeta}$ are uniformly bounded by Assumption INNOV(iv) and $E(|\varepsilon_{i+1}|^{2\zeta}\ |\mathcal{G}_i) = E|\varepsilon_{i+1}|^{2\zeta}$ is uniformly bounded it is enough to show that $E|U_{-u}U_{-v}U_{i-t}(\varepsilon^2_1 + 1)U_{i-s}|^{\zeta}$ is uniformly bounded. Again, by the LIE and $i - s > 1$, we have

$$\begin{aligned} &E|U_{-u}U_{-v}U_{i-t}(\varepsilon^2_1 + 1)U_{i-s}|^{\zeta} \\ &= EE(|U_{-u}U_{-v}U_{i-t}(\varepsilon^2_1 + 1)U_{i-s}|^{\zeta}|\mathcal{G}_{i-s-1}) \\ &= E|U_{-u}U_{-v}U_{i-t}(\varepsilon^2_1 + 1)\sigma_{i-s}|^{\zeta}\ E|\varepsilon_{i-s}|^{\zeta}\ . \end{aligned} \tag{22}$$

By Hölder's inequality $E|U_{-u}U_{-v}U_{i-t}(\varepsilon^2_1+1)\sigma_{i-s}|^{\zeta} \leq (E|U^2_{-u}U^2_{-v}U^2_{i-t} \times (\varepsilon^2_1 + 1)^2|^{\zeta}E|\sigma^2_{i-s}|^{\zeta})^{1/2}$. By the generalized H ölder inequality we finally obtain

$$\begin{aligned} &E|U^2_{-u}U^2_{-v}U^2_{i-t}(\varepsilon^2_1 + 1)^2|^{\zeta} = E|U^2_{-u}U^2_{-v}U^2_{i-t}|^{\zeta}E(\varepsilon^2_1 + 1)^{2\zeta} \\ &\leq (E|U_{-u}|^{6\zeta}E|U_{-v}|^{6\zeta}E|U_{i-t}|^{6\zeta})^{1/3}E(\varepsilon^2_1 + 1)^{2\zeta} \\ &= E|U_1|^{6\zeta}E(\varepsilon^2_1 + 1)^{2\zeta}, \end{aligned} \tag{23}$$

where in the last line we used stationarity. Now, $E|U_1|^{6\zeta} = E|\varepsilon_1|^{6\zeta}E|\sigma_1|^{6\zeta}$ which is bounded by assumption. Also, $\sup_{n\geq 1} E(\varepsilon_1^2+1)^{2\zeta} < \infty$ because $\sup_{n\geq 1} E \mid \varepsilon_1 \mid^{6\zeta} < \infty$ by assumption. This proves the first claim.

Next, we specify conditions on (c_n, α_n, β_n) that imply $\sup_{n\geq 1} E_{F_n} |\sigma_{n,1}|^{6\zeta} < \infty$. By Lindner (2009, Eq. (10)) we have $\sigma_{n,t}^2 = \sum_{i=0}^{\infty}\prod_{j=0}^{i-1} c_n(\beta_n + \alpha_n\varepsilon_{n,t-1-j}^2)$. Therefore, using Minkowski's inequality and $\{\varepsilon_{n,i}\}$ i.i.d. we have

$$(E_{F_n}|\sigma_{n,1}^2|^{3\zeta})^{1/(3\zeta)} \leq c_n \sum_{i=0}^{\infty}(E_{F_n}(\beta_n+\alpha_n\varepsilon_{n,1}^2)^{3\zeta})^{i/(3\zeta)}, \tag{24}$$

see Lindner (2009, first equation p. 57). Therefore $\sup_{n\geq 1} E_{F_n}|\sigma_{n,1}|^{6\zeta} < \infty$ if $\sup_{n\geq 1} c_n < \infty$ and $\sup_{n\geq 1} E_{F_n}(\beta_n+\alpha_n\varepsilon_{n,1}^2)^{3\zeta} < 1$.

For the case where $\varepsilon_{n,1}$ is $N(0, 1)$ we simulate $E_{F_n}(\beta_n+\alpha_n\varepsilon_{n,1}^2)^{3\zeta}$ for a grid with stepsize 0.01 of parameter combinations for (α_n, β_n) for which $\alpha_n, \beta_n \geq 0$ and $\alpha_n + \beta_n < 1$ using 2000,000 draws from $\varepsilon_{n,1}$ and $\zeta = 3+\varepsilon$ with $\varepsilon = 1/30$. The expectation is smaller than 1 for the parameter combinations (α_n, β_n) reported in the footnote below (5).

Assumption INNOV(iv) is clearly satisfied if $\inf_{n\geq 1} c_n > 0$. □

4.2. Proof of Theorem 1

To simplify notation, in the remainder of the paper we omit the subscript F_n on expectations.

4.2.1. Lemmas 2–9

The proof of Theorem 1 uses eight lemmas that we state in this section. The first four lemmas deal with the case of $h_1 \in [0, \infty)$. The last four deal with the case of $h_1 = \infty$.

In integral expressions below, we often leave out the lower and upper limits zero and one, the argument r, and dr to simplify notation when there is no danger of confusion. For example, $\int_0^1 I_h(r)^2 dr$ is typically written as $\int I_h^2$. By "⇒" we denote weak convergence of a stochastic process as $n \to \infty$.

Lemma 2. *Suppose Assumptions INNOV and STAT hold, $\rho_n \in (-1, 1)$ and $\rho_n = 1-h_{n,1}/n$ where $h_{n,1} \to h_1 \in [0, \infty)$ as $n \to \infty$. Then,*

$$(2h_{n,1}/n)^{1/2}Y_{n,0}^*/\mathrm{StdDev}_{F_n}(U_{n,0}) \to_d Z_1 \sim N(0, 1).$$

Define $h_{n,1}^* \geq 0$ by $\rho_n = \exp(-h_{n,1}^*/n)$. As shown in the proof of Lemma 2, $h_{n,1}^*/h_{n,1} \to 1$ when $h_1 \in [0, \infty)$. By recursive substitution, we have

$$Y_{n,i}^* = \widetilde{Y}_{n,i} + \exp(-h_{n,1}^* i/n)Y_{n,0}^*, \quad \text{where}$$

$$\widetilde{Y}_{n,i} = \sum_{j=1}^{i}\exp(-h_{n,1}^*(i-j)/n)U_{n,j}. \tag{25}$$

Let $BM(\Omega)$ denote a bivariate Brownian motion on [0, 1] with variance matrix Ω. The next lemma is used to establish the simplified form of the asymptotic distribution that appears in Theorem 1(a).

Lemma 3. *Suppose $(h_{2,1}^{1/2}W(r), M(r))' = BM(\Omega)$, where*

$$\Omega = \begin{bmatrix} h_{2,1} & h_{2,3} \\ h_{2,3} & h_{2,2} \end{bmatrix}.$$

Then, $M(r)$ can be written as $M(r) = h_{2,2}^{1/2}\big(h_{2,7}W(r) + (1-h_{2,7}^2)^{1/2} W_2(r)\big)$, where $(W(r), W_2(r))' = BM(I_2)$ and $h_{2,7} = h_{2,3}/(h_{2,1} h_{2,2})^{1/2}$ is the correlation that arises in the variance matrix Ω.

The following lemma states some general results on weak convergence of certain statistics to stochastic integrals. It is proved using Theorems 4.2 and 4.4 of Hansen (1992) and Lemma 2 above. Let ⊗ denote the Kronecker product.

Lemma 4. *Suppose $\{v_{n,i} : i \leq n, n \geq 1\}$ is a triangular array of row-wise strictly-stationary strong-mixing random d_v-vectors with (i) strong-mixing numbers $\{\alpha_n(m) : m \geq 1, n \geq 1\}$ that satisfy $\alpha(m) = \sup_{n\geq 1}\alpha_n(m) = O(m^{-\zeta\tau/(\zeta-\tau)})$ as $m \to \infty$ for some $\zeta > \tau > 2$, and (ii) $\sup_{n\geq 1}\|v_{n,i}\|_\zeta < \infty$, where $\|\cdot\|_\zeta$ denotes the L^ζ-norm. Suppose $n^{-1}EV_nV_n' \to \Omega_0$ as $n \to \infty$, where $V_n = \sum_{i=1}^n v_{n,i}$, and Ω_0 is some $d_v \times d_v$ variance matrix. Let $X_{n,i} = \rho_n X_{n,i-1} + v_{n,i}$, where $n(1-\rho_n) \to h_1 \in [0, \infty)$. If $h_1 > 0$, the first element of $X_{n,i}$ has a stationary initial condition and all of the other elements have zero initial conditions. If $h_1 = 0$, all of the elements of $X_{n,i}$ have zero initial conditions, i.e., $X_{n,0} = 0$. Let $\Lambda = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n\sum_{j=i+1}^n Ev_{n,i}v_{n,j}'$. Let $K_h(r) = \int_0^r \exp((r-s)h_1)dB(s)$, where $B(\cdot)$ is a d_v-vector $BM(\Omega_0)$ on [0, 1]. If $h_1 > 0$, let $K_h^*(r) = K_h(r) + e_1(2h_1)^{-1/2}\exp(-h_1 r)\Omega_{0,1,1}^{1/2}Z_1$, where $Z_1 \sim N(0, 1)$ is independent of $B(\cdot)$, $e_1 = (1, 0, \ldots, 0)' \in R^{d_v}$, and $\Omega_{0,1,1}$ denotes the (1, 1) element of Ω_0. If $h_1 = 0$, let $K_h^*(r) = K_h(r)$. Then, the following results hold jointly,*

(a) $n^{-1/2}X_{n,[nr]} \Rightarrow K_h^*(r)$,
(b) $n^{-1}\sum_{i=1}^n X_{n,i-1}v_{n,i}' \to_d \int K_h^* dB' + \Lambda$,
(c) *for* $\tau \geq 3$, $n^{-3/2}\sum_{i=1}^n (X_{n,i-1}\otimes X_{n,i-1})v_{n,i}' \to_d \int (K_h^*\otimes K_h^*)dB' + \big(\Lambda\otimes\int K_h^*\big) + \big(\int K_h^*\otimes\Lambda\big)$, *and*
(d) $n^{-1/2}\sum_{i=1}^n v_{n,i}' \to_d \int dB'$.

We now use Lemma 4 to establish the following results which are key in the proof of Theorem 1(a). Let $[a]$ denote the integer part of a.

