Bootstrapping autoregressions with conditional heteroskedasticity of unknown form

Sílvia Gonçalves[a,*,1], Lutz Kilian[b,c]

[a]*CIRANO, CIREQ and Département de Sciences Économiques, Université de Montréal, C.P. 6128, succ. Centre-Ville, Montréal, Canada QC H3C 3J7*
[b]*CEPR and Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220, USA*
[c]*Directorate General Research, European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany*

Abstract

Conditional heteroskedasticity is an important feature of many macroeconomic and financial time series. Standard residual-based bootstrap procedures for dynamic regression models treat the regression error as i.i.d. These procedures are invalid in the presence of conditional heteroskedasticity. We establish the asymptotic validity of three easy-to-implement alternative bootstrap proposals for stationary autoregressive processes with martingale difference errors subject to possible conditional heteroskedasticity of unknown form. These proposals are the fixed-design wild bootstrap, the recursive-design wild bootstrap and the pairwise bootstrap. In a simulation study all three procedures tend to be more accurate in small samples than the conventional large-sample approximation based on robust standard errors. In contrast, standard residual-based bootstrap methods for models with i.i.d. errors may be very inaccurate if the i.i.d. assumption is violated. We conclude that in many empirical applications the proposed robust bootstrap procedures should routinely replace conventional bootstrap procedures for autoregressions based on the i.i.d. error assumption.
© 2003 Elsevier B.V. All rights reserved.

_____

* Corresponding author. Département de Sciences Économiques, Université de Montréal, CP 6128, succ. Centre-Ville, Montréal Canada QC H3C 3J7.

1. Introduction

There is evidence of conditional heteroskedasticity in the residuals of many estimated dynamic regression models in finance and in macroeconomics (see, e.g., Engle, 1982; Bollerslev, 1986; Weiss, 1988). This evidence is particularly strong for regressions involving monthly, weekly and daily data. Standard residual-based bootstrap methods of inference for autoregressions treat the error term as independent and identically distributed (i.i.d.) and are invalidated by conditional heteroskedasticity. In this paper, we analyze two main proposals for dealing with conditional heteroskedasticity of unknown form in autoregressions.

The first proposal is very easy to implement. It involves an application of the wild bootstrap (WB) to the residuals of the dynamic regression model. The WB method allows for regression errors that follow martingale difference sequences (m.d.s.) with possible conditional heteroskedasticity. We investigate both the fixed-design and the recursive-design implementation of the WB for autoregressions. We prove their first-order asymptotic validity for the autoregressive parameters (and smooth functions thereof) under fairly general conditions including, for example, stationary ARCH, GARCH and stochastic volatility error processes (see, e.g., Bollerslev, 1986; Shephard, 1996).

There are several fundamental differences between this paper and earlier work on the WB in regression models. First, existing theoretical work has largely focused on providing first- and second-order theoretical justification for the wild bootstrap in the classical linear regression model (see, e.g., Wu, 1986; Liu, 1988; Mammen, 1993; Davidson and Flachaire, 2001). Second, the previous literature has mainly focused on the problem of unconditional heteroskedasticity in cross-sections, whereas we focus on the problem of conditional heteroskedasticity in time series. Third, much of the earlier work has dealt with models restricted under the null hypothesis of a test, whereas we focus on the construction of bootstrap confidence intervals from unrestricted regression models (see Davidson and Flachaire, 2001; Godfrey and Orme, 2001).

The work most closely related to ours is Kreiss (1997). Kreiss established the asymptotic validity of a fixed-design WB for stationary autoregressions with known finite lag order when the error term exhibits a specific form of conditional heteroskedasticity. We provide a generalization of this result to m.d.s. errors with possible conditional heteroskedasticity of unknown form. Our results cover as special cases the N-GARCH, *t*-GARCH and asymmetric GARCH models, as well as stochastic volatility models. Kreiss (1997) also proposed a recursive-design WB, under the name of "modified wild bootstrap", but he did not establish the consistency of this bootstrap proposal for autoregressive processes with conditional heteroskedasticity. We prove the first-order asymptotic validity of the recursive-design WB for finite-order autoregressions with m.d.s. errors subject to possible conditional heteroskedasticity of unknown form. The proof holds under slightly stronger assumptions than the proof for the fixed-design WB.

Tentative simulation evidence shows that the recursive-design WB scheme works well in practice for a wide range of models of conditional heteroskedasticity. In contrast, conventional residual-based resampling schemes for autoregressions based on the i.i.d. error assumption may be very inaccurate in the presence of conditional

heteroskedasticity. Moreover, the accuracy of the recursive-design WB method is comparable to that of the recursive-design i.i.d. bootstrap when the true errors are i.i.d. The recursive-design WB method is typically more accurate in small samples than the fixed-design WB method. It also tends to be more accurate than the Gaussian large-sample approximation based on robust standard errors.

The second proposal for dealing with conditional heteroskedasticity of unknown form involves the pairwise resampling of the observations. This method was originally suggested by Freedman (1981) for cross-sectional models. We establish the asymptotic validity of this method in the autoregressive context and compare its performance to that of the fixed-design and of the recursive-design WB. The pairwise bootstrap is less efficient than the residual-based WB, but—like the fixed-design WB—it remains valid for a broader range of GARCH processes than the recursive-design WB, including EGARCH, AGARCH and GJR-GARCH processes, which have been proposed specifically to capture asymmetric responses to shocks in asset returns (see, e.g., Engle and Ng (1993) for a review). We find in Monte Carlo simulations that the pairwise bootstrap is typically more accurate than the fixed-design WB method, but in small samples tends to be somewhat less accurate than the recursive-design WB when the data are persistent. For large samples these differences vanish, and the pairwise bootstrap is as accurate as the recursive-design WB.

A third proposal for dealing with conditional heteroskedasticity of unknown form is the resampling of blocks of autoregressive residuals (see, e.g., Berkowitz et al., 2000). No formal theoretical results exist that would justify such a bootstrap proposal. We do not consider this proposal for two reasons. First, in the context of a well-specified parametric model this proposal involves a loss of efficiency relative to the WB because it allows for serial correlation in the error term in addition to conditional heteroskedasticity. Second, the residual-based block bootstrap requires the choice of an additional tuning parameter in the form of the block size. In practice, results may be sensitive to the choice of block size. Although there are data-dependent rules for block size selection, these procedures are very computationally intensive and little is known about their accuracy in small samples. In contrast, the methods we propose are no more computationally burdensome than the standard residual-based algorithm and very easy to implement.

The paper is organized as follows. In Section 2 we provide empirical evidence that casts doubt on the use of the i.i.d. error assumption for autoregressions, and we highlight the limitations of existing bootstrap and asymptotic methods of inference when the autoregressive errors are conditionally heteroskedastic. In Section 3 we describe the bootstrap algorithms and state our main theoretical results. Details of the proofs are relegated to the appendix. In Section 4, we provide some tentative simulation evidence for the small-sample performance of alternative bootstrap proposals. We conclude in Section 5.

2. Evidence against the assumption of i.i.d. errors

Standard residual-based bootstrap methods of inference for dynamic regression models treat the error term as i.i.d. The i.i.d. assumption does not follow naturally from

Table 1
Approximate finite-sample p-values of the Engle (1982) LM test of the No-ARCH(q) hypothesis (in percent) for monthly autoregressions

q	1	2	3	4	5
Industrial output growth	1.58	2.40	3.28	1.61	1.47
M1 growth	0.00	0.01	0.01	0.02	0.01
CPI inflation	0.50	1.13	1.79	2.35	2.05
Real T-bill rate	0.08	0.18	0.29	0.37	0.34
Federal funds rate	3.37	0.45	0.71	0.94	0.90
Percent change in oil price	2.39	3.77	5.25	4.60	6.44

Source: Based on 20 000 bootstrap replications under i.i.d. error null hypothesis. All data have been filtered by a univariate AR model, the lag order of which has been selected by the AIC subject to an upper bound of 12 lags.

economic models. Nevertheless, in many cases it has proved convenient for theoretical purposes to treat the error term of dynamic regression models as i.i.d. This would be of little concern if actual data were well represented by models with i.i.d. errors. Unfortunately, this is not the case in many empirical studies. One approach in applied work has been simply to ignore the problem and to treat the error term as i.i.d. (see, e.g., Goetzmann and Jorion, 1993, 1995). An alternative approach has been to impose a parametric model of conditional heteroskedasticity. For example, Bollerslev (1986) models inflation as an autoregressive process with GARCH(1,1) errors. Similarly, Hodrick (1992) and Bekaert and Hodrick (2001) postulate a VAR model with conditionally Gaussian GARCH(1,1) errors. This approach is not without risks. First, it is not clear whether the class of GARCH models adequately captures the conditional heteroskedasticity in the data. Second, even when the class of GARCH models is appropriate, in practice, the precise form of the GARCH model will be unknown and different specifications may yield different results (see Wolf, 2000). Further difficulties arise in the multivariate case. For multivariate GARCH models it is often difficult to obtain reliable numerical estimates of the GARCH parameters. In response, researchers typically impose ad hoc restrictions on the covariance structure of the model (see, e.g., Bollerslev et al., 1988; Bollerslev, 1990; Bekaert et al., 1997) that call into question the theoretical validity of the estimates (see Ledoit et al., 2001). For these reasons, we argue for a nonparametric treatment of conditional heteroskedasticity in dynamic regression models.

Whereas the failure of the i.i.d. assumption is well-documented in empirical finance, it is less well known that many monthly macroeconomic variables also exhibit evidence of conditional heteroskedasticity. In fact, both the ARCH and the GARCH model were originally motivated by macroeconometric applications (see Engle, 1982; Bollerslev, 1986). The workhorse model of empirical macroeconomics is the linear autoregression. Table 1 illustrates that the errors of monthly autoregressions typically cannot be treated as i.i.d. It shows the results of LM tests of the null of no ARCH in the errors of six univariate monthly autoregressive models (see Engle, 1982). The data are the growth rate of U.S. industrial output, M1 growth, CPI inflation, the real 3-month T-Bill rate, the

nominal Federal Funds rate and the percent change in the price of oil. The data source is FRED, the sample period 1959.1–2001.8, and the autoregressive lag orders have been selected by the AIC. The LM tests strongly reject the assumption of conditional homoskedasticity for the errors of the AR models. Similar results are obtained for a fixed number of 12 lags or of 24 lags.

The evidence of non-i.i.d. errors in Table 1 is important because many methods of inference developed for smooth functions of autoregressive parameters (such as impulse responses) do not allow for conditional heteroskedasticity. For example, standard residual-based bootstrap methods for autoregressions rely on the i.i.d. error assumption and are invalid in the presence of conditional heteroskedasticity, as we will show in the next section. Similarly, the grid bootstrap of Hansen (1999) is based on the assumption of an autoregression with i.i.d. errors. Likewise, standard asymptotic methods for inference in autoregressions rely if not on the i.i.d. assumption, then on the assumption of conditional homoskedasticity. For example, the closed-form solutions for the asymptotic normal approximation of impulse response distributions proposed by Lütkepohl (1990) are based on the assumption of conditional homoskedasticity and hence will be inconsistent in the presence of conditional heteroskedasticity.