Lemma 5. *Suppose Assumptions INNOV and STAT hold, $\rho_n \in (-1, 1]$, $\rho_n = 1-h_{n,1}/n$ where $h_{n,1} \to h_1 \in (0, \infty)$. Then, the following results (a)–(k) hold jointly,*

(a) $n^{-1/2}Y_{n,[nr]}^* \Rightarrow h_{2,1}^{1/2}I_h^*(r)$,
(b) $n^{-1}\sum_{i=1}^n \phi_{n,i}^{-j} \to_p \lim_{n\to\infty} E\phi_{n,i}^{-j} = h_{2,(j+3)}$ *for* $j = 1, 2, 4$,
(c) $n^{-1}\sum_{i=1}^n U_{n,i}/\phi_{n,i}^4 \to_p \lim_{n\to\infty} E(U_{n,i}/\phi_{n,i}^4) = 0$,
(d) $n^{-1}\sum_{i=1}^n U_{n,i}^2/\phi_{n,i}^4 \to_p \lim_{n\to\infty} E(U_{n,i}^2/\phi_{n,i}^4) = h_{2,2}$,
(e) $n^{-1/2}\sum_{i=1}^n U_{n,i}/\phi_{n,i}^2 \to_d M(1) = \int dM = h_{2,2}^{1/2}\int d[h_{2,7}W(r) + (1-h_{2,7}^2)^{1/2}W_2(r)]$,
(f) $n^{-3/2}\sum_{i=1}^n Y_{n,i-1}^*/\phi_{n,i}^2 = n^{-3/2}\sum_{i=1}^n Y_{n,i-1}^* E\phi_{n,1}^{-2} + O_p(n^{-1/2}) \to_d h_{2,5}h_{2,1}^{1/2}\int I_h^*$,
(g) $n^{-1}\sum_{i=1}^n Y_{n,i-1}^*U_{n,i}/\phi_{n,i}^2 \to_d h_{2,1}^{1/2}\int I_h^* dM = h_{2,2}^{1/2}h_{2,1}^{1/2}\int I_h^* d[h_{2,7} W(r) + (1-h_{2,7}^2)^{1/2}W_2(r)]$,
(h) $n^{-2}\sum_{i=1}^n Y_{n,i-1}^{*2}/\phi_{n,i}^2 = n^{-2}\sum_{i=1}^n Y_{n,i-1}^{*2}E\phi_{n,1}^{-2} + O_p(n^{-1/2}) \to_d h_{2,5}h_{2,1}\int I_h^{*2}$,
(i) $n^{-3/2}\sum_{i=1}^n Y_{n,i-1}^*U_{n,i}^2/\phi_{n,i}^4 = n^{-3/2}\sum_{i=1}^n Y_{n,i-1}^*E(U_{n,1}^2/\phi_{n,1}^4) + O_p(n^{-1/2}) \to_d h_{2,2}h_{2,1}^{1/2}\int I_h^*$,
(j) $n^{-2}\sum_{i=1}^n Y_{n,i-1}^{*2}U_{n,i}^2/\phi_{n,i}^4 = n^{-2}\sum_{i=1}^n Y_{n,i-1}^{*2}E(U_{n,1}^2/\phi_{n,1}^4) + O_p(n^{-1/2}) \to_d h_{2,2}h_{2,1}\int I_h^{*2}$,
(k) $n^{-1-\ell_1/2}\sum_{i=1}^n Y_{n,i-1}^{*\ell_1}U_{n,i}^{\ell_2}/\phi_{n,i}^4 = o_p(n)$ *for* $(\ell_1, \ell_2) = (1, 0), (1, 1), (2, 0), (2, 1), (3, 0), (3, 1)$, *and* $(4, 0)$, *and*
(l) *when* $h_1 = 0$, *parts* (a) *and* (f)–(k) *hold with* $Y_{n,i-1}^*$ *replaced by* $\widetilde{Y}_{n,i-1}$.

In the proof of Theorem 1(b), we use the following well-known strong-mixing covariance inequality, see e.g. Doukhan (1994, Theorem 3, p. 9). Let X and Y be strong-mixing random variables with respect to σ-fields $\mathcal{F}_i^j$ (for integers $i \leq j$) such that $X \in \mathcal{F}_{-\infty}^n$ and $Y \in \mathcal{F}_{n+m}^\infty$ with strong-mixing numbers $\{\alpha(m) : m \geq 1\}$. For $p, q > 0$ such that $1 - p^{-1} - q^{-1} > 0$, let $\|X\|_p = (E|X|^p)^{1/p}$ and $\|Y\|_q = (E|Y|^q)^{1/q}$. Then, the following inequality holds

$$\mathrm{Cov}(X, Y) \leq 8\|X\|_p\|Y\|_q\alpha(k)^{1-p^{-1}-q^{-1}}. \tag{26}$$

The proof of Theorem 1(b) uses the following technical Lemmas. The lemmas make repeated use of the mixing inequality (26) applied with $p = q = \zeta > 3$, where ζ appears in Assumption INNOV.

Lemma 6. *Suppose $n(1 - \rho_n) \to \infty$, $\rho_n \to 1$, and Assumptions INNOV and STAT hold, then we have*

$$E(Y_{n,0}^{*2}U_{n,1}^2/\phi_{n,1}^4) - (1 - \rho_n^2)^{-1}(EU_{n,1}^2)E(U_{n,1}^2/\phi_{n,1}^4) = O(1),$$
$$E(Y_{n,0}^{*2}/\phi_{n,1}^2) - (1 - \rho_n^2)^{-1}EU_{n,1}^2E\phi_{n,1}^{-2} = O(1),$$
$$E(Y_{n,0}^{*}/\phi_{n,1}^2) = O(1), \quad \text{and}$$
$$E(Y_{n,0}^{*}U_{n,1}^2/\phi_{n,1}^4) = O(1).$$

Lemma 7. *Suppose $n(1 - \rho_n) \to \infty$, $\rho_n \to 1$ and Assumptions INNOV and STAT hold, then we have*

$$E\left(\sum_{i=1}^n [E\zeta_{n,i}^2 - E(\zeta_{n,i}^2|\mathcal{G}_{n,i-1})]\right)^2 \to 0,$$
$$\text{where } \zeta_{n,i} \equiv n^{-1/2}\frac{Y_{n,i-1}^*U_{n,i}/\phi_{n,i}^2}{(E(Y_{n,0}^{*2}U_{n,1}^2/\phi_{n,1}^4))^{1/2}}.$$

In Lemma 8, X_1, X_2, INNOV and U are defined as in the paragraph containing (4), but with $\phi_{n,i}$ in place of $\widehat{\phi}_{n,i}$.

Lemma 8. *Suppose $n(1 - \rho_n) \to \infty$, $\rho_n \to 1$, and Assumptions INNOV and STAT hold, then we have*

(a) $n^{-1}(1 - \rho_n)^{1/2}X_1'X_2 = o_p(1)$,
(b) $E(Y_{n,0}^{*2}/\phi_{n,1}^2)^{-1}n^{-1}X_1'X_1 \to_p 1$,
(c) $(E(Y_{n,0}^{*2}U_{n,1}^2/\phi_{n,1}^4))^{-1}n^{-1}\sum_{i=1}^n(Y_{n,i-1}^{*2}U_{n,i}^2/\phi_{n,i}^4) \to_p 1$,
(d) $(X'X)^{-1}X'U = (O_p((1 - \rho_n)^{1/2}n^{-1/2}), O_p(n^{-1/2}))'$,
(e) $(E(Y_{n,0}^{*2}U_{n,1}^2/\phi_{n,1}^4))^{-1}n^{-1}X_1'\Delta^2X_1 \to_p 1$,
(f) $(1 - \rho_n)^{1/2}n^{-1}(X_2'\Delta^2X_1) = O_p(1)$, *and*
(g) $n^{-1}(X_2'\Delta^2X_2) = O_p(1)$.

Lemma 9. *Suppose $n(1 - \rho_n) \to \infty$, $\rho_n \to 1$, and Assumptions INNOV and STAT hold. Then, we have $\sum_{i=1}^n E(\zeta_{n,i}^2 1(|\zeta_{n,i}| > \delta)|\mathcal{G}_{n,i-1}) \to_p 0$ for any $\delta > 0$.*

4.2.2. Proof of Theorem 1

To simplify notation, in the remainder of the paper we often leave out the subscript n. For example, instead of ρ_n, $\sigma_{U,n}^2$, $Y_{n,i}^*$, $U_{n,i}$, $\phi_{n,i}$, $\widehat{\phi}_{n,i}$, and $\zeta_{n,i}$, we write ρ, σ_U^2, Y_i^*, U_i, ϕ_i, $\widehat{\phi}_i$, and ζ_i. We do not drop n from $h_{n,1}$ because $h_{n,1}$ and h_1 are different quantities. As above, we omit the subscript F_n on expectations.

In the proof of Theorem 1, X_1, X_2, U, Δ, and Y are defined as in the paragraph containing (4), but with ϕ_i in place of $\widehat{\phi}_{n,i}$.

Proof of Theorem 1. First we prove part (a) of the theorem when $h_1 > 0$. In this case, $a_n = n$ and $d_n = n^{1/2}$. We can write

$$n(\widehat{\rho}_n - \rho) = \left(n^{-2}X_1'M_{X_2}X_1\right)^{-1}n^{-1}X_1'M_{X_2}U \quad \text{and}$$
$$n\widehat{\sigma}_n^2 = \left(n^{-2}X_1'M_{X_2}X_1\right)^{-1}\left(n^{-2}X_1'M_{X_2}\Delta^2M_{X_2}X_1\right) \times \left(n^{-2}X_1'M_{X_2}X_1\right)^{-1}. \tag{27}$$

We consider the terms in (27) one at a time. First, we have

$$\begin{aligned}
&n^{-2}X_1'M_{X_2}X_1\\
&= n^{-2}\sum_{i=1}^n\left(Y_{i-1}^*/\phi_i - \left(\sum_{j=1}^n Y_{j-1}^*/\phi_j^2\right)\left(\sum_{j=1}^n\phi_j^{-2}\right)^{-1}\phi_i^{-1}\right)^2\\
&= n^{-2}\sum_{i=1}^n Y_{i-1}^{*2}/\phi_i^2 - \left(n^{-3/2}\sum_{j=1}^n Y_{j-1}^*/\phi_j^2\right)^2\left(n^{-1}\sum_{j=1}^n\phi_j^{-2}\right)^{-1}\\
&\to_d h_{2,5}h_{2,1}\int I_h^{*2} - \left(h_{2,5}h_{2,1}^{1/2}\int I_h^*\right)^2 h_{2,5}^{-1}\\
&= h_{2,5}h_{2,1}\int I_{D,h}^{*2},
\end{aligned} \tag{28}$$

where the first two equalities hold by definitions and some algebra, and the convergence holds by Lemma 5(b), (f), and (h) with $j = 2$ in part (b).

Similarly, we have

$$\begin{aligned}
n^{-1}X_1'M_{X_2}U &= n^{-1}\sum_{i=1}^n\left(Y_{i-1}^*/\phi_i - \left(\sum_{j=1}^n Y_{j-1}^*/\phi_j^2\right)\right.\\
&\quad\times\left.\left(\sum_{j=1}^n\phi_j^{-2}\right)^{-1}\phi_i^{-1}\right)U_i/\phi_i\\
&= n^{-1}\sum_{i=1}^n Y_{i-1}^*U_i/\phi_i^2 - \left(n^{-3/2}\sum_{j=1}^n Y_{j-1}^*/\phi_j^2\right)\\
&\quad\times\left(n^{-1}\sum_{j=1}^n\phi_j^{-2}\right)^{-1}n^{-1/2}\sum_{i=1}^n U_i/\phi_i^2\\
&\to_d h_{2,1}^{1/2}\int I_h^*dM - h_{2,1}^{1/2}\int I_h^*\int dM\\
&= h_{2,1}^{1/2}\int I_{D,h}^*dM,
\end{aligned} \tag{29}$$

where the first two equalities hold by definitions and some algebra, and the convergence holds by Lemma 5(b) and (e)–(g) with $j = 2$ in part (b).

To determine the asymptotic distribution of $n^{-2}X_1'M_{X_2}\Delta^2M_{X_2}X_1$, we make the following preliminary calculations. Let $\widehat{U}_i/\phi_i$ denote the ith element of $M_XY = M_XU$. That is,

$$\widehat{U}_i/\phi_i = U_i/\phi_i - A_n'B_n^{-1}\begin{pmatrix} n^{-1/2}\phi_i^{-1} \\ n^{-1}Y_{i-1}^*/\phi_i \end{pmatrix}, \quad \text{where}$$

$$A_n = \begin{pmatrix} n^{-1/2}\sum_{j=1}^n U_j/\phi_j^2 \\ n^{-1}\sum_{j=1}^n Y_{j-1}^*U_j/\phi_j^2 \end{pmatrix} \quad \text{and}$$

$$B_n = \begin{pmatrix} n^{-1}\sum_{j=1}^n \phi_j^{-2} & n^{-3/2}\sum_{j=1}^n Y_{j-1}^*/\phi_j^2 \\ n^{-3/2}\sum_{j=1}^n Y_{j-1}^*/\phi_j^2 & n^{-2}\sum_{j=1}^n Y_{j-1}^{*2}/\phi_j^2 \end{pmatrix}. \quad (30)$$

Using (30), we have

$$\begin{aligned} & n^{-2}\sum_{i=1}^n Y_{i-1}^{*2}\widehat{U}_i^2/\phi_i^4 \\ &= n^{-2}\sum_{i=1}^n Y_{i-1}^{*2}U_i^2/\phi_i^4 - 2n^{-1}A_n' B_n^{-1} \\ &\quad \times \begin{pmatrix} n^{-3/2}\sum_{i=1}^n Y_{i-1}^{*2}U_i/\phi_i^4 \\ n^{-2}\sum_{i=1}^n Y_{i-1}^{*3}U_i/\phi_i^4 \end{pmatrix} + n^{-1}A_n' B_n^{-1} \\ &\quad \times \begin{pmatrix} n^{-2}\sum_{i=1}^n Y_{i-1}^{*2}/\phi_i^4 & n^{-5/2}\sum_{i=1}^n Y_{i-1}^{*3}/\phi_i^4 \\ n^{-5/2}\sum_{i=1}^n Y_{i-1}^{*3}/\phi_i^4 & n^{-3}\sum_{i=1}^n Y_{i-1}^{*4}/\phi_i^4 \end{pmatrix} B_n^{-1}A_n \\ &= n^{-2}\sum_{i=1}^n Y_{i-1}^{*2}U_i^2/\phi_i^4 + o_p(1), \end{aligned} \quad (31)$$

where the second equality holds using Lemma 5(k) with $(\ell_1, \ell_2) = (2, 1), (3, 1), (2, 0), (3, 0)$, and $(4, 0)$ and to show that A_n and B_n^{-1} are $O_p(1)$ we use Lemma 5(b) and (e)–(h) with $j = 2$ in part (b).

Similarly to (31) but with Y_{i-1}^* in place of Y_{i-1}^{*2}, and then with Y_{i-1}^{*2} deleted, we have

$$n^{-3/2}\sum_{i=1}^n Y_{i-1}^*\widehat{U}_i^2/\phi_i^4 = n^{-3/2}\sum_{i=1}^n Y_{i-1}^*U_i^2/\phi_i^4 + o_p(1) \quad \text{and}$$
$$n^{-1}\sum_{i=1}^n \widehat{U}_i^2/\phi_i^4 = n^{-1}\sum_{i=1}^n U_i^2/\phi_i^4 + o_p(1) \quad (32)$$

using Lemma 5 as above to show that A_n and B_n^{-1} are $O_p(1)$, using Lemma 5(k) with $(\ell_1, \ell_2) = (1, 1), (2, 1), (1, 0), (2, 0)$, and $(3, 0)$ for the first result, and using Lemma 5(k) with $(\ell_1, \ell_2) = (1, 1), (1, 0)$, and $(2, 0)$, Lemma 5(b) with $j = 4$, and Lemma 5(c) for the second result.

We now have

$$\begin{aligned} & n^{-2}X_1' M_{X_2}\Delta^2 M_{X_2}X_1 \\ &= n^{-2}\sum_{i=1}^n \left(\widehat{U}_i^2/\phi_i^2\right)\left(Y_{i-1}^*/\phi_i - \left(\sum_{j=1}^n Y_{j-1}^*/\phi_j^2\right)\right. \\ &\quad \left.\times \left(\sum_{j=1}^n \phi_j^{-2}\right)^{-1}\phi_i^{-1}\right)^2 \\ &= n^{-2}\sum_{i=1}^n Y_{i-1}^{*2}\widehat{U}_i^2/\phi_i^4 - 2\left(n^{-3/2}\sum_{j=1}^n Y_{j-1}^*/\phi_j^2\right) \\ &\quad \times \left(n^{-1}\sum_{j=1}^n \phi_j^{-2}\right)^{-1} n^{-3/2}\sum_{i=1}^n Y_{i-1}^*\widehat{U}_i^2/\phi_i^4 \\ &\quad + \left(n^{-3/2}\sum_{j=1}^n Y_{j-1}^*/\phi_j^2\right)^2\left(n^{-1}\sum_{j=1}^n \phi_j^{-2}\right)^{-2} n^{-1}\sum_{i=1}^n \widehat{U}_i^2\phi_i^{-4} \\ &= n^{-2}\sum_{i=1}^n Y_{i-1}^{*2}U_i^2/\phi_i^4 - 2\left(n^{-3/2}\sum_{j=1}^n Y_{j-1}^*/\phi_j^2\right) \\ &\quad \times \left(n^{-1}\sum_{j=1}^n \phi_j^{-2}\right)^{-1} n^{-3/2}\sum_{i=1}^n Y_{i-1}^*U_i^2/\phi_i^4 \\ &\quad + \left(n^{-3/2}\sum_{j=1}^n Y_{j-1}^*/\phi_j^2\right)^2\left(n^{-1}\sum_{j=1}^n \phi_j^{-2}\right)^{-2} \\ &\quad \times n^{-1}\sum_{i=1}^n U_i^2/\phi_i^4 + O_p(n^{-1}) \\ &\to_d h_{2,2}h_{2,1}\int I_h^{*2} - 2h_{2,1}^{1/2}\int I_h^* \cdot \left(h_{2,2}h_{2,1}^{1/2}\int I_h^*\right) \\ &\quad + \left(h_{2,1}^{1/2}\int I_h^*\right)^2 h_{2,2} \\ &= h_{2,2}h_{2,1}\int\left(I_h^* - \int I_h^*\right)^2 = h_{2,2}h_{2,1}\int I_{D,h}^{*2}, \end{aligned} \quad (33)$$

where the first two equalities follow from definitions and some algebra, the third equality holds by (31), (32), and Lemma 5(b), (d), (f), (i), and (j) with $j = 2$ in part (b), and the convergence holds by the same parts of Lemma 5.

Putting the results of (27)–(29) and (33), and Lemma 3 together gives

$$\begin{aligned} T_n^*(\rho_n) \to_d & \frac{h_{2,1}^{1/2}\int I_{D,h}^* dM}{\left(h_{2,2}h_{2,1}\int I_{D,h}^{*2}\right)^{1/2}} \\ = & \frac{h_{2,2}^{1/2}\int I_{D,h}^* d\left(h_{2,7}W + (1-h_{2,7}^2)^{1/2}W_2\right)}{h_{2,2}^{1/2}\left(\int I_{D,h}^{*2}\right)^{1/2}} \\ = & h_{2,7}\left(\int I_{D,h}^{*2}\right)^{-1/2}\int I_{D,h}^* dW \\ & + (1-h_{2,7}^2)^{1/2}Z_2, \end{aligned} \quad (34)$$

where the last equality uses the definition of Z_2 in (15). This completes the proof of part (a) of the theorem when $h_1 > 0$.

Next, we consider the case where $h_1 = 0$. In this case, (27)–(34) hold except that the convergence results in (28), (29) and (33) only hold with Y_{i-1}^* replaced by $\widetilde{Y}_{i-1}$ because Lemma 5(l) only applies to random variables based on a zero initial condition when $h_1 = 0$. Hence, we need to show that the difference between the second last line of (28) with Y_{i-1}^* appearing and with $\widetilde{Y}_{i-1}$ appearing is $o_p(1)$ and that analogous results hold for (29) and (33).

For $h_1 = 0$, by a mean value expansion, we have

$$\begin{aligned} \max_{0\le j\le 2n}|1-\rho^j| &= \max_{0\le j\le 2n}|1-\exp(-h_{n,1}^* j/n)| \\ &= \max_{0\le j\le 2n}|1-(1-h_{n,1}^* j\exp(m_j)/n)| \\ &\le 2h_{n,1}^*\max_{0\le j\le 2n}|\exp(m_j)| = O(h_{n,1}^*), \end{aligned} \quad (35)$$

for $0 \le |m_j| \le h_{n,1}^* j/n \le 2h_{n,1}^* \to 0$, where $h_{n,1}^*$ is defined just above (25).

Using the decomposition in (25), we have $Y_{i-1}^* = \widetilde{Y}_{i-1} + \rho^{i-1}Y_0^*$. To show the desired result for (28), we write the second last line of (28) as

$$n^{-2}\sum_{i=1}^n\left(Y_{i-1}^*/\phi_i - \left(\sum_{j=1}^n Y_{j-1}^*/\phi_j^2\right)\left(\sum_{j=1}^n \phi_j^{-2}\right)^{-1}\phi_i^{-1}\right)^2$$

$$= n^{-2}\sum_{i=1}^{n}\left(\widetilde{Y}_{i-1}/\phi_i + \rho^{i-1}Y_0^*/\phi_i - \left(\sum_{j=1}^{n}\widetilde{Y}_{j-1}/\phi_j^2 + \rho^{j-1}Y_0^*/\phi_j^2\right)\left(\sum_{j=1}^{n}\phi_j^{-2}\right)^{-1}\phi_i^{-1}\right)^2$$

$$= n^{-2}\sum_{i=1}^{n}\left(\widetilde{Y}_{i-1}/\phi_i - \left(\sum_{j=1}^{n}\widetilde{Y}_{j-1}/\phi_j^2\right)\left(\sum_{j=1}^{n}\phi_j^{-2}\right)^{-1}\phi_i^{-1} + O_p(h_{n,1}^* Y_0^*)/\phi_i\right)^2$$

$$= n^{-2}\sum_{i=1}^{n}\left(\widetilde{Y}_{i-1}/\phi_i - \left(\sum_{j=1}^{n}\widetilde{Y}_{j-1}/\phi_j^2\right)\left(\sum_{j=1}^{n}\phi_j^{-2}\right)^{-1}\phi_i^{-1}\right)^2 + O_p(n^{-1/2}h_{n,1}^* Y_0^*), \quad (36)$$

where the second equality holds because $\rho^{i-1} = 1 + O(h_{n,1}^*)$ uniformly in $i \le n$ by (35), and the third equality holds using Lemma 5. Next, Lemma 2 and $h_{n,1}^*/h_{n,1} \to 1$ (which is established at the beginning of the proof of Lemma 2 in Andrews and Guggenberger (2010b)) show that $n^{-1/2}h_{n,1}^* Y_0^* = O_p(h_{n,1}^{*1/2}) = o_p(1)$. This completes the proof of the desired result for (28) when $h_1 = 0$. The proofs for (29) and (33) are similar. This completes the proof of part (a) of the theorem.

It remains to consider the case where $h_1 = \infty$, i.e., part (b) of the theorem. The results in part (b) generalize the results in Giraitis and Phillips (2006) in the following ways: (i) from a no-intercept model to a model with an intercept, (ii) to a case in which the innovation distribution depends on n, (iii) to allow for conditional heteroskedasticity in the error distribution, (iv) to cover a quasi-GLS estimator in place of the LS estimator, and (v) to cover the standard deviation estimator as well as the GLS/LS estimator itself.