In this paper we study several easy-to-implement bootstrap methods that allow inference in autoregressions with possible conditional heteroskedasticity of unknown form. Unlike the standard residual-based bootstrap for models with i.i.d. innovations these bootstrap methods remain valid under the much weaker assumption of m.d.s. innovations, and they do not require the researcher to take a stand on the existence or specific form of conditional heteroskedasticity. For expository purposes we focus on univariate autoregressive models. Analogous results for the multivariate case are possible at the cost of additional notation.

3. Theory

Let $(\Omega, \mathscr{F}, P)$ be a probability space and $\{\mathscr{F}_t\}$ a sequence of increasing σ-fields of $\mathscr{F}$. The sequence of martingale differences $\{\varepsilon_t, \ t \in \mathbb{Z}\}$ is defined on $(\Omega, \mathscr{F}, P)$, where each ε_t is assumed to be measurable with respect to $\mathscr{F}_t$. We observe a sample of data $\{y_{-p+1}, \ldots, y_0, y_1, \ldots, y_n\}$ from the following data generating process (DGP) for the time series y_t,

$$\phi(L)y_t = \varepsilon_t, \tag{3.1}$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$, $\phi_p \neq 0$, is assumed to have all roots outside the unit circle and the lag order p is finite and known. $\phi = (\phi_1, \ldots, \phi_p)'$ is the parameter of interest, which we estimate by ordinary least squares (OLS) using observations 1 through n:

$$\hat{\phi} = \left(n^{-1} \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} \right)^{-1} n^{-1} \sum_{t=1}^{n} Y_{t-1} y_t,$$

where $Y_{t-1} = (y_{t-1}, \ldots, y_{t-p})'$. In this paper we focus on bootstrap confidence intervals for ϕ that are robust to the presence of conditional heteroskedasticity of unknown form in the innovations $\{\varepsilon_t\}$. More specifically, we assume the following condition:

Assumption A.

   (i) $E(\varepsilon_t | \mathscr{F}_{t-1}) = 0$, almost surely, where $\mathscr{F}_{t-1} = \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$ is the σ-field generated by $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$.

   (ii) $E(\varepsilon_t^2) = \sigma^2 < \infty$.

   (iii) $\lim_{n \to \infty} n^{-1} \sum_{t=1}^n E(\varepsilon_t^2 | \mathscr{F}_{t-1}) = \sigma^2 > 0$ in probability.

   (iv) $\tau_{r,s} \equiv \sigma^{-4} E(\varepsilon_t^2 \varepsilon_{t-r} \varepsilon_{t-s})$ is uniformly bounded for all $t, r \geqslant 1$, $s \geqslant 1$; $\tau_{r,r} > 0$ for all r.

   (v) $\lim_{n \to \infty} n^{-1} \sum_{t=1}^n \varepsilon_{t-r} \varepsilon_{t-s} E(\varepsilon_t^2 | \mathscr{F}_{t-1}) = \sigma^4 \tau_{r,s}$ in probability for any $r \geqslant 1$, $s \geqslant 1$.

   (vi) $E|\varepsilon_t|^{4r}$ is uniformly bounded, for some $r > 1$.

Assumption A replaces the usual i.i.d. assumption on the errors $\{\varepsilon_t\}$ by the less restrictive martingale difference sequence assumption. In particular, Assumption A allows for dependent, but uncorrelated errors. It does not impose conditional homoskedasticity on the sequence $\{\varepsilon_t\}$, although it requires $\{\varepsilon_t\}$ to be covariance stationary. Assumption A covers a variety of conditionally heteroskedastic models such as ARCH, GARCH, EGARCH and stochastic volatility models (see, e.g. Deo (2000), who shows that a stronger version of Assumption A is satisfied for stochastic volatility and GARCH models). Assumptions (iv) and (v) restrict the fourth-order cumulants of ε_t.

   Recently, Kuersteiner (2001) derived the asymptotic distribution of efficient instrumental variables estimators in the context of ARMA models with martingale difference errors that are strictly stationary and ergodic, and that satisfy a summability condition on the fourth-order cumulants. His result also applies to the OLS estimator in the AR model as a special case. In Theorem 3.1, we provide an alternative derivation of the asymptotic distribution of the OLS estimator of the AR model under the slightly less restrictive Assumption A. We use Kuersteiner's (2001) notation to characterize the asymptotic covariance matrix of $\hat{\phi}$. Using $\phi^{-1}(L) = \sum_{j=0}^\infty \psi_j L^j$, we let $b_j = (\psi_{j-1}, \ldots, \psi_{j-p})'$ with $\psi_0 = 1$ and $\psi_j = 0$ for $j < 0$. The coefficients ψ_j satisfy the recursion $\psi_s - \phi_1 \psi_{s-1} - \cdots - \phi_p \psi_{s-p} = 0$ for all $s > 0$ and $\psi_0 = 1$. We let $\Rightarrow$ denote convergence in distribution throughout.

Theorem 3.1. *Under Assumption* A, $\sqrt{n}(\hat{\phi} - \phi) \Rightarrow N(0, C)$, *where*

$$C = A^{-1} B A^{-1},$$

$$A = \sigma^2 \sum_{j=1}^\infty b_j b_j' \quad and \quad B = \sigma^4 \sum_{i=1}^\infty \sum_{j=1}^\infty b_i b_j' \tau_{i,j}.$$

   The asymptotic covariance matrix of $\hat{\phi}$ is of the traditional "sandwich" form, where $A = E(n^{-1} \sum_{t=1}^n Y_{t-1} Y_{t-1}')$ and $B = Var(n^{-1/2} \sum_{t=1}^n Y_{t-1} \varepsilon_t)$. Under conditional homoskedasticity, $B = \sigma^2 A$. In particular, by application of the law of iterated expectations, we have that $\tau_{i,i} \equiv \sigma^{-4} E(\varepsilon_t^2 \varepsilon_{t-i}^2) = \sigma^{-4} E(\varepsilon_{t-i}^2 E(\varepsilon_t^2 | \mathscr{F}_{t-1})) = \sigma^{-4} E(\varepsilon_{t-i}^2 \sigma^2) =$

1 for all $i = 1, 2, \ldots$. Similarly, we can show that $\tau_{i,j} = 0$ for all $i \neq j$. Thus, for instance in the AR(1) case, the asymptotic variance of $\hat{\phi} = \hat{\phi}_1$ simplifies to $C = (\sigma^2 \sum_{i=0}^{\infty} \psi_i^2)^{-2} (\sigma^4 \sum_{i=0}^{\infty} \psi_i^2) = 1 - \phi_1^2$.

The validity of any bootstrap method in the context of autoregressions with conditional heteroskedasticity depends crucially on the ability of the bootstrap to allow consistent estimation of the asymptotic covariance matrix C. The standard residual-based bootstrap method fails to do so by not correctly mimicking the behavior of the fourth-order cumulants of ε_t in the conditionally heteroskedastic case, as we now show. Let $\hat{\varepsilon}_t^*$ be resampled with replacement from the centered residuals. The standard residual-based bootstrap builds y_t^* recursively from $\hat{\varepsilon}_t^*$ according to

$$y_t^* = Y_{t-1}^{*\prime} \, \hat{\phi} + \hat{\varepsilon}_t^*, \quad t = 1, \ldots, n,$$

where $Y_{t-1}^* = (y_{t-1}^*, \ldots, y_{t-p}^*)'$, given appropriate initial conditions. The recursive-design i.i.d. bootstrap analogues of A and B are $A_{\mathrm{riid}}^* = n^{-1} \sum_{t=1}^{n} \mathrm{E}^*(Y_{t-1}^* Y_{t-1}^{*\prime})$ and $B_{\mathrm{riid}}^* = Var^*(n^{-1/2} \sum_{t=1}^{n} Y_{t-1}^* \hat{\varepsilon}_t^*)$, respectively. Because $\hat{\varepsilon}_t^*$ is i.i.d. $(0, \hat{\sigma}^2)$, where $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^{n} (\hat{\varepsilon}_t - \bar{\hat{\varepsilon}})^2$, $\hat{\varepsilon}_t^*$ and Y_{t-1}^* are (conditionally) independent, and

$$B_{\mathrm{riid}}^* = n^{-1} \sum_{t=1}^{n} \mathrm{E}^*(Y_{t-1}^* Y_{t-1}^{*\prime} \, \hat{\varepsilon}_t^{*2}) = n^{-1} \sum_{t=1}^{n} \mathrm{E}^*(Y_{t-1}^* Y_{t-1}^{*\prime}) \, \mathrm{E}^*(\hat{\varepsilon}_t^{*2}) = \hat{\sigma}^2 A_{\mathrm{riid}}^*.$$

Thus, the bootstrap analogue of C, $C_{\mathrm{riid}}^* \equiv A_{\mathrm{riid}}^{*-1} B_{\mathrm{riid}}^* A_{\mathrm{riid}}^{*-1} = \hat{\sigma}^2 A_{\mathrm{riid}}^{*-1}$, converges in probability to $\sigma^2 A^{-1}$, implying that the limiting distribution of the recursive i.i.d. bootstrap is $\mathrm{N}(0, \sigma^2 A^{-1})$. As Theorem 3.1 shows, $\sigma^2 A^{-1}$, however, is *not* the correct asymptotic covariance matrix of $\hat{\phi}$ without imposing further conditions, e.g., that ε_t is conditionally homoskedastic. In the general, conditionally heteroskedastic case, B depends on $\sigma^4 \tau_{i,j}$. The recursive-design i.i.d. bootstrap implies $\mathrm{E}^*(\hat{\varepsilon}_{t-i}^* \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^{*2}) = \hat{\sigma}^4$ when $i = j$ and zero otherwise, and thus implicitly sets $\tau_{i,j} = 1$ for $i = j$ and $\tau_{i,j} = 0$ for $i \neq j$.

Given the failure of the standard-residual based bootstrap, we are interested in establishing the first-order asymptotic validity of three alternative bootstrap methods in this environment. Two of the bootstrap methods we study rely on an application of the wild bootstrap (WB). The WB has been originally developed by Wu (1986), Liu (1988) and Mammen (1993) in the context of static linear regression models with (unconditionally) heteroskedastic errors. We consider both a recursive-design and a fixed-design version of the WB. The third method is a natural generalization of the pairwise bootstrap for linear regression first suggested by Freedman (1981) for cross-sectional data.

3.1. Recursive-design wild bootstrap

The recursive-design WB is a simple modification of the usual recursive-design bootstrap method for autoregressions (see, e.g., Bose, 1988) which consists of replacing Efron's i.i.d. bootstrap by the wild bootstrap when bootstrapping the errors of the AR model. More specifically, the recursive-design WB bootstrap generates a pseudo time series $\{y_t^*\}_{t=1}^{n}$ according to the autoregressive process:

$$y_t^* = Y_{t-1}^{*\prime} \, \hat{\phi} + \hat{\varepsilon}_t^*, \quad t = 1, \ldots, n,$$

where $\hat{\varepsilon}_t^* = \hat{\varepsilon}_t \eta_t$, with $\hat{\varepsilon}_t = \hat{\phi}(L)y_t$, and where η_t is an i.i.d. sequence with mean zero and variance one such that $E^*|\eta_t|^4 \leqslant \Delta < \infty$. We let $y_t^* = 0$ for all $t \leqslant 0$. Kreiss (1997) suggested this method in the context of autoregressive models with i.i.d. errors, but did not investigate its theoretical justification in more general models. Here, we will provide conditions for the asymptotic validity of the recursive-design WB proposal for finite-order autoregressive processes with possibly conditionally heteroskedastic errors.