It is enough to consider the two cases $\rho \to \rho^* < 1$ and $\rho \to 1$. First, assume $\rho \to 1$ and $n(1-\rho) \to \infty$. In this case, the sequences a_n and d_n are equal to the expressions in (11) up to lower order terms. We first prove $a_n(\widehat{\rho}_n - \rho) \to_d N(0,1)$. Note that

$$a_n(\widehat{\rho}_n - \rho) = \left(n^{-1}\frac{X_1'M_{X_2}X_1}{E(Y_0^{*2}/\phi_1^2)}\right)^{-1}\frac{n^{-1/2}X_1'M_{X_2}U}{(E(Y_0^{*2}U_1^2/\phi_1^4))^{1/2}} \equiv \nu_n\xi_n, \quad (37)$$

where ν_n and ξ_n have been implicitly defined. We now show $\nu_n \to_p 1$ and $\xi_n \to_d N(0,1)$.

To show the latter, define the martingale difference sequence

$$\zeta_i \equiv n^{-1/2}\frac{Y_{i-1}^* U_i/\phi_i^2}{(E(Y_0^{*2}U_1^2/\phi_1^4))^{1/2}}. \quad (38)$$

We show that

$$\frac{n^{-1/2}X_1'P_{X_2}U}{(E(Y_0^{*2}U_1^2/\phi_1^4))^{1/2}} \to_p 0 \quad \text{and} \quad \sum_{i=1}^{n}\zeta_i \to_d N(0,1). \quad (39)$$

To show the first result, note that $n^{-1/2}X_2'U = n^{-1/2}\sum_{i=1}^n U_i/\phi_i^2 = O_p(1)$ by a CLT for a triangular array of martingale difference random variables U_i/ϕ_i^2 for which $E|U_i/\phi_i^2|^3 < \infty$ and $n^{-1}\sum_{i=1}^n (U_i^2/\phi_i^4 - EU_i^2/\phi_i^4) \to_p 0$. The latter convergence in probability condition holds by Lemma 5(d). Furthermore, $(n^{-1}X_2'X_2)^{-1} = O_p(1)$ by Lemma 5(d) and Assumption INNOV(vi). Finally, $n^{-1}(1-\rho)^{1/2}X_1'X_2 = n^{-1}(1-\rho)^{1/2}\sum_{i=1}^n Y_{i-1}^*/\phi_i^2 = o_p(1)$ by Lemma 8(a). The first result in (39) then follows because $E(Y_0^{*2}U_1^2/\phi_1^4) = O((1-\rho)^{-1})$ by Lemma 6.

To show the latter we adjust the proof of Lemma 1 in Giraitis and Phillips (2006). It is enough to prove the analogue of Eqs. (11) and (12) in Giraitis and Phillips (2006), namely the Lindeberg condition $\sum_{i=1}^n E(\zeta_i^2 1(|\zeta_i| > \delta)|\mathcal{G}_{i-1}) \to_p 0$ for any $\delta > 0$ and $\sum_{i=1}^n E(\zeta_i^2|\mathcal{G}_{i-1}) \to_p 1$. Lemma 9 shows the former and Lemma 7 implies the latter, because by stationarity (within rows) we have $\sum_{i=1}^n E\zeta_i^2 = 1$.

By Lemma 8(b) and Lemma 6

$$\frac{n^{-1}X_1'X_1}{E(Y_0^{*2}/\phi_1^2)} \to_p 1 \quad \text{and} \quad \frac{n^{-1}X_1'P_{X_2}X_1}{E(Y_0^{*2}/\phi_1^2)} \to_p 0 \quad (40)$$

which imply $\nu_n \to_p 1$.

We next show that $d_n\widehat{\sigma}_n \to_p 1$. By (40) it is enough to show that

$$\frac{n^{-1}X_1'M_{X_2}\Delta^2 M_{X_2}X_1}{E(Y_0^{*2}U_1^2/\phi_1^4)} \to_p 1. \quad (41)$$

Lemma 8(e)–(g) shows that $(E(Y_0^{*2}U_1^2/\phi_1^4))^{-1}n^{-1}X_1'\Delta^2X_1 \to_p 1$, $(1-\rho)^{1/2}n^{-1}(X_2'\Delta^2X_1) = O_p(1)$, and $n^{-1}(X_2'\Delta^2X_2) = O_p(1)$. These results combined with Lemma 6, $(n^{-1}X_2'X_2)^{-1} = O_p(1)$, and $n^{-1}(1-\rho)^{1/2}X_1'X_2 = o_p(1)$ imply (41).

In the case $\rho \to \rho^* < 1$, Theorem 1(b) follows by using appropriate CLTs for martingale difference sequences and weak laws of large numbers. For example, the analogue to the expression in parentheses in (37) satisfies

$$\frac{n^{-1}X_1'M_{X_2}X_1}{E(Y_0^{*2}/\phi_1^2) - (E(Y_0^*/\phi_1^2))^2/E(\phi_1^{-2})} \to_p 1. \quad (42)$$

This follows by a weak law of large numbers for triangular arrays of mean zero, $L^{1+\delta}$ bounded (for some $\delta > 0$), near-epoch dependent random variables. Andrews (1988, p. 464) shows that the latter conditions imply that the array is a uniformly integrable L^1 mixingale for which a WLLN holds, see Andrews (1988, Theorem 2). For example, to show $n^{-1}X_1'X_1 - E(Y_0^{*2}/\phi_1^2) \to_p 0$, note that $Y_{i-1}^{*2}/\phi_1^2 - EY_0^{*2}/\phi_1^2$ is near-epoch dependent with respect to the σ-field $\mathcal{G}_i$ using the moment conditions in Assumption INNOV(iii), $\sum_{j=0}^{\infty}\rho^{*j} = (1-\rho^*)^{-1} < \infty$, and $\rho \to \rho^* < 1$. □

4.3. *Proof of Theorem 2*

Proof of Theorem 2. Suppose $h_1 \in [0,\infty)$. Inspection of the proof of Theorem 1 shows that it suffices to show that Lemma 5 holds with $\widehat{\phi}_i$ in place of ϕ_i. The difference between the lhs quantity in Lemma 5(b) with $j=1$ and the corresponding quantity with $\widehat{\phi}_i$ in place of ϕ_i is $o_p(1)$ by Assumption CHE(ii)(b) with $(d,j) = (0,1)$. The same result holds for $j=2$ because

$$\left|n^{-1}\sum_{i=1}^{n}\widehat{\phi}_i^{-2} - \phi_i^{-2}\right| \le n^{-1}\sum_{i=1}^{n}\widehat{\phi}_i^{-1}|\widehat{\phi}_i^{-1} - \phi_i^{-1}| + n^{-1}\sum_{i=1}^{n}\phi_i^{-1}|\widehat{\phi}_i^{-1} - \phi_i^{-1}| \le 2\varepsilon^{-1/2}n^{-1}\sum_{i=1}^{n}|\widehat{\phi}_i^{-1} - \phi_i^{-1}| = o_p(1), \quad (43)$$

where the first inequality holds by the triangle inequality, the second inequality holds by Assumption CHE(i), and the equality holds by Assumption CHE(ii)(b) with $(d,j) = (0,1)$. For $j=4$, the same result holds by the same argument as just given with 4 in place of 2 in the first line and 2 in place of 1 in the second and third lines.

The differences between the lhs quantities in Lemma 5(c) and (d) and the corresponding quantities with $\widehat{\phi}_i$ in place of ϕ_i are $o_p(1)$ by the same argument as in (43) (with 4 in place of 2 in the first line and 2 in place of 1 in the second and third lines) using Assumption CHE(ii)(b) with $(d, j) = (1, 2)$ and $(2, 2)$, respectively.

The differences between the lhs quantities in Lemma 5(e) and (g) and the corresponding quantities with $\widehat{\phi}_i$ in place of ϕ_i are $o_p(1)$ by Assumption CHE(ii)(a) with $j = 0$ and $j = 1$, respectively.

The difference between the lhs quantity in Lemma 5(f) and the corresponding quantity with $\widehat{\phi}_i$ in place of ϕ_i is $o_p(1)$ because

$$\begin{aligned}&\left|n^{-3/2}\sum_{i=1}^{n}Y_{i-1}^{*}(\widehat{\phi}_i^{-2}-\phi_i^{-2})\right|\\&\quad\leq\sup_{i\leq n,n\geq1}|n^{-1/2}Y_{i-1}^{*}|\cdot n^{-1}\sum_{i=1}^{n}|\widehat{\phi}_i^{-2}-\phi_i^{-2}|=o_p(1),\end{aligned}\tag{44}$$

where the equality holds by (43) and $\sup_{i\leq n,n\geq1}|n^{-1/2}Y_{i-1}^{*}| = O_p(1)$, which holds by Lemma 5(a) and the continuous mapping theorem. Analogous results hold for Lemma 5(h)–(j) using Assumption CHE(ii)(b) with $(d, j) = (2, 2)$ for parts (i) and (j).

Next, we show that the lhs quantity in Lemma 5(k) with $\widehat{\phi}_i$ in place of ϕ_i is $o_p(n)$. We have

$$\begin{aligned}&\left|n^{-1-\ell_1/2}\sum_{i=1}^{n}Y_{i-1}^{*\ell_1}U_i^{\ell_2}/\widehat{\phi}_i^{4}\right|\\&\quad\leq\varepsilon^{-2}\sup_{i\leq n,n\geq1}|n^{-1/2}Y_{i-1}^{*}|^{\ell_1}\cdot n^{-1}\sum_{i=1}^{n}|U_i|=O_p(1),\end{aligned}\tag{45}$$

using Assumption CHE(i), $\sup_{i\leq n,n\geq1}|n^{-1/2}Y_{i-1}^{*}| = O_p(1)$, and a WLLN for strong-mixing triangular arrays of $L^{1+\delta}$-bounded random variables, see Andrews (1988), which relies on Assumption INNOV(iii). The results in Lemma 5(l) hold by the same arguments as given above.

Next, suppose $h_1 = \infty$. Lemma 6 shows that $E(Y_0^{*2}/\phi_1^2) = O((1-\rho)^{-1})$ and $E(Y_0^{*2}U_1^2/\phi_1^4) = O((1-\rho)^{-1})$, where $O((1-\rho)^{-1}) = O(1)$ in the case where $\rho \to \rho^* < 1$. Inspection of the proof of Theorem 1 then shows that it suffices to show that the equivalent of (39)–(41) holds when ϕ_i is replaced by $\widehat{\phi}_i$. More precisely, by Lemma 6, for (40) it is sufficient to show that

$$\text{(i) } n^{-1}(1-\rho)\sum_{i=1}^{n}(Y_{i-1}^{*})^2(\widehat{\phi}_i^{-2}-\phi_i^{-2})=o_p(1),\tag{46}$$

(ii) $n^{-1}(1-\rho)^{1/2}\sum_{i=1}^{n}Y_{i-1}^{*}(\widehat{\phi}_i^{-2}-\phi_i^{-2}) = o_p(1)$, and (iii) $n^{-1}\sum_{i=1}^{n}(\widehat{\phi}_i^{-2}-\phi_i^{-2}) = o_p(1)$. In addition, for (39), it is sufficient to show that (iv) $n^{-1/2}\sum_{i=1}^{n}((1-\rho)^{1/2}Y_{i-1}^{*})^jU_i \times (\widehat{\phi}_i^{-2}-\phi_i^{-2}) = o_p(1)$ for $j = 0, 1$. To show (41), it is enough to show that in addition $n^{-1}(1-\rho)X_1'\Delta^2X_1 \to_p 1$, $n^{-1}(1-\rho)^{1/2}(X_2'\Delta^2X_1) = O_p(1)$, and $n^{-1}(X_2'\Delta^2X_2) = O_p(1)$ hold (with X_1, X_2, and Δ defined with $\widehat{\phi}_i$, not ϕ_i). Inspecting the proof of Lemma 8(e)–(g) in Andrews and Guggenberger (2010b) carefully, it follows that to show the latter three conditions, it is enough to show that in addition to (i)–(iv), we have (v) $n^{-1}(1-\rho)\sum_{i=1}^{n}(Y_{i-1}^{*})^2U_i^2(\widehat{\phi}_i^{-4}-\phi_i^{-4}) = o_p(1)$ and (vi) $n^{-r_1}(1-\rho)^{r_2}\sum_{i=1}^{n}(Y_{i-1}^{*})^{r_3}U_i^{r_4}(\widehat{\phi}_i^{-4}-\phi_i^{-4}) = o_p(1)$ for $(r_1,\ldots,r_4) = (3/2, 1, 2, 1), (2, 1, 2, 0), (3/2, 3/2, 3, 1), (2, 3/2, 3, 0)$, and $(2, 3/2, 4, 0)$. These conditions come from the proof of Lemma 8.