Establishing the validity of the recursive-design WB requires a strengthening of Assumption A. Specifically, we need Assumption A′ below in order to ensure convergence of the bootstrap estimator of the asymptotic covariance matrix C to its correct limit. In contrast, the fixed-design WB and the pairwise bootstrap to be discussed later are valid under the less restrictive Assumption A.

Assumption A′.
 (iv′) $E(\varepsilon_t^2 \varepsilon_{t-r} \varepsilon_{t-s}) = 0$ for all $r \neq s$, for all t, $r \geqslant 1$, $s \geqslant 1$.
 (vi′) $E|\varepsilon_t|^{4r}$ is uniformly bounded for some $r \geqslant 2$ and for all t.

Assumption A′ restricts the class of conditionally heteroskedastic autoregressive models in two dimensions. First, Assumption A′(iv′) requires $\tau_{r,s} = 0$ for all $r \neq s$. Milhøj (1985) shows that this assumption is satisfied for the ARCH(p) model with innovations having a symmetric distribution. Bollerslev (1986) and He and Teräsvirta (1999) extend the argument to the GARCH(p,q) case. In addition, Deo (2000) shows that this assumption is satisfied by certain stochastic volatility models. Assumption A′(iv′) excludes some non-symmetric parametric models such as asymmetric EGARCH. Second, we now require the existence of at least eight moments for the martingale difference sequence $\{\varepsilon_t\}$ as opposed to only $4r$ moments, for some $r > 1$, as in Assumption A. A similar moment condition was used by Kreiss (1997) in his Theorem 4.3, which shows the validity of the recursive-design WB for possibly infinite-order AR processes with i.i.d. innovations.

The strengthening of Assumption A is crucial to showing the asymptotic validity of the recursive-design WB in the martingale difference context. In particular, conditional on the data, and given the independence of $\{\eta_t\}$, $\{Y_{t-1}^* \hat{\varepsilon}_t^*, \mathscr{F}_t^*\}$ can be shown to be a vector m.d.s., where $\mathscr{F}_t^* = \sigma(\eta_t, \eta_{t-1}, \ldots, \eta_1)$. We use Assumption A′(vi′) to ensure convergence of $n^{-1} \sum_{t=1}^{n} Y_{t-1}^* Y_{t-1}^{*\prime} \hat{\varepsilon}_t^{*2}$ to $B_{\text{rwb}}^* \equiv Var^*(n^{-1/2} \sum_{t=1}^{n} Y_{t-1}^* \hat{\varepsilon}_t^*)$, thus verifying one of the conditions of the CLT for m.d.s. Assumption A′(iv′) ensures convergence of the recursive-design WB variance B_{rwb}^* to the correct limiting variance of $n^{-1/2} \sum_{t=1}^{n} Y_{t-1} \varepsilon_t$. More specifically, letting $Y_{t-1}^* \equiv \sum_{j=1}^{t-1} \hat{b}_j \hat{\varepsilon}_{t-j}^*$ with $\hat{b}_j \equiv (\hat{\psi}_{j-1}, \ldots, \hat{\psi}_{j-p})'$, $\hat{\psi}_0 = 1$ and $\hat{\psi}_j = 0$ for $j < 0$, it follows by direct evaluation that

$$B_{\text{rwb}}^* = n^{-1} \sum_{t=1}^{n} \sum_{j=1}^{t-1} \sum_{i=1}^{t-1} \hat{b}_j \hat{b}_i' E^*(\hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_{t-i}^* \hat{\varepsilon}_t^{*2}),$$

where $E^*(\hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_{t-i}^* \hat{\varepsilon}_t^{*2}) = \hat{\varepsilon}_{t-i}^2 \hat{\varepsilon}_t^2$ for $i = j$ and zero otherwise. We can rewrite B_{rwb}^* as $\sum_{j=1}^{n-1} \hat{b}_j \hat{b}_j' n^{-1} \sum_{t=1+j}^{n} \hat{\varepsilon}_t^2 \hat{\varepsilon}_{t-j}^2$, which converges in probability to $\tilde{B} \equiv \sum_{j=1}^{\infty} b_j b_j' \sigma^4 \tau_{jj}$ under Assumption A. Without Assumption A′(iv′) an asymptotic bias term appears in the estimation of $B \equiv \sigma^4 \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} b_i b_j' \tau_{i,j}$, which is equal to $-\sigma^4 \sum_{i \neq j} b_i b_j' \tau_{i,j}$.

Assumption $A'(iv')$ sets $\tau_{i,j}$ equal to zero for $i \neq j$, and thus ensures that the recursive-design WB consistently estimates B.

Theorem 3.2 formally establishes the asymptotic validity of the recursive-design WB for finite-order autoregressions with conditionally heteroskedastic errors. Let $\hat{\phi}^*_{\text{rwb}}$ denote the recursive-design WB OLS estimator, i.e., $\hat{\phi}^*_{\text{rwb}} = (n^{-1} \sum_{t=1}^n Y^*_{t-1} Y^{*\prime}_{t-1})^{-1} n^{-1} \sum_{t=1}^n Y^*_{t-1} y^*_t$.

Theorem 3.2. *Under Assumption* A *strengthened by Assumption* $A'(iv')$ *and* (vi'), *it follows that*

$$\sup_{x \in \mathbb{R}^p} |P^*(\sqrt{n}(\hat{\phi}^*_{\text{rwb}} - \hat{\phi}) \leqslant x) - P(\sqrt{n}(\hat{\phi} - \phi) \leqslant x)| \overset{P}{\to} 0,$$

where P^* *denotes the probability measure induced by the recursive-design WB.*

3.2. Fixed-design wild bootstrap

The fixed-design WB generates $\{y^*_t\}_{t=1}^n$ according to the equation

$$y^*_t = Y'_{t-1} \hat{\phi} + \hat{\varepsilon}^*_t, \quad t = 1, \ldots, n, \tag{3.2}$$

where $\hat{\varepsilon}^*_t = \hat{\varepsilon}_t \eta_t$, $\hat{\varepsilon}_t = \hat{\phi}(L) y_t$, and where η_t is an i.i.d. sequence with mean zero and variance one such that $E^* |\eta_t|^{2r} \leqslant \Delta < \infty$, for some $r > 1$. The fixed-design WB estimator is $\hat{\phi}^*_{\text{fwb}} = (n^{-1} \sum_{t=1}^n Y_{t-1} Y'_{t-1})^{-1} n^{-1} \sum_{t=1}^n Y_{t-1} y^*_t$. The fixed-design WB corresponds to a regression-type bootstrap method in that (3.2) is a fixed-design regression model, conditional on the original sample. A similar "fixed-regressor bootstrap" has also been proposed by Hansen (2000) in the context of testing for structural change in regression models. The fixed-design WB was originally suggested by Kreiss (1997). Kreiss' (1997) Theorem 4.2 proves the first-order asymptotic validity of the fixed-design WB for finite-order autoregressions with conditional heteroskedasticity of a specific form. More specifically, he postulates a DGP of the form $y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sigma(y_{t-1}) v_t$, where v_t is i.i.d. $(0,1)$ with finite fourth moment. The i.i.d. assumption on the rescaled innovations v_t is violated if for instance the conditional moments of v_t depend on past observations. We prove the first-order asymptotic validity of the fixed-design WB of Kreiss (1997) under a broader set of regularity conditions, namely Assumption A.

Theorem 3.3. *Under Assumption* A,

$$\sup_{x \in \mathbb{R}^p} \left| P^*(\sqrt{n}(\hat{\phi}^*_{\text{fwb}} - \hat{\phi}) \leqslant x) - P(\sqrt{n}(\hat{\phi} - \phi) \leqslant x) \right| \overset{P}{\to} 0,$$

where P^* *denotes the probability measure induced by the fixed-design WB.*

In contrast to the recursive-design WB, the ability of the fixed-design WB to estimate consistently the variance, and hence the limiting distribution, of $\hat{\phi}$ does not require a strengthening of Assumption A. Specifically, the variance of the limiting conditional bootstrap distribution of $\hat{\phi}^*_{\text{fwb}}$ is given by $A^{*-1}_{\text{fwb}} B^*_{\text{fwb}} A^{*-1}_{\text{fwb}}$, where $A^*_{\text{fwb}} = n^{-1} \sum_{t=1}^n Y_{t-1} Y'_{t-1}$ and $B^*_{\text{fwb}} \equiv Var^*(n^{-1/2} \sum_{t=1}^n Y_{t-1} \hat{\varepsilon}^*_t) = n^{-1} \sum_{t=1}^n Y_{t-1} Y'_{t-1} \hat{\varepsilon}^2_t$.

Under Assumption A one can show that $A^*_{\text{fwb}} \overset{P}{\to} A$ and $B^*_{\text{fwb}} \overset{P}{\to} B$, thus ensuring that $A^{*-1}_{\text{fwb}} B^*_{\text{fwb}} A^{*-1}_{\text{fwb}} \overset{P}{\to} A^{-1} B A^{-1} \equiv C$.

3.3. Pairwise bootstrap

Another bootstrap method that captures the presence of conditional heteroskedasticity in autoregressive models consists of bootstrapping "pairs", or tuples, of the dependent and the explanatory variables in the autoregression. This method is an extension of Freedman's (1981) bootstrap method for the correlation model to the autoregressive context. In the AR(p) model, it amounts to resampling with replacement from the set of tuples $(y_t, Y'_{t-1}) = (y_t, y_{t-1}, \dots, y_{t-p})$, $t = 1, \dots, n$. Let $\{(y^*_t, Y^{*\prime}_{t-1}) = (y^*_t, y^*_{t-1}, \dots, y^*_{t-p}), \ t = 1, \dots, n\}$ be an i.i.d. resample from this set. Then the pairwise bootstrap estimator is defined by $\hat{\phi}^*_{\text{pb}} = (n^{-1} \sum_{t=1}^n Y^*_{t-1} Y^{*\prime}_{t-1})^{-1} n^{-1} \sum_{t=1}^n Y^*_{t-1} y^*_t$. The bootstrap analogue of ϕ is $\hat{\phi}$, since $\hat{\phi}$ is the parameter value that minimizes $E^*[(y^*_t - Y^{*\prime}_{t-1} \phi)^2]$. The following theorem establishes the asymptotic validity of the pairwise bootstrap for the AR(p) process with m.d.s. errors satisfying Assumption A.

Theorem 3.4. *Under Assumption* A, *it follows that*

$$\sup_{x \in \mathbb{R}^p} |P^*(\sqrt{n}(\hat{\phi}^*_{\text{pb}} - \hat{\phi}) \leqslant x) - P(\sqrt{n}(\hat{\phi} - \phi) \leqslant x)| \overset{P}{\to} 0,$$

where P^* denotes the probability measure induced by the pairwise bootstrap.

3.4. Asymptotic validity of bootstrapping the studentized slope parameter

Corollary 3.1 below establishes the asymptotic validity of bootstrapping the *t-statistic* for the elements of ϕ. To conserve space, we let $\hat{\phi}^*$ denote the OLS estimator of ϕ obtained under any of the three robust bootstrap resampling schemes studied above. Similarly, we use $(y^*_t, Y^{*\prime}_{t-1})$ to denote bootstrap data in general. In particular, we implicitly set $Y^*_{t-1} = Y_{t-1}$ for the fixed-design WB.

For a typical element ϕ_j a bootstrap percentile-t confidence interval is based on $t_{\hat{\phi}^*_j} = \sqrt{n}(\hat{\phi}^*_j - \hat{\phi}_j)/\sqrt{\hat{C}^*_{jj}}$, the bootstrap analogue of the t-statistic $t_{\hat{\phi}_j} = \sqrt{n}(\hat{\phi}_j - \phi_j)/\sqrt{\hat{C}_{jj}}$. In the context of (conditional) heteroskedasticity, $\hat{C}_{jj}$ and $\hat{C}^*_{jj}$ are the heteroskedasticity-consistent variance estimators evaluated on the original and on the bootstrap data, respectively. Specifically, for the bootstrap t-statistic let

$$\hat{C}^* = \hat{A}^{*-1} \hat{B}^* \hat{A}^{*-1}, \quad \text{with}$$

$$\hat{A}^* = n^{-1} \sum_{t=1}^n Y^*_{t-1} Y^{*\prime}_{t-1} \quad \text{and} \quad \hat{B}^* = n^{-1} \sum_{t=1}^n Y^*_{t-1} Y^{*\prime}_{t-1} \ \tilde{\varepsilon}^{*2}_t,$$

where $\tilde{\varepsilon}^*_t = y^*_t - \hat{\phi}^{*\prime} Y^*_{t-1}$ are the bootstrap residuals.

Corollary 3.1. *Assume Assumption* A *holds. Then, for the fixed-design WB and the pairwise bootstrap, it follows that*

$$\sup_{x \in \mathbb{R}} \left| P^*(t_{\hat{\phi}_j^*} \leqslant x) - P(t_{\hat{\phi}_j} \leqslant x) \right| \xrightarrow{P} 0, \quad j = 1, \ldots, p.$$

If Assumption A *is strengthened by Assumption* A$'$(iv$'$) *and* (vi$'$), *then the above result also holds for the recursive-design WB.*

4. Simulation evidence

In this section, we study the accuracy of the bootstrap approximation proposed in Section 3 for sample sizes of interest in applied work. We focus on the AR(1) model as the leading example of an autoregressive process. The DGP is $y_t = \phi_1 y_{t-1} + \varepsilon_t$ with $\phi_1 \in \{0, 0.9\}$. In our simulation study we allow for GARCH(1,1) errors of the form $\varepsilon_t = \sqrt{h_t} v_t$, where v_t is i.i.d. N(0,1) and $h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}$, $t = 1, \ldots, n$. We normalize the unconditional variance of ε_t to one. In addition to conditional N(0,1) innovations we also consider GARCH models with conditional t_5-errors (suitably normalized to have unit variance). For $\beta = 0$ this model reduces to an ARCH(1) model. For $\alpha = 0$ and $\beta = 0$ the error sequence reduces to a sequence of (possibly non-Gaussian) i.i.d errors. We allow for varying degrees of volatility persistence modeled as GARCH processes with $\alpha + \beta \in \{0, 0.5, 0.95, 0.99\}$. The parameter settings for α and β are similar to settings found in applied work. In addition, we consider AR(1) models with exponential GARCH errors (EGARCH), asymmetric GARCH errors (AGARCH) and with the GJR-GARCH errors proposed by Glosten et al. (1993). Our parameter settings are based on Engle and Ng (1993).

Finally, we also consider the stochastic volatility model $\varepsilon_t = v_t \exp(h_t)$ with $h_t = \lambda h_{t-1} + 0.5 u_t$, where $|\lambda| < 1$ and (u_t, v_t) is a sequence of independent bivariate normal random variables with zero mean and covariance matrix $diag(\sigma_u^2, 1)$. This model is a m.d.s. model and satisfies Assumption A. We follow Deo (2000) in postulating the values $(0.936, 0.424)$ and $(0.951, 0.314)$ for (λ, σ_u). These are values obtained by Shephard (1996) by fitting this stochastic volatility model to real exchange rate data.

We generate repeated trials of length $n \in \{50, 100, 200, 400\}$ from these processes and conduct bootstrap inference based on the fitted AR(1) model for each trial. All fitted models include an intercept. For the recursive-design bootstrap methods, we generate the start-up values by randomly drawing observations with replacement from the original data set (see, e.g. Berkowitz and Kilian, 2000). The number of Monte Carlo trials is 10,000 with 999 bootstrap replications each. The fixed-design and recursive-design WB involve applying the WB to the residuals of the fitted model. Recall that the WB innovation is $\hat{\varepsilon}_t^* = \hat{\varepsilon}_t \eta_t$, with $\hat{\varepsilon}_t = y_t - \hat{\phi}_0 - \hat{\phi}_1 y_{t-1}$, where η_t is an i.i.d. sequence with mean zero and variance one such that $E^* |\eta_t|^4 \leqslant \Delta < \infty$. In practice, there are several choices for η_t that satisfy these conditions. In the baseline simulations we use $\eta_t \sim N(0,1)$. Our results are robust to alternative choices, as will be shown at the end of this section.

We are interested in studying the coverage accuracy of nominal 90% symmetric percentile-*t* bootstrap confidence intervals for the slope parameter ϕ_1. We also considered equal-tailed percentile-*t* intervals, but found that symmetric percentile-*t* intervals of the form

$$\left(\hat{\phi}_1 - t_{0.9}^* n^{-1/2} \sqrt{\hat{C}_{11}}, \hat{\phi}_1 + t_{0.9}^* n^{-1/2} \sqrt{\hat{C}_{11}} \right),$$

where $\Pr(|t_{\hat{\phi}_1^*}| \leqslant t_{0.9}^*) = 0.9$, virtually always were slightly more accurate. Unlike the percentile interval, the construction of the bootstrap-*t* interval requires the use of an estimate of the standard error. We use the heteroskedasticity-robust estimator of the covariance proposed by Nicholls and Pagan (1983) based on work by Eicker (1963) and White (1980):

$$(X'X)^{-1} X' \, diag(\hat{\varepsilon}_t^2) X (X'X)^{-1},$$

where X denotes the regressor matrix of the AR model. We also experimented with several modified robust covariance estimators (see MacKinnon and White, 1985; Chesher and Jewitt, 1987; Davidson and Flachaire, 2001). For our sample sizes, none of these estimators performed better than the basic estimator proposed by Nicholls and Pagan (1983). Finally, virtually identical results were obtained based on WB bootstrap standard error estimates. The latter approach involves a nested bootstrap loop and is not recommended for computational reasons. As a benchmark we also include the coverage rates of the Gaussian large-sample approximation based on Nicholls–Pagan robust standard errors.

The simulation results are in Tables 2–5. Starting with the results for N-GARCH errors in Table 2, several broad tendencies emerge. First, the accuracy of the standard recursive-design bootstrap procedure based on i.i.d. resampling of the residuals is high when the model errors are truly i.i.d., but can be very poor in the presence of N-GARCH. In the latter case, accuracy tends to deteriorate for large *n*. Second, for sample sizes of 100 or larger, conventional large-sample approximations based on robust standard errors tend to be more accurate than the recursive-design i.i.d. bootstrap in the presence of N-GARCH, but less accurate for models with i.i.d. errors. In either case, the coverage rates may be substantially below the nominal level. Third, all three robust bootstrap methods tend to be more accurate than the i.i.d. bootstrap or the conventional Gaussian approximation, when the errors are conditionally heteroskedastic. Fourth, for persistent processes, the accuracy of the recursive-design WB is typically higher than that of the pairwise bootstrap. For large *n* these differences vanish and both methods are about equally accurate. The accuracy of the recursive-design wild bootstrap is comparable to that of the recursive-design i.i.d. bootstrap for models with i.i.d. errors. The fixed-design WB is typically less accurate than the recursive-design WB and the pairwise bootstrap, although the discrepancies diminish for large *n*.

The results for the AR(1) model with t_5-GARCH errors in Table 3 are qualitatively similar, except that the accuracy of the recursive-design i.i.d. bootstrap tends to be even lower than for N-GARCH processes. In Table 4 we explore a number of additional models of conditional heteroskedasticity that have been used primarily to model returns in empirical finance. The results for the stochastic volatility model

Table 2
Coverage rates of nominal 90% symmetric percentile-t intervals for ϕ_1: AR(1)-N-GARCH model

DGP: $y_t = \phi_1 y_{t-1} + \varepsilon_t, \ \varepsilon_t = h_t^{1/2} v_t, \ h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \ v_t \sim N(0,1)$

n	ϕ_1	$\alpha + \beta$	α	β	Recursive i.i.d.	Recursive WB	Fixed WB	Pairwise	Robust SE Gaussian
50	0	0	0	0	89.1	90.1	89.0	88.9	86.0
		0.5	0.5	0	77.5	88.9	87.9	89.5	84.8
		0.95	0.3	0.65	81.4	89.2	88.5	89.4	85.2
		0.99	0.2	0.79	84.1	89.5	88.7	89.2	85.5
		0.99	0.05	0.94	88.6	90.1	89.2	88.8	86.0
	0.9	0	0	0	83.9	83.2	78.7	79.7	76.0
		0.5	0.5	0	80.4	84.4	80.5	82.0	76.6
		0.95	0.3	0.65	80.1	84.0	80.5	81.4	76.8
		0.99	0.2	0.79	80.8	83.6	80.2	80.7	76.1
		0.99	0.05	0.94	83.7	83.3	79.0	79.6	75.7
100	0	0	0	0	89.7	90.2	89.4	89.5	88.0
		0.5	0.5	0	73.6	89.3	88.5	89.3	86.1
		0.95	0.3	0.65	77.2	89.6	88.8	89.5	86.7
		0.99	0.2	0.79	80.6	90.1	89.4	89.4	86.8
		0.99	0.05	0.94	88.7	90.4	89.6	89.6	87.9
	0.9	0	0	0	87.4	87.5	84.8	84.0	82.5
		0.5	0.5	0	82.7	87.8	85.0	85.5	82.7
		0.95	0.3	0.65	81.5	87.9	85.6	85.3	82.5
		0.99	0.2	0.79	83.1	87.8	85.5	85.1	82.6
		0.99	0.05	0.94	86.9	87.5	85.0	84.2	82.3
200	0	0	0	0	89.6	90.5	89.9	89.7	89.2
		0.5	0.5	0	70.7	89.3	88.5	89.4	87.2
		0.95	0.3	0.65	72.9	89.4	88.9	89.2	87.3
		0.99	0.2	0.79	76.4	89.7	89.0	89.6	87.8
		0.99	0.05	0.94	87.9	90.4	89.6	89.6	88.9
	0.9	0	0	0	89.3	88.9	87.0	87.1	86.4
		0.5	0.5	0	83.6	88.6	87.0	88.1	86.7
		0.95	0.3	0.65	79.9	89.4	88.3	88.1	86.5
		0.99	0.2	0.79	81.2	89.8	88.5	88.5	86.9
		0.99	0.05	0.94	88.0	89.3	87.3	87.3	86.4
400	0	0	0	0	90.3	90.8	90.6	90.2	89.8
		0.5	0.5	0	68.5	90.0	89.4	89.9	88.3
		0.95	0.3	0.65	68.6	90.2	89.8	90.0	88.4
		0.99	0.2	0.79	72.2	90.6	90.0	90.0	88.7
		0.99	0.05	0.94	87.4	90.8	90.3	90.0	89.7
	0.9	0	0	0	90.0	89.7	88.3	88.6	88.2
		0.5	0.5	0	83.4	89.3	88.2	89.6	88.5
		0.95	0.3	0.65	76.2	89.5	88.8	89.5	88.2
		0.99	0.2	0.79	76.8	89.7	89.0	89.6	88.6
		0.99	0.05	0.94	87.9	89.7	88.6	89.0	88.5

are qualitatively the same as for N-GARCH and t-GARCH. For the other three models, we find that there is little to choose between the recursive-design WB and the pairwise bootstrap. Their coverage probability for small samples and highly persistent