Conditions (iii) and (iv) are assumed in Assumption CHE(ii)(c) and (d). Immediately below we prove (i) in (46) using Assumption CHE(ii)(d) with $(d, j, k) = (2, 2, 0)$; (ii), (v), and (vi) can be shown using exactly the same approach by applying Assumption CHE(ii)(d) with $(d, j, k) = (1, 2, 0), (2, 4, 0), (2, 4, 2)$, and $(2, 4, 4)$, respectively.

We now prove (i) in (46). Note that by the Cauchy–Schwarz inequality we have

$$\begin{aligned}&n^{-1}(1-\rho)\sum_{i=1}^{n}(Y_{i-1}^{*})^2(\widehat{\phi}_i^{-2}-\phi_i^{-2})\\&\quad\leq\left(n^{-1}(1-\rho)^2\sum_{i=1}^{n}(Y_{i-1}^{*})^4\right)^{1/2}\left(n^{-1}\sum_{i=1}^{n}(\widehat{\phi}_i^{-2}-\phi_i^{-2})^2\right)^{1/2}\end{aligned}\tag{47}$$

and therefore by Assumption CHE(ii)(d) it is enough to show that $n^{-1}(1-\rho)^2\sum_{i=1}^{n}(Y_{i-1}^{*})^4 = O_p(1)$. By Markov's inequality, we have

$$\begin{aligned}&P\left(n^{-1}(1-\rho)^2\sum_{i=1}^{n}(Y_{i-1}^{*})^4>M\right)\\&\quad\leq M^{-2}n^{-2}(1-\rho)^4\sum_{i,j=1}^{n}E(Y_{i-1}^{*}Y_{j-1}^{*})^4.\end{aligned}\tag{48}$$

Thus, it is enough to show that for

$$\begin{aligned}E_{ijstuvabcd}&=E(U_{i-1-s}U_{i-1-t}U_{i-1-u}U_{i-1-v}\\&\quad\times U_{j-1-a}U_{j-1-b}U_{j-1-c}U_{j-1-d}),\end{aligned}\tag{49}$$

we have

$$\begin{aligned}&n^{-2}(1-\rho)^4\sum_{i,j=1}^{n}\sum_{s,t,u,v=0}^{\infty}\sum_{a,b,c,d=0}^{\infty}\rho^{a+b+c+d+s+t+u+v}E_{ijstuvabcd}\\&\quad=O(1).\end{aligned}\tag{50}$$

In the case where $\rho \to \rho^* < 1$, (50) holds by Assumption INNOV(iii). Next consider the case when $\rho \to 1$. Note that when the largest subindex $i-1-s, \ldots, j-1-d$ in (50) appears only once in $E_{ijstuvabcd}$, then the expectation equals zero because U_i is a martingale difference sequence. As in some proofs of Lemmas 2–9, one can then show that it is enough to consider the case where the largest subindex appears twice and all other subindices are different from each other. One has to consider different subcases regarding the order of the subindices. We consider only one case here, namely the case where $i-1-s < i-1-t < \cdots < j-1-b < j-1-c = j-1-d$ and thus $c = d$. The other cases are handled using an analogous approach. We make use of the mixing inequality in (26) and apply Assumption INNOV(iii). Note that

$$\begin{aligned}&n^{-2}(1-\rho)^4\sum_{i,j=1}^{n}\sum_{s>t>u>v=0}^{\infty}\sum_{a>b>c=0}^{\infty}\rho^{a+b+2c+s+t+u+v}E_{ijstuvabcc}\\&\quad=O(n^{-2}(1-\rho)^4)\sum_{i,j=1}^{n}\sum_{s>t>u>v=0}^{\infty}\sum_{a>b>c=0}^{\infty}\rho^{a+b+2c+s+t+u+v}\\&\qquad\times(\max\{s-t,t-u,b-c\})^{-3-\varepsilon}\\&\quad=O(n^{-2}(1-\rho)^3)\sum_{i,j=1}^{n}\sum_{s>t=0}^{\infty}\rho^s(s-t)^{-1-\varepsilon/3}\\&\qquad\times\sum_{u>v=0}^{\infty}\rho^s(s-t)^{-1-\varepsilon/3}\sum_{b>c=0}^{\infty}\rho^b(b-c)^{-1-\varepsilon/3}\\&\quad=O(1),\end{aligned}\tag{51}$$

where the last equality holds because $\sum_{b>c=0}^{\infty}\rho^b(b-c)^{-1-\varepsilon/3} = \sum_{c=0}^{\infty}\rho^c\sum_{b=1}^{\infty}\rho^b b^{-1-\varepsilon/3} = O((1-\rho)^{-1})$. This completes the proof of (i) in (46). □

4.4. Proof of Lemma 1

Proof of Lemma 1. By Assumption CHE2(i) and (v), Assumption CHE(i) holds. We verify Assumption CHE(ii)(a) (which applies when $h_1 \in [0, \infty)$) for $j = 1$. The proof for $j = 0$ is similar. We need to show that

$$n^{-1/2}\sum_{i=1}^{n}(n^{-1/2}Y_{i-1}^{*})U_i[\widehat{\phi}_i^{-2} - \phi_i^{-2}] = o_p(1). \quad (52)$$

To do so, we need to take account of the fact that under Assumption CHE2, $\widehat{\phi}_i^2$ differs from ϕ_i^2 in three ways. First, $\widehat{\phi}_i^2$ is based on the estimated conditional heteroskedasticity parameter $\widetilde{\pi}_n$, not the pseudo-true value π_n; second, $\widehat{\phi}_i^2$ is based on residuals, i.e., it uses $(\widetilde{\alpha}_n, \widetilde{\rho}_n)$, not the true values $(0, \rho_n)$; and third $\widehat{\phi}_i^2$ is defined using the truncated-at-time-period-one value L_i, not L.

Assumption CHE2(iii) and (iv) implies that $\|\widetilde{\pi}_n - \pi_n\| \leq Cn^{-\delta_2}$ wp $\to$ 1 for some constant $C < \infty$. Hence, $\widetilde{\pi}_n \in \Pi_{n,0} = \Pi_n \cap B(\pi_n, Cn^{-\delta_2})$ wp $\to$ 1 (where $B(\pi, \delta)$ denotes a ball with center at π and radius δ). The set $\Pi_{n,0}$ contains a finite number of elements and the number is bounded over $n \geq 1$. Without loss of generality, we can assume that $\Pi_{n,0}$ contains $K < \infty$ elements for each $n \geq 1$. We order the elements in each set $\Pi_{n,0}$ and call them $\pi_{n,k}$ for $k = 1, \ldots, K$. This yields K sequences $\{\pi_{n,k} : n \geq 1\}$ for $k = 1, \ldots, K$.

To show (52), we use the following argument. Suppose for some random variables $\{(Z_{n,0}, Z_n(\pi_{n,1}), \ldots, Z_n(\pi_{n,K}))' : n \geq 1\}$ and Z, we have

$$(Z_{n,0}, Z_n(\pi_{n,1}), \ldots, Z_n(\pi_{n,K}))' \to_d (Z, \ldots, Z)' \quad (53)$$

as $n \to \infty$. In addition, suppose $\widetilde{\pi}_n \in \{\pi_{n,1}, \ldots, \pi_{n,K}\}$ wp $\to$ 1. Then, by the continuous mapping theorem,

$$\min_{k\leq K} Z_n(\pi_{n,k}) - Z_{n,0} \to_d \left(\min_{k\leq K} Z\right) - Z = 0,$$
$$\max_{k\leq K} Z_n(\pi_{n,k}) - Z_{n,0} \to_d \left(\max_{k\leq K} Z\right) - Z = 0,$$
$$Z_n(\widetilde{\pi}_n) - Z_{n,0} \in [\min_{k\leq K} Z_n(\pi_{n,k}) - Z_{n,0}, \max_{k\leq K} Z_n(\pi_{n,k}) - Z_{n,0}] \text{ wp} \to 1, \quad \text{and hence,}$$
$$Z_n(\widetilde{\pi}_n) - Z_{n,0} \to_d 0. \quad (54)$$

Since convergence in distribution to zero is equivalent to convergence in probability to zero, this gives $Z_n(\widetilde{\pi}_n) - Z_{n,0} \to_p 0$. We apply this argument with

$$Z_{n,0} = n^{-1/2}\sum_{i=1}^{n}(n^{-1/2}Y_{i-1}^{*})U_i\phi_i^{-2} \quad \text{and}$$
$$Z_n(\pi_{n,k}) = n^{-1/2}\sum_{i=1}^{n}(n^{-1/2}Y_{i-1}^{*})U_i\phi_i^{-2}(\widetilde{\alpha}_n, \widetilde{\rho}_n, \pi_{n,k}) \quad (55)$$

for $k = 1, \ldots, K$.

Hence, it suffices to show (53), where $\{\pi_{n,k} : n \geq 1\}$ is a fixed sequence such that $\pi_{n,k} \to \pi_0$ for $k = 1, \ldots, K$. To do so, we show below that

$$Z_n(\pi_{n,k}) - \overline{Z}_n(\pi_{n,k}) = o_p(1), \quad \text{where}$$
$$\overline{Z}_n(\pi_{n,k}) = n^{-1/2}\sum_{i=1}^{n}(n^{-1/2}Y_{i-1}^{*})U_i\phi_i^{-2}(0, \rho_n, \pi_{n,k}). \quad (56)$$

(By definition, $\overline{Z}_n(\pi_{n,k})$ is the same as $Z_n(\pi_{n,k})$ except that it is defined using the true parameters $(0, \rho_n)$ rather than the estimated parameters $(\widetilde{\alpha}_n, \widetilde{\rho}_n)$.) It is then enough to show that (53) holds with $\overline{Z}_n(\pi_{n,k})$ in place of $Z_n(\pi_{n,k})$.