Table 3
Coverage rates of nominal 90% symmetric percentile-t intervals for ϕ_1: AR(1)-t_5-GARCH model

DGP: $y_t = \phi_1 y_{t-1} + \varepsilon_t$, $\varepsilon_t = h_t^{1/2} v_t$, $h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}$, $v_t \sim t_5$

n	ϕ_1	$\alpha + \beta$	α	β	Recursive i.i.d.	Recursive WB	Fixed WB	Pairwise	Robust SE Gaussian
50	0	0	0	0	90.6	89.1	88.2	89.5	86.0
		0.5	0.5	0	75.5	87.6	86.3	89.4	83.1
		0.95	0.3	0.65	80.9	88.2	86.9	89.5	83.9
		0.99	0.2	0.79	83.5	88.4	87.3	89.1	84.2
		0.99	0.05	0.94	89.5	89.1	87.9	89.4	85.9
	0.9	0	0	0	84.5	83.8	80.0	81.1	77.4
		0.5	0.5	0	79.5	84.3	81.0	83.0	77.4
		0.95	0.3	0.65	79.4	84.4	80.8	82.9	77.2
		0.99	0.2	0.79	80.7	84.3	80.3	82.5	76.9
		0.99	0.05	0.94	84.3	83.6	80.0	81.0	76.9
100	0	0	0	0	90.3	89.7	89.0	89.5	88.0
		0.5	0.5	0	70.6	88.0	87.8	89.0	84.8
		0.95	0.3	0.65	75.3	88.7	88.3	88.9	86.1
		0.99	0.2	0.79	78.1	89.0	88.7	88.8	86.4
		0.99	0.05	0.94	88.3	89.5	89.2	89.2	87.8
	0.9	0	0	0	88.6	88.0	84.0	85.5	82.7
		0.5	0.5	0	82.3	88.7	85.3	86.9	83.2
		0.95	0.3	0.65	81.4	88.7	85.4	86.1	83.2
		0.99	0.2	0.79	82.3	88.2	85.3	85.9	83.4
		0.99	0.05	0.94	87.3	87.9	84.4	85.0	83.0
200	0	0	0	0	90.6	90.3	89.5	89.6	88.8
		0.5	0.5	0	66.2	88.8	88.0	89.8	85.5
		0.95	0.3	0.65	70.6	89.1	88.5	89.6	86.9
		0.99	0.2	0.79	74.1	89.4	88.9	89.8	87.2
		0.99	0.05	0.94	87.2	90.1	88.8	89.4	88.0
	0.9	0	0	0	89.4	89.0	87.2	87.2	86.6
		0.5	0.5	0	80.7	89.4	87.7	89.0	86.6
		0.95	0.3	0.65	77.3	88.8	88.1	88.4	86.8
		0.99	0.2	0.79	78.7	89.0	87.9	88.2	86.6
		0.99	0.05	0.94	87.6	89.1	87.2	87.4	86.4
400	0	0	0	0	90.1	90.1	89.3	90.1	88.8
		0.5	0.5	0	61.2	89.3	87.7	90.5	85.9
		0.95	0.3	0.65	64.6	89.8	88.5	90.4	87.0
		0.99	0.2	0.79	68.4	89.7	89.1	90.3	87.8
		0.99	0.05	0.94	84.6	90.1	89.7	90.4	88.9
	0.9	0	0	0	89.5	89.5	88.6	88.7	88.4
		0.5	0.5	0	79.2	89.9	88.4	89.9	87.7
		0.95	0.3	0.65	72.5	89.7	88.8	90.3	87.8
		0.99	0.2	0.79	74.0	89.6	89.0	89.8	88.1
		0.99	0.05	0.94	85.6	89.6	88.8	89.2	88.3

data tends to be too low, but consistently higher than that of any alternative method. In all other cases, both methods are highly accurate. Neither the recursive-design i.i.d. bootstrap nor the conventional Gaussian approximation perform well. The high

Table 4
Coverage rates of nominal 90% symmetric percentile-t intervals for ϕ_1

(a) AR(1)-EGARCH model (Engle and Ng, 1993)

DGP: $y_t = \phi_1 y_{t-1} + \varepsilon_t,\ \varepsilon_t = h_t^{1/2} v_t,\ ln(h_t) = -0.23 + 0.9\, ln(h_{t-1}) + 0.25[|v_{t-1}^2| - 0.3 v_{t-1}]$
$v_t \sim N(0,1)$

n	ϕ_1	Recursive i.i.d.	Recursive WB	Fixed WB	Pairwise	Robust SE Gaussian
50	0	79.4	88.7	88.2	89.6	85.3
	0.9	79.5	84.6	81.2	82.3	77.4
100	0	73.8	90.0	89.3	89.4	86.1
	0.9	80.1	87.4	85.1	86.6	83.3
200	0	68.7	89.7	89.1	90.0	87.3
	0.9	78.3	88.7	87.4	88.6	86.6
400	0	63.8	89.8	89.1	90.2	88.0
	0.9	74.5	89.3	88.3	89.4	88.2

(b) AR(1)-AGARCH model (Engle, 1990)

DGP: $y_t = \phi_1 y_{t-1} + \varepsilon_t,\ \varepsilon_t = h_t^{1/2} v_t,\ h_t = 0.0216 + 0.6896 h_{t-1} + 0.3174[\varepsilon_{t-1} - 0.1108]^2$
$v_t \sim N(0,1)$

n	ϕ_1	Recursive i.i.d.	Recursive WB	Fixed WB	Pairwise	Robust SE Gaussian
50	0	80.7	89.2	88.4	89.8	85.6
	0.9	80.3	84.5	81.2	82.6	77.4
100	0	74.8	89.8	89.3	89.5	86.2
	0.9	79.8	87.4	85.6	86.5	83.8
200	0	68.5	90.0	89.3	90.0	87.5
	0.9	76.5	88.9	87.8	88.7	86.8
400	0	62.0	89.8	89.1	89.8	87.9
	0.9	68.8	89.3	88.6	90.0	88.2

(c) AR(1)-GJR GARCH model (Glosten et al., 1993)

DGP: $y_t = \phi_1 y_{t-1} + \varepsilon_t,\ \varepsilon_t = h_t^{1/2} v_t,\ h_t = 0.005 + 0.7 h_{t-1} + 0.28[|\varepsilon_{t-1}| - 0.23\varepsilon_{t-1}]^2$
$v_t \sim N(0,1)$

n	ϕ_1	Recursive i.i.d.	Recursive WB	Fixed WB	Pairwise	Robust SE Gaussian
50	0	81.8	89.3	88.5	90.0	85.8
	0.9	80.0	84.4	81.4	82.3	77.4
100	0	75.8	90.2	89.6	89.3	86.2
	0.9	79.7	87.7	85.4	86.3	83.6
200	0	70.1	90.2	89.5	89.9	87.8
	0.9	77.2	89.0	87.8	89.0	87.0
400	0	64.1	90.1	89.5	90.2	88.5
	0.9	70.5	89.6	88.9	90.2	88.8

Table 4
Contd.

(d) AR(1)-stochastic volatility model (Shephard, 1996)

DGP: $y_t = \phi_1 y_{t-1} + \varepsilon_t, \ \varepsilon_t = v_t \exp(h_t), \ h_t = \lambda h_{t-1} + 0.5 u_t, \ (u_t, v_t) \sim N[0, diag(\sigma_u^2, 1)]$

n	ϕ_1	λ	σ_u	Recursive i.i.d.	Recursive WB	Fixed WB	Pairwise	Robust SE Gaussian
50	0	0.936	0.424	82.3	88.0	87.2	89.3	85.8
		0.951	0.314	84.9	89.9	87.8	89.4	85.8
	0.9	0.936	0.424	80.5	84.4	80.7	83.0	77.4
		0.951	0.314	82.0	83.9	80.2	81.8	77.4
100	0	0.936	0.424	78.2	89.5	88.8	89.7	86.2
		0.951	0.314	81.5	89.8	88.9	89.6	86.2
	0.9	0.936	0.424	82.0	87.7	85.7	86.3	83.6
		0.951	0.314	83.5	87.6	85.1	85.8	83.6
200	0	0.936	0.424	73.0	89.7	89.0	89.4	87.8
		0.951	0.314	78.1	89.7	89.2	89.6	87.4
	0.9	0.936	0.424	79.6	89.2	87.5	88.4	87.0
		0.951	0.314	82.2	89.0	87.5	88.0	87.0
400	0	0.936	0.424	69.3	89.8	89.2	90.0	88.5
		0.951	0.314	74.7	90.0	89.5	89.6	88.5
	0.9	0.936	0.424	76.4	89.7	89.0	89.4	88.8
		0.951	0.314	79.9	89.5	88.7	89.2	88.8

accuracy of the recursive-design WB even for EGARCH, AGARCH and GJR-GARCH error processes is surprising, given its lack of theoretical support for these DGPs. Apparently, the failure of the sufficient conditions for the asymptotic validity of the recursive-design WB method has little effect on its performance in small samples. Fortunately, applications in finance, for which such asymmetric volatility models have been developed, invariably involve large sample sizes, conditions under which pairwise resampling is just as accurate as the recursive-design WB and theoretically justified.

We conclude this section with a sensitivity analysis of the effect that the choice of η_t has on the performance of the wild bootstrap. To conserve space, we focus on the recursive-design WB only. In the baseline simulations we used $\eta_t \sim N(0,1)$. Table 5 shows additional results based on the two-point distribution $\eta_t = -(\sqrt{5}-1)/2$ with probability $p = (\sqrt{5}+1)/(2\sqrt{5})$ and $\eta_t = (\sqrt{5}+1)/2$ with probability $1-p$, as proposed by Mammen (1993), and the two-point distribution $\eta_t = 1$ with probability 0.5 and $\eta_t = -1$ with probability 0.5, as proposed by Liu (1988). The DGPs involve N-GARCH errors as in Table 2. The baseline results for $\eta_t \sim N(0,1)$ are also included for comparison. Table 5 shows that the coverage results are remarkably robust to the choice of η_t. Moreover, none of the three WB resampling schemes clearly dominates the others.

Given the computational costs of the simulation study, we have chosen to focus on a stylized autoregressive model, but have explored a wide range of conditionally heteroskedastic errors. Although our simulation results are necessarily tentative, they suggest that the recursive-design WB for autoregressions should replace conventional

Table 5
Coverage rates of nominal 90% symmetric percentile-t intervals for ϕ_1: AR(1)-N-GARCH model

DGP: $y_t = \phi_1 y_{t-1} + \varepsilon_t$, $\varepsilon_t = h_t^{1/2} v_t$, $h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}$, $v_t \sim N(0,1)$

n	ϕ_1	$\alpha + \beta$	α	β	Alternative recursive-design WB schemes		
					N(0,1)	Mammen	Liu
50	0	0	0	0	90.1	89.2	88.9
		0.5	0.5	0	88.9	88.9	88.6
		0.95	0.3	0.65	89.2	88.9	88.7
		0.99	0.2	0.79	89.5	89.1	88.8
		0.99	0.05	0.94	90.1	89.1	88.7
	0.9	0	0	0	83.2	83.8	84.3
		0.5	0.5	0	84.4	85.2	85.4
		0.95	0.3	0.65	84.0	84.0	84.6
		0.99	0.2	0.79	83.6	83.7	84.3
		0.99	0.05	0.94	83.3	83.7	84.3
100	0	0	0	0	90.2	90.0	89.4
		0.5	0.5	0	89.3	89.3	88.7
		0.95	0.3	0.65	89.6	89.4	89.2
		0.99	0.2	0.79	90.1	89.4	89.1
		0.99	0.05	0.94	90.4	89.8	89.4
	0.9	0	0	0	87.5	87.0	87.3
		0.5	0.5	0	87.8	87.9	88.1
		0.95	0.3	0.65	87.9	87.2	87.6
		0.99	0.2	0.79	87.8	87.4	87.9
		0.99	0.05	0.94	87.5	87.1	87.4
200	0	0	0	0	90.5	90.3	89.9
		0.5	0.5	0	89.3	89.3	89.0
		0.95	0.3	0.65	89.4	89.6	89.2
		0.99	0.2	0.79	89.7	89.8	89.4
		0.99	0.05	0.94	90.4	90.0	89.6
	0.9	0	0	0	88.9	88.9	89.0
		0.5	0.5	0	88.6	89.5	89.7
		0.95	0.3	0.65	89.4	89.5	89.5
		0.99	0.2	0.79	89.8	89.5	89.7
		0.99	0.05	0.94	89.3	89.4	89.4
400	0	0	0	0	90.8	90.4	90.1
		0.5	0.5	0	90.0	89.9	89.6
		0.95	0.3	0.65	90.2	90.0	89.7
		0.99	0.2	0.79	90.6	90.2	89.8
		0.99	0.05	0.94	90.8	90.3	90.2
	0.9	0	0	0	89.7	90.0	89.7
		0.5	0.5	0	89.3	90.2	90.2
		0.95	0.3	0.65	89.5	90.0	90.2
		0.99	0.2	0.79	89.7	90.1	90.1
		0.99	0.05	0.94	89.7	90.0	90.0

recursive design i.i.d. bootstrap methods in many applications. The pairwise bootstrap provides a suitable alternative when sample sizes are at least moderately large and the possibility of asymmetric forms of GARCH is a practical concern. Even for moderate

sample sizes the accuracy of the pairwise bootstrap is slightly higher than that of the fixed-design bootstrap.

5. Concluding remarks

The aim of the paper has been to extend the range of applications of autoregressive bootstrap methods in empirical finance and macroeconometrics. We analyzed the theoretical properties of three bootstrap procedures for stationary autoregressions that are robust to conditional heteroskedasticity of unknown form: the fixed-design WB, the recursive-design WB and the pairwise bootstrap. Throughout the paper, we established conditions for the first-order asymptotic validity of these bootstrap procedures. We did not attempt to address the issue of the existence of higher-order asymptotic refinements provided by the bootstrap approximation. Arguments aimed at proving asymptotic refinements require the existence of an Edgeworth expansion for the distribution of the estimator of interest. Establishing the existence of such an Edgeworth expansion is beyond the scope of this paper. Moreover, the quality of the finite-sample approximation provided by analytic Edgeworth expansions often is poor and less accurate than bootstrap approximations. Thus, Edgeworth expansions in general are imperfect guides to the relative accuracy of alternative bootstrap methods (see Härdle et al., 2001). Indeed, preliminary simulation evidence indicates that wild bootstrap methods based on two-point distributions, which may be expected to yield asymptotic refinements in our context, do not perform systematically better than the first-order accurate methods studied in this paper. Nevertheless, we found that the robust bootstrap approximation is typically more accurate in small samples than the usual first-order asymptotic approximation based on robust standard errors. Our simulation results also highlighted the dangers of incorrectly modelling the error term in dynamic regression models as i.i.d. We found that conventional residual-based bootstrap methods may be very inaccurate in the presence of conditional heteroskedasticity.

Based on the theoretical and simulation results in this paper, no single bootstrap method for dealing with conditional heteroskedasticity of unknown form will be optimal in all cases. The recursive-design WB seems best suited for applications in empirical macroeconomics. This method performs well, whether the error term of the autoregression is i.i.d. or conditionally heteroskedastic, but it lacks theoretical justification for some forms of asymmetric GARCH that have figured prominently in the literature on high-frequency returns. When the sample size is at least moderately large and asymmetric forms of GARCH are a practical concern, the pairwise bootstrap method provides a suitable alternative. The fixed-design WB has the same theoretical justification as the pairwise bootstrap for parametric models, but appears to be less accurate in practice.

There are several interesting extensions of the approach taken in this paper. One possible extension is the development of bootstrap methods for conditionally heteroskedastic stationary autoregressions of possibly infinite order. This extension is considered in Gonçalves and Kilian (2003). Another useful extension would be to establish the

validity of the recursive-design WB for regression parameters in I(1) autoregressions that can be written in terms of zero mean stationary regressors, generalizing recent work by Inoue and Kilian (2002) on I(1) autoregressive models with i.i.d. errors. Yet another useful extension would be to establish the asymptotic validity of robust versions of the grid bootstrap of Hansen (1999). These extensions are nontrivial and left for future research.

Acknowledgements

Appendix.

Throughout this appendix, K denotes a generic constant independent of n. We use *u.i.* to mean uniformly integrable. Given an $m \times n$ matrix A; let $\|A\| = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|$; for an $m \times 1$ vector a, let $|a| = \sum_{i=1}^{m} |a_i|$. For any $n \times n$ matrix A, $diag(a_{11}, \ldots, a_{nn})$ denotes a diagonal matrix with a_{ii}, $i = 1, \ldots, n$ in the main diagonal. Similarly, let $[a_{ij}]_{i,j=1,\ldots,n}$ denote a matrix A with typical element a_{ij}. For any bootstrap statistic T_n^* we write $T_n^* \xrightarrow{P^*} 0$ in probability when $\lim_{n \to \infty} P[P^*(|T_n^*| > \delta) > \delta] = 0$ for any $\delta > 0$, i.e., $P^*(|T_n^*| > \delta) = o_P(1)$. We write $T_n^* \Rightarrow^{d_{P^*}} D$, in probability, for any distribution D, when weak convergence under the bootstrap probability measure occurs in a set with probability converging to one.

The following CLT will be useful in proving results for the bootstrap (cf. White, 1999, p. 133; the Lindeberg condition there has been replaced by the stronger Lyapunov condition here):

Theorem A.1 (Martingale difference arrays CLT). *Let $\{Z_{nt}, \mathscr{F}_{nt}\}$ be a martingale difference array such that $\sigma_{nt}^2 = E(Z_{nt}^2)$, $\sigma_{nt}^2 \neq 0$, and define $\bar{Z}_n \equiv n^{-1} \sum_{t=1}^{n} Z_{nt}$ and $\bar{\sigma}_n^2 \equiv Var(\sqrt{n}\bar{Z}_n) = n^{-1} \sum_{t=1}^{n} \sigma_{nt}^2$. If*

1. $n^{-1} \sum_{t=1}^{n} Z_{nt}^2 / \bar{\sigma}_n^2 - 1 \xrightarrow{P} 0$, *and*
2. $\lim_{n \to \infty} \bar{\sigma}_n^{-2(1+\delta)} n^{-(1+\delta)} \sum_{t=1}^{n} E|Z_{nt}|^{2(1+\delta)} = 0$ *for some $\delta > 0$,*

then $\sqrt{n}\bar{Z}_n / \bar{\sigma}_n \Rightarrow N(0,1)$.

The following lemma generalizes Kuersteiner's (2001) Lemma A.1. Kuersteiner's Assumption A.1 is stronger than our Assumption A in that it assumes that $\{\varepsilon_t\}$ is strictly stationary and ergodic, and in that it imposes a summability condition on the fourth-order cumulants.

Lemma A.1. *Under Assumption* A, *for each* $m \in \mathbb{N}$, *m fixed, the vector*

$$n^{-1/2} \sum_{t=1}^{n} (\varepsilon_t \varepsilon_{t-1}, \dots, \varepsilon_t \varepsilon_{t-m})' \Rightarrow \mathrm{N}(0, \Omega_m),$$

where $\Omega_m = \sigma^4 [\tau_{r,s}]_{r,s=1,\dots,m}$.

Lemmas A.2–A.5 are used to prove the asymptotic validity of the recursive-design WB (cf. Theorem 3.2). In these lemmas, $\hat{\varepsilon}_t^* = \hat{\varepsilon}_t \eta_t$, $t = 1, \dots, n$, where $\hat{\varepsilon}_t = y_t - \hat{\phi}' Y_{t-1}$, and η_t is i.i.d. $(0,1)$ such that $\mathrm{E}^* |\eta_t|^4 \leqslant \Delta < \infty$.

Lemma A.2. *Under Assumption* A, *for fixed* $j \in \mathbb{N}$,

(i) $n^{-1} \sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^{*2} \xrightarrow{P^*} \sigma^2$, *in probability;*

(ii) $n^{-1} \sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^* \xrightarrow{P^*} 0$, *in probability.*

If we strengthen Assumption A *by* A′(vi′), *then for fixed* $i, j \in \mathbb{N}$,

(iii) $n^{-1} \sum_{t=\max(i,j)+1}^{n} \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_{t-i}^* \hat{\varepsilon}_t^{*2} \xrightarrow{P^*} \sigma^4 \tau_{i,j} 1(i=j)$, *in probability, where* $1(i=j)$ *is* 1 *if* $i = j$, *and* 0 *otherwise.*

The following lemma is the WB analogue of Lemma A.1.

Lemma A.3. *Under Assumption* A *strengthened by* A(vi′), *for all fixed* $m \in \mathbb{N}$,

$$n^{-1/2} \sum_{t=m+1}^{n} (\hat{\varepsilon}_t^* \hat{\varepsilon}_{t-1}^*, \dots, \hat{\varepsilon}_t^* \hat{\varepsilon}_{t-m}^*)' \Rightarrow^{d_{P^*}} \mathrm{N}(0, \tilde{\Omega}_m),$$

in probability, where $\tilde{\Omega}_m \equiv \sigma^4 \, diag(\tau_{1,1}, \dots, \tau_{m,m})$ *and* $\Rightarrow^{d_{P^*}}$ *denotes weak convergence under the bootstrap probability measure.*

Lemma A.4. *Suppose Assumption* A *holds. Then* $n^{-1} \sum_{t=1}^{n} Y_{t-1}^* Y_{t-1}^{*\prime} \xrightarrow{P^*} A$, *in probability, where* $A \equiv \sigma^2 \sum_{j=1}^{\infty} b_j b_j'$.