For the case $h_1 \in [0, \infty)$ considered here, we do the latter by applying Lemma 4 with

$$v_{n,i} = (U_i, U_i\phi_i^{-2}, U_i\phi_i^{-2}(0, \rho_n, \pi_{n,1}), \ldots, U_i\phi_i^{-2}(0, \rho_n, \pi_{n,K}))'. \quad (57)$$

Conditions (i) and (ii) of Lemma 4 hold by Assumptions INNOV and CHE2(v) (which guarantees that $\widehat{\phi}_i^{-2}$ and $\phi_i^{-2}(0, \rho_n, \pi_{n,k})$ are uniformly bounded above). In addition, $\Lambda = 0$ because $\{(v_{n,i}, \mathcal{G}_{n,i-1}) : i = \ldots, 0, 1, \ldots; n \geq 1\}$ is a martingale difference triangular array. Using Assumption CHE2(vi), for all $k_1, k_2, k_3, k_4 = 0, \ldots, K$, we have

$$\lim_{n\to\infty} n^{-1}EV_{n,k_1}V_{n,k_2}' = \lim_{n\to\infty} n^{-1}EV_{n,k_3}V_{n,k_4}', \quad \text{where}$$
$$V_{n,0} = \sum_{i=1}^{n} U_i\phi_i^{-2} = \sum_{i=1}^{n} U_i\left(\omega_n + \sum_{j=1}^{L}\mu_j(\pi_n)U_{i-j}^2\right) \quad \text{and}$$
$$V_{n,k} = \sum_{i=1}^{n} U_i\phi_i^{-2}(0, \rho_n, \pi_{n,k}) = \sum_{i=1}^{n} U_i\left(\omega_{n,k} + \sum_{j=1}^{L_i}\mu_j(\pi_{n,k})U_{i-j}^2\right) \quad (58)$$

for $k = 1, \ldots, K$. In consequence, the matrix Ω_0 in Lemma 4 has all elements that are not in the first row or column equal to each other. For this reason, the elements in the limit random vector in (53) are equal to each other. We conclude that (53) holds when $\overline{Z}_n(\pi_{n,k})$ appears in place of $Z_n(\pi_{n,k})$ by Lemma 4(b). In this case, $Z = h_{2,1}^{1/2}\int I_h^* dM$, see Lemma 5(g) and its proof in Andrews and Guggenberger (2010b). The verification of Assumption CHE(ii)(a) when $j = 0$ is the same as that above because one of the elements of X_{i-1} in Lemma 4(b) can be taken to equal 1 and the latter result still holds with the corresponding element of K_h^* being equal to 1, see Hansen (1992, Theorem 3.1).

It remains to show (56) holds in the case $h_1 \in [0, \infty)$ considered here. We only deal with the case $j = 1$. The case $j = 0$ can be handled analogously. To evaluate $\phi_i^{-2}(\widetilde{\alpha}_n, \widetilde{\rho}_n, \pi_{n,k}) - \phi_i^{-2}(0, \rho_n, \pi_{n,k})$, we use the Taylor expansion

$$(x+\delta)^{-1} = x^{-1} - x^{-2}\delta + x_*^{-3}\delta^2, \quad (59)$$

where x_* is between $x + \delta$ and x, applied with $x + \delta = \phi_i^2(\widetilde{\alpha}_n, \widetilde{\rho}_n, \pi_{n,k})$, $x = \phi_i^2(0, \rho_n, \pi_{n,k})$, and

$$\delta = \delta_i = \phi_i^2(\widetilde{\alpha}_n, \widetilde{\rho}_n, \pi_{n,k}) - \phi_i^2(0, \rho_n, \pi_{n,k}). \quad (60)$$

Thus, to show Assumption CHE(ii)(a), it suffices to show that

$$n^{-1/2}\sum_{i=1}^{n}(n^{-1/2}Y_{i-1}^{*})U_i(\phi_i^{-4}(0, \rho_n, \pi_{n,k})\delta - x_*^{-3}\delta^2) = o_p(1). \quad (61)$$

Note that in the Taylor expansion, x^{-2} and x_*^{-3} are both bounded above (uniformly in i) because both $x + \delta$ and x are bounded away from zero by Assumption CHE2(v). Simple algebra gives

$$\delta = \sum_{t=1}^{L_i}\mu_t(\pi_{n,k})[-2U_{i-t}\widetilde{\alpha}_n - 2Y_{i-t-1}^{*}U_{i-t}(\widetilde{\rho}_n - \rho_n) + \widetilde{\alpha}_n^2 + 2Y_{i-t-1}^{*}(\widetilde{\rho}_n - \rho_n)\widetilde{\alpha}_n + Y_{i-t-1}^{*2}(\widetilde{\rho}_n - \rho_n)^2]. \quad (62)$$

The effect of truncation by L_i rather than L only affects the finite number of summands with $i \leq L$ and hence its effect is easily seen to be asymptotically negligible and hence without loss of generality we can set $L_i = L$ for the rest of the proof.

We first deal with the contributions from $\phi_i^{-4}(0, \rho_n, \pi_{n,k})\delta$ in (61). Rather than considering the sum $\sum_{t=1}^{L_i}$ in (62) when showing

(61), it is enough to show that for every fixed $t = 1, \ldots, L$ the resulting expression in (61) is $o_p(1)$. Fix $t \in \{1, \ldots, L\}$ and set $b_i = \phi_i^{-4}(0, \rho, \pi_{n,k})$. It is enough to show that

$$n^{-1/2} \sum_{i=1}^{n} (n^{-1/2} Y_{i-1}^{*}) U_i b_i c_{it} = o_p(1), \tag{63}$$

where c_{it} equals

$$\text{(i) } U_{i-t}\widetilde{\alpha}_n, \quad \text{(ii) } Y_{i-t-1}^{*} U_{i-t}(\widetilde{\rho}_n - \rho), \quad \text{(iii) } \widetilde{\alpha}_n^2,$$
$$\text{(iv) } Y_{i-t-1}^{*}(\widetilde{\rho}_n - \rho)\widetilde{\alpha}_n, \quad \text{or} \quad \text{(v) } Y_{i-t-1}^{*2}(\widetilde{\rho}_n - \rho)^2. \tag{64}$$

By Assumption CHE2(iii) and because $h_1 \in [0, \infty)$, we have (1) $\widetilde{\alpha}_n = O_p(n^{-1/2})$ and $\widetilde{\rho}_n - \rho = O_p(n^{-1})$. Terms of the form (2) $n^{-1} \sum_{i=1}^{n} Y_{i-1}^{*} U_i b_i U_{i-t}$ and $n^{-3/2} \sum_{i=1}^{n} Y_{i-1}^{*} Y_{i-t-1}^{*} U_i U_{i-t} b_i$ are $O_p(1)$ by Lemma 4(b)–(c) applied with $v_{n,i} = (U_i, U_{i-t}, U_i U_{i-t} b_i)'$. Note here that b_i is an element of the σ-field $\sigma(U_{i-L}, \ldots, U_{i-1})$ by definition of $\phi_i^2(0, \rho, \pi_{n,k})$ in (19) and by Assumption CHE2(i) and (v), (3) $\sup_{i \le n, n \ge 1} |n^{-1/2} Y_{i-1}^{*}| = O_p(1)$ by Lemma 5(a), (4) terms of the form $n^{-1} \sum_{i=1}^{n} |U_i U_{i-1}^{j}|$ for $j = 1, 2$ are $O_p(1)$ by a WLLN for strong-mixing triangular arrays, see Andrews (1988), and (5) the b_i are $O_p(1)$ uniformly in i. The result in (63) for cases (i)–(ii) of (64) follows from (2). Cases (iii)–(v) are established by $|n^{-1/2} \sum_{i=1}^{n} (n^{-1/2} Y_{i-1}^{*}) U_i b_i c_{it}| \le \sup_{i \le n, n \ge 1} |n^{-1/2} Y_{i-1}^{*}| n^{-3/2} \times \sum_{i=1}^{n} |U_i| = o_p(1)$ using (1) and (3)–(5).

Next, we deal with the contributions from $x_*^{-3}\delta^2$ in (61). Because x_*^{-3} and $\mu_t(\pi_{n,k})$ are both $O_p(1)$ uniformly in i, it is enough to show that

$$n^{-1/2} \sum_{i=1}^{n} |n^{-1/2} Y_{i-1}^{*} U_i c_{ij_1} d_{ij_2}| = o_p(1), \tag{65}$$

where c_{ij} and $d_{ij} \in \{U_{i-j}\widetilde{\alpha}_n, Y_{i-j-1}^{*} U_{i-j}(\widetilde{\rho}_n - \rho), \widetilde{\alpha}_n^2, Y_{i-j-1}^{*}(\widetilde{\rho}_n - \rho)\widetilde{\alpha}_n, Y_{i-j-1}^{*2}(\widetilde{\rho}_n - \rho)^2\}$ and $j_1, j_2 \in \{1, \ldots, L_i\}$. Conditions (1), (3), and (4) imply (65). This completes the proof of Assumption CHE(ii)(a).

Next, we verify Assumption CHE(ii)(b) (which applies when $h_1 \in [0, \infty)$). For the cases of $(d, j) = (0, 2), (1, 2)$, and $(2, 2)$, the proof is similar to that given below for Assumption CHE(ii)(d) but with $a_n = O(n^{1/2}(1-\rho)^{-1/2})$ replaced by $a_n = n$ and using the results above that (i) $\sup_{i \le n, n \ge 1} |n^{-1/2} Y_{i-1}^{*}| = O_p(1)$ and (ii) terms of the form $n^{-1} \sum_{i=1}^{n} |U_i^{j_1} U_{i-1}^{j_2}|$ for $j_1 = 1, 2$ and $j_2 = 1, 2$ are $O_p(1)$, which holds using Assumption INNOV(iii). (Note that the case of $(d, j) = (0, 2)$ is not needed for Assumption CHE(ii) but is used in the verification of Assumption CHE(ii)(b) for the case where $(d, j) = (0, 1)$, which follows.)

We now verify Assumption CHE(ii)(b) for $(d, j) = (0, 1)$. We have

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} |\widehat{\phi}_i^{-1} - \phi_i^{-1}| &= n^{-1} \sum_{i=1}^{n} |\widehat{\phi}_i - \phi_i| / (\widehat{\phi}_i \phi_i) \\ &\le \varepsilon^{-1} n^{-1} \sum_{i=1}^{n} |\widehat{\phi}_i - \phi_i| \\ &\le \varepsilon^{-3/2} n^{-1} \sum_{i=1}^{n} |\widehat{\phi}_i^2 - \phi_i^2|, \end{aligned} \tag{66}$$

where the first inequality holds because $\widehat{\phi}_i^2$ and ϕ_i^2 are bounded away from zero by some $\varepsilon > 0$ by Assumption CHE2(i), (ii), and (v) and the second inequality holds by the mean-value expansion $(x + \delta)^{1/2} = x^{1/2} + (1/2)x_*^{-1/2}\delta$, where x_* lies between $x + \delta$ and x, applied with $x + \delta = \widehat{\phi}_i^2, x = \phi_i^2, \delta = \widehat{\phi}_i^2 - \phi_i^2$, and $x_*^{-1/2} = \phi_{i,*}^{-1} \le \varepsilon^{-1/2}$ using Assumption CHE2(v), where $\phi_{i,*}^2$ lies between $\widehat{\phi}_i^2$ and ϕ_i^2. The rhs of (66) is $o_p(1)$ by the result above that Assumption CHE(ii)(b) holds for $(d, j) = (0, 2)$.