Lemma A.5. *Suppose Assumption* A *strengthened by* A(vi′) *holds. Then*

$$n^{-1/2} \sum_{t=1}^{n} Y_{t-1}^* \hat{\varepsilon}_t^* \Rightarrow^{d_{P^*}} \mathrm{N}(0, \tilde{B}),$$

in probability, where $\tilde{B} = \sum_{j=1}^{\infty} b_j b_j' \sigma^4 \tau_{j,j}$.

Proof of Theorem 3.1. We show that (i) $A_{1n} \equiv n^{-1} \sum_{t=1}^{n} Y_{t-1} Y_{t-1}' \xrightarrow{P} A$; and (ii) $A_{2n} \equiv n^{-1/2} \sum_{t=1}^{n} Y_{t-1} \varepsilon_t \Rightarrow \mathrm{N}(0, B)$. First, notice that for any stationary AR(p) process we have $y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, where $\{\psi_j\}$ satisfies the recursion $\psi_s - \phi_1 \psi_{s-1} - \cdots - \phi_p \psi_{s-p} = 0$ with $\psi_0 = 1$ and $\psi_j = 0$ for $j < 0$, implying that $\sum_{j=0}^{\infty} j |\psi_j| < \infty$. We can write $Y_{t-1} =$

$(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-1-j}, \ldots, \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-p-j})' = \sum_{j=1}^{\infty} b_j \varepsilon_{t-j}$ with $b_j = (\psi_{j-1}, \ldots, \psi_{j-p})'$, where $\psi_{-j} = 0$ for all $j > 0$. Hence, by direct evaluation,

$$A \equiv \mathrm{E}(Y_{t-1} Y_{t-1}') = \mathrm{E}\left[\left(\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} b_j b_i' \varepsilon_{t-j} \varepsilon_{t-i}\right)\right] = \sigma^2 \sum_{j=1}^{\infty} b_j b_j'$$

$$= \left[\sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k-l|}\right]_{k,l=1,\ldots,p},$$

since $\mathrm{E}(\varepsilon_{t-i} \varepsilon_{t-j}) = 0$ for $i \neq j$ under the m.d.s. assumption, and $\sum_{j=0}^{\infty} |\psi_j \psi_{j+|k-l|}| \leqslant \sum_{j=0}^{\infty} |\psi_j| \sum_{j=0}^{\infty} |\psi_{j+|k-l|}| < \infty$ for all k, l. To show (i), for fixed $m \in \mathbb{N}$, define $A_{1n}^m \equiv n^{-1} \sum_{t=1}^{n} Y_{t-1,m} Y_{t-1,m}'$, where $Y_{t-1,m} = \sum_{j=1}^{m} b_j \varepsilon_{t-j}$. It suffices to show: (a) $A_{1n}^m \xrightarrow{P} A_1^m \equiv \sigma^2 \sum_{j=1}^{m} b_j b_j'$ as $n \to \infty$, for each fixed m; (b) $A_1^m \to A$ as $m \to \infty$, and (c) $\lim_{m \to \infty} \limsup_{n \to \infty} P[\|A_{1n} - A_{1n}^m\| \geqslant \delta] = 0$ for all $\delta > 0$ (cf. Proposition 6.3.9 of Brockwell and Davis (BD), 1991, p. 207). For (a), we have $A_{1n}^m = \sum_{j=1}^{m} \sum_{i=1}^{m} b_j b_i' n^{-1} \sum_{t=1}^{n} \varepsilon_{t-j} \varepsilon_{t-i}$. For fixed $i \neq j$ it follows that $n^{-1} \sum_{t=1}^{n} \varepsilon_{t-j} \varepsilon_{t-i} \xrightarrow{P} 0$ by Andrews' (1988) LLN for u.i. L_1-mixingales, since $\{\varepsilon_{t-j} \varepsilon_{t-i}\}$ is a m.d.s. with $\mathrm{E}|\varepsilon_{t-j} \varepsilon_{t-i}|^r \leqslant \|\varepsilon_{t-j}\|_{2r}^r \|\varepsilon_{t-i}\|_{2r}^r < \Delta^{2r} < \infty$ by Cauchy–Schwartz and Assumption A(vi). For fixed $i = j$, we can write $n^{-1} \sum_{t=1}^{n} \varepsilon_{t-j}^2 - \sigma^2 = n^{-1} \sum_{t=1}^{n} z_t + n^{-1} \sum_{t=1}^{n} \mathrm{E}(\varepsilon_{t-j}^2 | \mathscr{F}_{t-j-1}) - \sigma^2$, with $z_t = \varepsilon_{t-j}^2 - \mathrm{E}(\varepsilon_{t-j}^2 | \mathscr{F}_{t-j-1})$. Since z_t can be shown to be an u.i. m.d.s, the first term goes to zero in probability by Andrews' LLN. The second term vanishes in probability by Assumption A(iii). Thus, $n^{-1} \sum_{t=1}^{n} \varepsilon_{t-j}^2 - \sigma^2 \xrightarrow{P} 0$ for fixed j. It follows that $A_{1n}^m \xrightarrow{P} \sigma^2 \sum_{j=1}^{m} b_j b_j' \equiv A_1^m$, which completes the proof of (a). Part (b) follows from the dominated convergence theorem, given that $\|\sum_{j=1}^{\infty} b_j b_j'\| \leqslant \sum_{j=1}^{\infty} |b_j|^2 < \infty$. To prove (c), note that for any $\delta > 0$,

$$P[\|A_{1n} - A_{1n}^m\| \geqslant \delta] \leqslant \frac{1}{\delta} \mathrm{E}\|A_{1n} - A_{1n}^m\|$$

$$\leqslant \frac{2}{\delta} \left(\sum_{j > m}^{\infty} |b_j|\right) \left(\sum_{j=1}^{\infty} |b_j|\right) n^{-1} \sum_{t=1}^{n} \mathrm{E}|\varepsilon_{t-i} \varepsilon_{t-j}|$$

$$\leqslant \left(\sum_{j > m}^{\infty} |b_j|\right) K \to 0 \quad \text{as } m \to \infty,$$

since $\mathrm{E}|\varepsilon_{t-i} \varepsilon_{t-j}| \leqslant \Delta$ for some $\Delta < \infty$, and since $\sum_{j=1}^{\infty} |b_j| < \infty$. Next, we prove (ii). We apply Proposition 6.3.9 of BD. Let $Z_t = Y_{t-1} \varepsilon_t \equiv \sum_{j=1}^{\infty} b_j \varepsilon_{t-j} \varepsilon_t$. For fixed m, define $Z_t^m = Y_{t-1,m} \varepsilon_t = \sum_{j=1}^{m} b_j \varepsilon_{t-j} \varepsilon_t$, where $Y_{t-1,m}$ is defined as above. We first

show that $n^{-1/2} \sum_{t=1}^{n} Z_t^m \Rightarrow N(0, B_m)$, with $B_m = \sum_{j=1}^{m} \sum_{i=1}^{m} b_j b_i' \sigma^4 \tau_{j,i}$. We have

$$n^{-1/2} \sum_{t=1}^{n} Z_t^m = n^{-1/2} \sum_{t=1}^{n} \sum_{j=1}^{m} b_j \varepsilon_{t-j} \varepsilon_t = \sum_{j=1}^{m} b_j n^{-1/2} \sum_{t=1}^{n} \varepsilon_{t-j} \varepsilon_t \equiv \sum_{j=1}^{m} b_j \chi_{nj}.$$

By Lemma A.1 we have that $(\chi_{n1}, \ldots, \chi_{nm})' \Rightarrow N(0, \Omega_m)$. Thus, $\sum_{j=1}^{m} b_j \chi_{nj} \Rightarrow N(0, B_m)$, with $B_m = b' \Omega_m b$, $b' = (b_1, \ldots, b_m)$. Since $\| \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} b_j b_i' \sigma^4 \tau_{j,i} \| \leqslant \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} |b_j| |b_i| \sigma^4 |\tau_{j,i}| < \infty$, it follows that $B_m \to B \equiv \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} b_j b_i' \sigma^4 \tau_{j,i}$ as $m \to \infty$. Finally, for any $\lambda \in \mathbb{R}^p$ such that $\lambda' \lambda = 1$ and for any $\delta > 0$, we have

$$\lim_{m \to \infty} \lim_{n \to \infty} \sup P \left[\left| n^{-1/2} \sum_{t=1}^{n} \lambda' Z_t - n^{-1/2} \sum_{t=1}^{n} \lambda' Z_t^m \right| \geqslant \delta \right]$$

$$= \lim_{m \to \infty} \lim_{n \to \infty} \sup P \left[\left| n^{-1/2} \sum_{t=1}^{n} \sum_{j>m} \lambda' b_j \varepsilon_{t-j} \varepsilon_t \right| \geqslant \delta \right]$$

$$\leqslant \lim_{m \to \infty} \lim_{n \to \infty} \sup \frac{1}{n \delta^2} E \left(\left| \sum_{t=1}^{n} \sum_{j>m} \lambda' b_j \varepsilon_{t-j} \varepsilon_t \right|^2 \right)$$

$$= \lim_{m \to \infty} \frac{1}{\delta^2} \left(\sum_{j>m} \sum_{i>m} \lambda' b_j b_i' \lambda \sigma^4 \tau_{j,i} \right) = 0,$$

where the inequality holds by Chebyshev's inequality, the second-to-last equality holds by the fact that $E(\varepsilon_{t-j} \varepsilon_t \varepsilon_{s-i} \varepsilon_s) = 0$ for $s \neq t$, and all i, j, and the last equality holds by the summability of $\{\psi_j\}$ and the fact that $\tau_{j,i}$ are uniformly bounded. $\square$

Proof of Theorem 3.2. By Lemma A.4, $n^{-1} \sum_{t=1}^{n} Y_{t-1}^* Y_{t-1}^{*\prime} \xrightarrow{P^*} A$, in probability, whereas Lemma A.5 implies $n^{-1/2} \sum_{t=1}^{n} Y_{t-1}^* \hat{\varepsilon}_t^* \Rightarrow^{d_{P^*}} N(0, \tilde{B})$, in probability. Since under Assumption A(iv'), $B = \tilde{B}$, the result follows by Polya's Theorem, given that the normal distribution is everywhere continuous. $\square$