Next, we verify Assumption CHE(ii)(c) (which applies when $h_1 = \infty$). We only show the case $j = 1$, the case $j = 0$ is handled analogously. We use a very similar approach to the one in the proof of Assumption CHE(ii)(a). We show that (56) holds when $h_1 = \infty$ and that

$$Z_{n,0} - \overline{Z}_n(\pi_{n,k}) = o_p(1) \tag{67}$$

for every $k = 1, \ldots, K$, where

$$Z_{n,0} = n^{-1/2} \sum_{i=1}^{n} ((1-\rho)^{1/2} Y_{i-1}^{*}) U_i \phi_i^{-2},$$

$$Z_n(\pi_{n,k}) = n^{-1/2} \sum_{i=1}^{n} ((1-\rho)^{1/2} Y_{i-1}^{*}) U_i \phi_i^{-2}(\widetilde{\alpha}_n, \widetilde{\rho}_n, \pi_{n,k}), \quad \text{and}$$

$$\overline{Z}_n(\pi_{n,k}) = n^{-1/2} \sum_{i=1}^{n} ((1-\rho)^{1/2} Y_{i-1}^{*}) U_i \phi_i^{-2}(0, \rho, \pi_{n,k}). \tag{68}$$

We first show (56). By (59),

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} ((1-\rho)^{1/2} Y_{i-1}^{*}) U_i (\phi_i^{-2}(\widetilde{\alpha}_n, \widetilde{\rho}_n, \pi_{n,k}) \\ & \quad - \phi_i^{-2}(0, \rho, \pi_{n,k})) \\ & = n^{-1/2} \sum_{i=1}^{n} ((1-\rho)^{1/2} Y_{i-1}^{*}) U_i (-\phi_i^{-4}(0, \rho, \pi_{n,k})\delta \\ & \quad + x_*^{-3}\delta^2), \end{aligned} \tag{69}$$

where δ is defined in (62) and x_* in (59). Hence, it suffices to show that the expression in the second line of (69) is $o_p(1)$. First, we deal with the contributions from $-\phi_i^{-4}(0, \rho, \pi_{n,k})\delta$ in (69). Rather than considering the sum $\sum_{j=1}^{L_i}$ in (62) when showing (69), it is enough to show that for every fixed $j = 1, \ldots, L_i$ the expression in the second line of (69) is $o_p(1)$. Fix $j \in \{1, \ldots, L_i\}$, set $b_i = \phi_i^{-4}(0, \rho, \pi_{n,k})$, and note that $\mu_j(\pi_{n,k})$ is bounded by Assumption CHE2(vi). It is enough to show that

$$n^{-1/2} \sum_{i=1}^{n} ((1-\rho)^{1/2} Y_{i-1}^{*}) U_i b_i c_{ij} = o_p(1), \tag{70}$$

where c_{ij} equals

$$\text{(i) } U_{i-j}\widetilde{\alpha}_n, \quad \text{(ii) } Y_{i-j-1}^{*} U_{i-j}(\widetilde{\rho}_n - \rho), \quad \text{(iii) } \widetilde{\alpha}_n^2,$$
$$\text{(iv) } Y_{i-j-1}^{*}(\widetilde{\rho}_n - \rho)\widetilde{\alpha}_n, \quad \text{or} \quad \text{(v) } Y_{i-j-1}^{*2}(\widetilde{\rho}_n - \rho)^2. \tag{71}$$

In case (i) of (71), we use Assumption CHE2(iii) which implies $\widetilde{\alpha}_n = O_p(n^{-1/2})$. By Markov's inequality and Assumption STAT, we have

$$\begin{aligned} & P\left(\left| n^{-1}(1-\rho)^{1/2} \sum_{i=1}^{n} Y_{i-1}^{*} U_i b_i U_{i-j} \right| > \varepsilon \right) \\ & = O(n^{-2}(1-\rho)) \sum_{i,k=1}^{n} E b_i b_k Y_{i-1}^{*} Y_{k-1}^{*} U_i U_{i-j} U_k U_{k-j} \\ & = O(n^{-2}(1-\rho)) \sum_{i,k=1}^{n} \sum_{s,t=0}^{\infty} \rho^{s+t} E b_i b_k U_{i-s-1} \\ & \quad \times U_{k-t-1} U_i U_{i-j} U_k U_{k-j}. \end{aligned} \tag{72}$$

Note that b_i is an element of the σ-field $\sigma(U_{i-L}, \ldots, U_{i-1})$. The latter holds by definition of $\phi_i^2(0, \rho, \pi_{n,k})$ in (19) and by

Assumption CHE2(i) and (v). To show that the last expression in (72) is $o(1)$ we have to distinguish several subcases. As in several proofs above, we can assume that all subindices $i-s-1, k-t-1, \ldots, k-j$ are different. We only consider the case $i-s-1 < k-t-1 < i-j < k-j$. The other cases can be dealt with using an analogous approach. By Assumption INNOV(iii) and the mixing inequality in (26), we have

$$\sum_{k=1}^{n}\sum_{s,t=0}^{\infty}\sum_{i=1}^{n}\rho^{s+t}Eb_ib_kU_{i-s-1}U_{k-t-1}U_iU_{i-j}U_kU_{k-j}$$
$$= O(1)\sum_{k=1}^{n}\sum_{s,t=0}^{\infty}\sum_{i=1}^{k-t+s-1}\rho^{s+t}(k-t-i+s)^{-3-\varepsilon}$$
$$= O(1)\sum_{k=1}^{n}\sum_{s,t=0}^{\infty}\rho^{s+t}\sum_{i=1}^{k-t+s-1}i^{-3-\varepsilon}$$
$$= O(n(1-\rho)^{-2}), \tag{73}$$

where in the third line we do the change of variable $i \mapsto k-t-i+s$. This implies that the expression in (72) is $o(1)$ because $n(1-\rho) \to \infty$.

In case (ii) of (71), using $\widetilde{\rho}_n - \rho = O_p(n^{-1/2}(1-\rho)^{1/2})$ by Assumption CHE2(iii), (11), and Lemma 6, and using Markov's inequality as for case (i), it is enough to show that

$$\sum_{i,k=1}^{n}\sum_{s,t=0}^{\infty}\sum_{u,v=0}^{\infty}\rho^{s+t+u+v}Eb_{ij}b_{kj}U_{i-s-1}$$
$$\times U_{i-j-1-t}U_iU_{i-j}U_{k-u-1}U_{k-j-1-v}U_kU_{k-j} \tag{74}$$

is $o(n^2(1-\rho)^{-2})$. Again, one has to separately examine several subcases regarding the order of the subindices $i-s-1, \ldots, k-j$ on the random variables U_i. We can assume that all subindices are different. We only study the case $i-s-1 < i-j-1-t < k-u-1 < k-j-1-v < i-j$. The other cases can be handled analogously. By Assumption INNOV(iii), boundedness of b_i, and the mixing inequality in (26), the expression in (74) is of order

$$O(1)\sum_{k=1}^{n}\sum_{s,t=0}^{\infty}\sum_{u,v=0}^{\infty}\sum_{i=k-v}^{n}\rho^{s+t+u+v}$$
$$\times \max(s-t-j, i-k+v+1)^{-3-\varepsilon}$$
$$= O(1)\sum_{u,v=0}^{\infty}\rho^{u+v}\sum_{k=1}^{n}\sum_{i=k-v}^{n}(i-k+v+1)^{-3/2}$$
$$\times\sum_{s,t=0}^{\infty}\rho^{s+t}(s-t-j)^{-3/2}$$
$$= O((1-\rho)^{-3}n), \tag{75}$$

where in the first line we use $k-1-v < i$ and in the last line we use $\sum_{i=k-v}^{n}(i-k+v+1)^{-3/2} = \sum_{i=1}^{n-k+v+1}i^{-3/2} = O(1)$. The desired result then follows because $n(1-\rho) \to \infty$ implies $O((1-\rho)^{-3}n) = o(n^2(1-\rho)^{-2})$.

Cases (iii)–(v) of (71) can be handled analogously.

Next, we show that the contribution from $x_*^{-3}\delta^2$ in (69) is $o_p(1)$. Noting that x_*^{-3} and $\mu_j(\pi_{n,k})$ are $O_p(1)$ uniformly in i by Assumption CHE2(ii), (v), and (vi), it is enough to show that $n^{-1/2}(1-\rho)^{1/2}\sum_{i=1}^{n}|Y_{i-1}^*U_ic_{ij_1}d_{ij_2}| = o_p(1)$, where c_{ij} and $d_{ij} \in \{U_{i-j}\widetilde{\alpha}_n, Y_{i-j-1}^*U_{i-j}(\widetilde{\rho}_n-\rho), \widetilde{\alpha}_n^2, Y_{i-j-1}^*(\widetilde{\rho}_n-\rho)\widetilde{\alpha}_n, Y_{i-j-1}^{*2}(\widetilde{\rho}_n-\rho)^2\}$ and $j_1, j_2 \in \{1, \ldots, L_i\}$. Using $\widetilde{\alpha}_n = O_p(n^{-1/2})$ and $\widetilde{\rho}_n - \rho = O_p(n^{-1/2}(1-\rho)^{1/2})$ the latter follows easily from Markov's inequality. For example,

$$P\left(n^{-1/2}(1-\rho)^{1/2}\sum_{i=1}^{n}|Y_{i-1}^*U_i(U_{i-j_1}\widetilde{\alpha}_n)(U_{i-j_2}\widetilde{\alpha}_n)| > \varepsilon\right)$$
$$= O(n^{-3}(1-\rho))\sum_{i,k=1}^{n}\sum_{s,t=0}^{\infty}\rho^{s+t}E|U_{i-1-s}U_iU_{i-j_1}$$
$$\times U_{i-j_2}U_{k-1-t}U_kU_{k-j_1}U_{k-j_2}|$$
$$= O(n^{-3}(1-\rho))(1-\rho)^{-2}n^2$$
$$= o(1) \tag{76}$$

by Assumption INNOV(iii) and $n(1-\rho) \to \infty$.

Next we show that (67) holds. We have

$$Z_{n,0} - \overline{Z}_n(\pi_{n,k})$$
$$= n^{-1/2}\sum_{i=1}^{n}((1-\rho)^{1/2}Y_{i-1}^*)U_i(\phi_i^{-2} - \phi_i^{-2}(0,\rho,\pi_{n,k}))$$
$$= n^{-1/2}(1-\rho)^{1/2}\sum_{i=1}^{n}Y_{i-1}^*U_i(\phi_i^2(0,\rho,\pi_{n,k}) - \phi_i^2)$$
$$\times(\phi_i^{-2}\phi_i^{-2}(0,\rho,\pi_{n,k}))$$
$$= n^{-1/2}(1-\rho)^{1/2}\sum_{i=1}^{n}Y_{i-1}^*U_i\left(\omega_n - \omega_{n,k} + \sum_{j=1}^{L}(\mu_j(\pi_n)\right.$$
$$\left. - \mu_j(\pi_{n,k}))U_{i-j}^2\right)(\phi_i^{-2}\phi_i^{-2}(0,\rho,\pi_{n,k})) + o_p(1), \tag{77}$$

where ω_n is defined in Assumption CHE2(ii). Thus, it is enough to show that

$$D_1 = n^{-1/2}(1-\rho)^{1/2}\sum_{i=1}^{n}Y_{i-1}^*U_i(\omega_n - \omega_{n,k})$$
$$\times(\phi_i^{-2}\phi_i^{-2}(0,\rho,\pi_{n,k})) \quad \text{and}$$
$$D_{2j} = n^{-1/2}(1-\rho)^{1/2}\sum_{i=1}^{n}Y_{i-1}^*U_i((\mu_j(\pi_n) - \mu_j(\pi_{n,k}))U_{i-j}^2)$$
$$\times(\phi_i^{-2}\phi_i^{-2}(0,\rho,\pi_{n,k})) \tag{78}$$

are $o_p(1)$ for $j = 1, \ldots, L$. We can prove $D_{2j} = o_p(1)$ along the same lines as $D_1 = o_p(1)$ and we therefore only prove $D_1 = o_p(1)$. By Assumption CHE2(ii) and $\pi_{n,k} \to \pi_0$, we have $\omega_n - \omega_{n,k} \to 0$. Thus, by Markov's inequality and Assumption STAT,

$$P(|D_1| > \varepsilon)$$
$$= o(n^{-1}(1-\rho))\sum_{i,v=1}^{n}\sum_{s,t=0}^{\infty}\rho^{s+t}EU_{i-1-s}U_iU_{v-1-t}U_v$$
$$\times\phi_i^{-2}\phi_i^{-2}(0,\rho,\pi_{n,k})\phi_v^{-2}\phi_v^{-2}(0,\rho,\pi_{n,k}). \tag{79}$$

The random variable $e_{iv} = (\phi_i^{-2}\phi_i^{-2}(0,\rho,\pi_{n,k}))(\phi_v^{-2}\phi_v^{-2}(0,\rho,\pi_{n,k}))$ is an element of the σ-field $\sigma(U_{\min\{i,v\}-L}, \ldots, U_{\max\{i,v\}})$ by definition of $\phi_i^2(0,\rho,\pi_{n,k})$ in (19) and by Assumption CHE2(i) and (v). To prove that the rhs in (79) is $o_p(1)$ we have to study several subcases. We only examine the subcase where all subindices $i-1-s, i, v-1-t, v$ are different and where $i-1-s < i < v-1-t < v$. The other cases can be dealt with analogously. By Assumption INNOV(iii), boundedness of e_{iv}, and the mixing inequality in (26), the rhs in (79) for the particular subcase is of order

$$o(n^{-1}(1-\rho))\sum_{i,v=1}^{n}\sum_{s,t=0}^{\infty}\rho^{s+t}(s+1)^{-3/2}(v-1-t-i)^{-3/2}$$
$$= o(n^{-1}(1-\rho))\sum_{s,t=0}^{\infty}\rho^{s+t}(s+1)^{-3/2}$$
$$\times\sum_{v=1}^{n}\sum_{i=1}^{v-2-t}(v-1-t-i)^{-3/2}$$

$$= o(n^{-1}(1-\rho))O((1-\rho)^{-1})O(n)$$
$$= o(1), \tag{80}$$

where in the third line a change of variable $i \to -i-t-1+v$ was used. This completes the verification of Assumption CHE(ii)(c).

Finally, we show that Assumption CHE(ii)(d) holds. First, note that Assumptions CHE2(i), (ii), and (v) imply $\widehat{\phi}_i^{-j}\phi_i^{-j} = O_p(1)$ uniformly in i. Therefore, writing $|\widehat{\phi}_i^{-j} - \phi_i^{-j}|^d$ as $|(\phi_i^j - \widehat{\phi}_i^j)/(\widehat{\phi}_i^j\phi_i^j)|^d$ we have

$$n^{-1}\sum_{i=1}^{n}|U_i^k(\widehat{\phi}_i^{-j} - \phi_i^{-j})^d| = O_p(1)n^{-1}\sum_{i=1}^{n}|U_i^k| \cdot |\widehat{\phi}_i^j - \phi_i^j|^d. \tag{81}$$

We need to show that the quantity in (81) is $o_p(1)$. Note that by the definition of $\widehat{\phi}_i^2$ in (19) and ϕ_i^2 in Assumption CHE2(ii) we have

$$|\widehat{\phi}_i^j - \phi_i^j|^d = \left|\left(\widetilde{\omega}_n + \sum_{v=1}^{L_i}\mu_v(\widetilde{\pi}_n)\widehat{U}_{i-v}^2(\widetilde{\alpha}_n, \widetilde{\rho}_n)\right)^{j/2} - \left(\omega_n + \sum_{v=1}^{L}\mu_v(\pi_n)U_{i-v}^2\right)^{j/2}\right|^d \tag{82}$$

with $\widehat{U}_{i-v}^2(\widetilde{\alpha}_n, \widetilde{\rho}_n) = (-(\widetilde{\rho}_n - \rho)Y_{i-v-1}^* - \widetilde{\alpha}_n + U_{i-v})^2$. It can be shown that the additional terms in (81), that arise if we replace L_i by L in (82), are of order $o_p(1)$. We first study the case where $j = 2$. Multiplying out in (82), it follows that when $d = 1$, $\widehat{\phi}_i^2 - \phi_i^2$ can be bounded by a finite sum of elements in $S = \{|\widetilde{\omega}_n - \omega_n|, |\mu_v(\widetilde{\pi}_n) - \mu_v(\pi_n)|U_{i-v}^2, (\widetilde{\rho}_n - \rho)^2Y_{i-v-1}^{*2}, \widetilde{\alpha}_n^2, |(\widetilde{\rho}_n - \rho)Y_{i-v-1}^*\widetilde{\alpha}_n|, |(\widetilde{\rho}_n - \rho)Y_{i-v-1}^*U_{i-v}|, \widetilde{\alpha}_nU_{i-v} : \text{ for } v = 1, \ldots, L\}$. When $d = 2$, $(\widehat{\phi}_i^j - \phi_i^j)^2$ can be bounded by a finite sum of elements given as products of two terms in S. By Assumption CHE2(iii) and $a_n = O(n^{1/2}(1-\rho)^{-1/2})$, we have $\widetilde{\rho}_n - \rho = O_p(n^{-1/2}(1-\rho)^{1/2})$, $\widetilde{\alpha}_n = O_p(n^{-1/2})$, and $\widetilde{\omega}_n - \omega_n = O_p(n^{-\delta_2})$. To show the quantity in (81) is $o_p(1)$, it is enough to verify that $n^{-1}\sum_{i=1}^n|U_i^ks_{i1}s_{i2}| = o_p(1)$ where for $d = 1$, $s_{i1} \in S$ and $s_{i2} = 1$ and for $d = 2$, $s_{i1}, s_{i2} \in S$. We only show this for one particular choice of s_{i1}, s_{i2}, namely, $s_{i1} = s_{i2} = |\mu_v(\widetilde{\pi}_n) - \mu_v(\pi_n)|U_{i-v}^2$; the other cases can be handled analogously. In that case, we have $|\mu_v(\widetilde{\pi}_n) - \mu_v(\pi_n)|^2n^{-1}\sum_{i=1}^n|U_i^kU_{i-v}^2| = o_p(1)$ because $|\mu_v(\widetilde{\pi}_n) - \mu_v(\pi_n)|^2 = o(1)$ by Assumption CHE2(iii), (iv), and (vi), and $n^{-1}\sum_{i=1}^n|U_i^kU_{i-v}^2| = O_p(1)$ by a weak law of large numbers for triangular arrays of $L^{1+\delta}$-bounded strong-mixing random variables for $\delta > 0$, see Andrews (1988), using the moment conditions in Assumption INNOV(iii).

The case $j = 4$ can be proved analogously. □

Acknowledgments

Andrews gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-0417911 and SES-0751517. Guggenberger gratefully acknowledges the research support of the National Science Foundation via grant SES-0748922 and of the Sloan Foundation via a 2009 Sloan Fellowship. The authors would like to thank the editors and two referees for helpful comments and Brendan Beare, Tim Bollerslev, Rustam Ibragimov, and Alexander Lindner for helpful correspondence. This paper is a shortened version of Andrews and Guggenberger (2010b).

References

Andrews, D.W.K., 1988. Laws of large numbers for dependent non-identically distributed random variables. Econometric Theory 4, 458–467.
Andrews, D.W.K., 1993. Exactly median-unbiased estimation of first-order autoregressive/unit root models. Econometrica 61, 139–165.
Andrews, D.W.K., Chen, H.-Y., 1994. Approximately median-unbiased estimation of autoregressive models. Journal of Business and Economic Statistics 12, 187–204.
Andrews, D.W.K., Guggenberger, P., 2009. Hybrid and size-corrected subsampling methods. Econometrica 77, 721–762.
Andrews, D.W.K., Guggenberger, P., 2010a. Asymptotic size and a problem with subsampling and with the m out of n bootstrap. Econometric Theory 26, 426–468.
Andrews, D.W.K., Guggenberger, P., 2010b. Asymptotics for LS, GLS, and feasible GLS statistics in an AR(1) model with conditional heteroskedasticity. Cowles Foundation Discussion Paper No. 1665RR, Yale Univ. Available at: http://cowles.econ.yale.edu.
Bobkowski, M.J., 1983. Hypothesis testing in nonstationary time series. Unpublished Ph.D. Thesis, University of Wisconsin, Madison.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.
Cavanagh, C., 1985. Roots local to unity. Unpublished Manuscript, Harvard University.
Chan, N.H., Wei, C.Z., 1987. Asymptotic inference for nearly nonstationary AR(1) processes. Annals of Statistics 15, 1050–1063.
Chen, W.W., Deo, R.S., 2011. The restricted likelihood ratio test for autoregressive processes. Journal of Time Series Analysis. doi:10.1111/j.1467-9892.2011.00769.x.
Doukhan, P., 1994. Mixing Properties and Examples. In: Lecture Notes in Statistics, vol. 85. Springer, New York.
Elliott, G., 1999. Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. International Economic Review 40, 767–783.
Elliott, G., Stock, J.H., 2001. Confidence intervals for autoregressive coefficients near one. Journal of Econometrics 103, 155–181.
Giraitis, L., Phillips, P.C.B., 2006. Uniform limit theory for stationary autoregression. Journal of Time Series Analysis 27, 51–60. Corrigendum 27, pp. I–II.
Gonçalves, S., Kilian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. Journal of Econometrics 123, 89–120.
Gonçalves, S., Kilian, L., 2007. Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity. Econometric Reviews 26, 609–641.
Guo, B., Phillips, P.C.B., 2001. Testing for autocorrelation and unit roots in the presence of conditional heteroskedasticity of unknown form. Unpublished Working Paper, Department of Economics, UC Santa Cruz.
Hansen, B.E., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.
Hansen, B.E., 1999. The grid bootstrap and the autoregressive model. Review of Economics and Statistics 81, 594–607.
Lindner, A.M., 2009. Stationarity, mixing, distributional properties and moments of GARCH(p, q)-processes. In: Andersen, T.G., Davis, R.A., Kreiß, J.-P., Mikosch, Th. (Eds.), Handbook of Financial Time Series. Springer, Berlin, pp. 481–496.
Mikusheva, A., 2007. Uniform inferences in autoregressive models. Econometrica 75, 1411–1452. Corrigendum Econometrica 78, 1773.
Müller, U.K., Elliott, G., 2003. Tests for unit roots and the initial condition. Econometrica 71, 1269–1286.
Nankervis, J.C., Savin, N.E., 1996. The level and power of the bootstrap t-test in the trend model with AR(1) errors. Journal of Business and Economic Statistics 14, 161–168.
Park, J.Y., 2002. Weak unit roots. Unpublished Working Paper, Department of Economics, Rice University.
Phillips, P.C.B., 1977. Approximations to some finite sample distributions associated with a first-order stochastic difference equation. Econometrica 45, 463–485.
Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74, 535–547.
Phillips, P.C.B., Magdalinos, T., 2007. Limit theory for moderate deviations from a unit root. Journal of Econometrics 136, 115–130.
Seo, B., 1999. Distribution theory for unit root tests with conditional heteroskedasticity. Journal of Econometrics 91, 113–144.
Stock, J.H., 1991. Confidence intervals for the largest autoregressive root in US macroeconomic time series. Journal of Monetary Economics 28, 435–459.