Proof of Theorem 3.3. We need to show that (a) $n^{-1} \sum_{t=1}^{n} Y_{t-1} Y_{t-1}' \xrightarrow{P} A$, and (b) $n^{-1/2} \sum_{t=1}^{n} Y_{t-1} \hat{\varepsilon}_t^* \Rightarrow^{d_{P^*}} N(0, B)$ in probability. Part (a) was proved in Theorem 3.1. To show part (b) note that

$$n^{-1/2} \sum_{t=1}^{n} Y_{t-1} \hat{\varepsilon}_t^* = n^{-1/2} \sum_{t=1}^{n} Y_{t-1} \varepsilon_t \eta_t - n^{-1/2} \sum_{t=1}^{n} Y_{t-1} (\varepsilon_t - \hat{\varepsilon}_t) \eta_t$$

$$= n^{-1/2} \sum_{t=1}^{n} Y_{t-1} \varepsilon_t \eta_t - n^{-1} \sum_{t=1}^{n} Y_{t-1} Y_{t-1}' \eta_t \sqrt{n}(\hat{\phi} - \phi) \equiv A_1^* - A_2^*.$$

First, note that $A_2^* \xrightarrow{P^*} 0$, in probability, since $\sqrt{n}(\hat{\phi} - \phi) = O_P(1)$ and $n^{-1} \sum_{t=1}^{n} Y_{t-1} Y_{t-1}' \eta_t \xrightarrow{P^*} 0$, in probability. This follows from showing that $E^*(n^{-1} \sum_{t=1}^{n} Y_{t-1} Y_{t-1}' \eta_t) = 0$

and $Var^*(n^{-1} \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} \eta_t) = n^{-2} \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} Y_{t-1} Y'_{t-1} \xrightarrow{P} 0$, under Assumption A. We next show $A_1^* \Rightarrow^{d_{P^*}} N(0, B)$ in probability, where $B = Var(n^{-1/2} \sum_{t=1}^{n} Y_{t-1} \varepsilon_t) = n^{-1} \sum_{t=1}^{n} E(Y_{t-1} Y'_{t-1} \varepsilon_t^2)$. For any $\lambda \in \mathbb{R}^p$, $\lambda' \lambda = 1$, let $Z_t^* = \lambda' Y_{t-1} \varepsilon_t \eta_t$. $\{Z_t^*\}$ is (conditionally) independent such that $E^*(n^{-1/2} \sum_{t=1}^{n} Z_t^*) = 0$ and $Var^*(n^{-1/2} \sum_{t=1}^{n} Z_t^*) = \lambda' n^{-1} \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} \varepsilon_t^2 \lambda$. We now apply Lyapunov's Theorem (e.g. Durrett, 1996, p. 121). Let $\alpha_n^{*2} = \lambda' \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} \varepsilon_t^2 \lambda$. By arguments similar to Theorem 3.1, $n^{-1} \alpha_n^{*2} \xrightarrow{P} B$. If for some $r > 1$

$$\alpha_n^{*-2r} \sum_{t=1}^{n} E^* |Z_t^*|^{2r} \xrightarrow{P} 0 \qquad (A.1)$$

then $\alpha_n^{*-1} \sum_{t=1}^{n} Z_t^* \Rightarrow^{d_{P^*}} N(0,1)$ in probability. By Slutsky's Theorem, it follows that $n^{-1/2} \sum_{t=1}^{n} Z_t^* \Rightarrow^{d_{P^*}} N(0, \lambda' B \lambda)$. To show (A.1), note that the LHS can be written as

$$\left(\frac{\alpha_n^{*2}}{n} \right)^{-r} n^{-r} \sum_{t=1}^{n} |\lambda' Y_{t-1} \varepsilon_t|^{2r} E^* |\eta_t|^{2r}.$$

Thus, it suffices to show that $E|n^{-r} \sum_{t=1}^{n} |\lambda' Y_{t-1} \varepsilon_t|^{2r} E^* |\eta_t|^{2r}| \to 0$. Since $E^* |\eta_t|^{2r} \leqslant \Delta < \infty$, this holds provided $E|\lambda' Y_{t-1} \varepsilon_t|^{2r} \leqslant \Delta < \infty$, which follows under Assumption A. $\square$

Proof of Theorem 3.4. Let $\hat{\varepsilon}_t = y_t - \hat{\phi}' Y_{t-1}$, $\hat{\varepsilon}_t^* = y_t^* - \hat{\phi}' Y_{t-1}^*$, and $\varepsilon_t^* = y_t^* - \phi' Y_{t-1}^*$. We show that (i) $n^{-1} \sum_{t=1}^{n} Y_{t-1}^* Y_{t-1}^{*'} \xrightarrow{P^*} A$ in probability, and (ii) $n^{-1/2} \sum_{t=1}^{n} Y_{t-1}^* \hat{\varepsilon}_t^* \Rightarrow^{d_{P^*}} N(0, B)$ in probability. We can write,

$$n^{-1} \sum_{t=1}^{n} Y_{t-1}^* Y_{t-1}^{*'} - A = \left\{ n^{-1} \sum_{t=1}^{n} Y_{t-1}^* Y_{t-1}^{*'} - n^{-1} \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} \right\}$$

$$+ \left\{ n^{-1} \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} - A \right\} \equiv A_1^* + A_2.$$

Theorem 3.1 shows $A_2 \xrightarrow{P} 0$. Next we show $A_1^* \xrightarrow{P^*} 0$, in probability. Conditional on the data, by Chebyshev's inequality, it suffices that $E^*(A_1^* A_1^{*'}) = o_P(1)$. But

$$E^*(A_1^* A_1^{*'}) = n^{-1} E^* \left(n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \left(Y_{t-1}^* Y_{t-1}^{*'} - n^{-1} \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} \right) \right.$$

$$\times \left(Y_{s-1}^* Y_{s-1}^{*'} - n^{-1} \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} \right)' \right)$$

$$= n^{-1} \left\{ n^{-1} \sum_{t=1}^{n} \left(Y_{t-1} Y'_{t-1} - n^{-1} \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} \right) \right.$$

$$\times \left(Y_{t-1} Y'_{t-1} - n^{-1} \sum_{t=1}^{n} Y_{t-1} Y'_{t-1} \right)' \right\},$$

where the term in curly brackets is $O_P(1)$ given Assumption A (in particular, given A(vi)), delivering the result. Next we show (ii). We can write

$$n^{-1/2} \sum_{t=1}^{n} Y_{t-1}^* \hat{\varepsilon}_t^* = n^{-1/2} \sum_{t=1}^{n} \left(Y_{t-1}^* \varepsilon_t^* - n^{-1} \sum_{t=1}^{n} Y_{t-1} \varepsilon_t \right)$$

$$+ \left(n^{-1} \sum_{t=1}^{n} Y_{t-1} Y_{t-1}' - n^{-1} \sum_{t=1}^{n} Y_{t-1}^* Y_{t-1}^{*\prime} \right) \sqrt{n}(\hat{\phi} - \phi) \equiv B_1^* + B_2^*.$$

Since $B_2^* \xrightarrow{P^*} 0$ in probability, (ii) follows if we prove that $B_1^* \Rightarrow^{d_{P^*}} N(0, B)$ in probability. This follows straightforwardly by an application of Lyapunov's CLT, given that $Z_t^* \equiv Y_{t-1}^* \varepsilon_t^* - n^{-1} \sum_{t=1}^{n} Y_{t-1} \varepsilon_t$ is (conditionally) i.i.d. with mean zero and variance $Var^*(Z_t^*) = n^{-1} \sum_{t=1}^{n} Z_t Z_t'$, where $Z_t \equiv Y_{t-1} \varepsilon_t - n^{-1} \sum_{t=1}^{n} Y_{t-1} \varepsilon_t$, and by arguments similar to those used in the proof of Theorem 3.1, $n^{-1} \sum_{t=1}^{n} Y_{t-1} Y_{t-1}' \varepsilon_t^2 \xrightarrow{P} B$ and $n^{-1} \sum_{t=1}^{n} Y_{t-1} \varepsilon_t \xrightarrow{P} 0$. $\square$

Proof of Corollary 3.1. Given the previous results, it suffices to show that $\hat{C}^* \xrightarrow{P^*} C$, i.e., (i) $\hat{A}^* \xrightarrow{P^*} A$, and (ii) $\hat{B}^* \xrightarrow{P^*} B$, in probability, where $B = \tilde{B}$ for the recursive-design WB. We showed (i) in Lemma A.4 for the recursive-design WB, and in Theorems 3.3 and 3.4, for the fixed-design WB and pairwise bootstrap, respectively. Next, we sketch the proof of (ii). For simplicity we take $p = 1$. The proof for general p is similar. For each of the three bootstrap schemes, we can write $\tilde{\varepsilon}_t^* = \hat{\varepsilon}_t^* - (\hat{\phi}^* - \hat{\phi}) y_{t-1}^*$, where $\hat{\varepsilon}_t^* = \hat{\varepsilon}_t \eta_t$ for the recursive-design and fixed-design WB, and $\hat{\varepsilon}_t^* = y_t^* - \hat{\phi} y_{t-1}^*$ for the pairwise bootstrap. Thus,

$$\hat{B}^* = \hat{B}_1^* + \hat{B}_2^* + \hat{B}_3^*, \quad \text{with}$$

$$\hat{B}_1^* = n^{-1} \sum_{t=1}^{n} y_{t-1}^{*2} \hat{\varepsilon}_t^{*2}, \quad \hat{B}_2^* = -2(\hat{\phi}^* - \hat{\phi}) n^{-1} \sum_{t=1}^{n} y_{t-1}^{*3} \hat{\varepsilon}_t^*, \quad \text{and}$$

$$\hat{B}_3^* = (\hat{\phi}^* - \hat{\phi})^2 n^{-1} \sum_{t=1}^{n} y_{t-1}^{*4}.$$

It is enough to show that with probability approaching one, (a) $\hat{B}_1^* \xrightarrow{P^*} B$, (b) $\hat{B}_2^* \xrightarrow{P^*} 0$, and (c) $\hat{B}_3^* \xrightarrow{P^*} 0$. For the fixed-design WB, starting with (a), note that $y_{t-1}^* = y_{t-1}$, and therefore $\hat{B}_1^* - B = n^{-1} \sum_{t=1}^{n} y_{t-1}^2 \hat{\varepsilon}_t^2 (\eta_t^2 - 1) + n^{-1} \sum_{t=1}^{n} y_{t-1}^2 \hat{\varepsilon}_t^2 - B \equiv \chi_1 + \chi_2$. Under our assumptions $\chi_2 \xrightarrow{P} 0$. Since $\hat{\varepsilon}_t = \varepsilon_t - (\hat{\phi} - \phi) y_{t-1}$, we can write $\chi_1 = n^{-1} \sum_{t=1}^{n} y_{t-1}^2 \varepsilon_t^2 (\eta_t^2 - 1) - 2(\hat{\phi} - \phi) n^{-1} \sum_{t=1}^{n} y_{t-1}^3 \varepsilon_t (\eta_t^2 - 1) + (\hat{\phi} - \phi)^2 n^{-1} \sum_{t=1}^{n} y_{t-1}^4 (\eta_t^2 - 1)$. We can show that each of these terms is $o_{P^*}(1)$ in probability. For the first term, write $z_t = y_{t-1}^2 \varepsilon_t^2 (\eta_t^2 - 1)$, and note that z_t is (conditionally) a m.d.s. with respect to $\mathscr{F}_t^* = \sigma(\eta_t, \ldots, \eta_1)$. Thus, by Andrews' (1988) LLN, it follows that $n^{-1} \sum_{t=1}^{n} z_t \xrightarrow{P^*} 0$, in probability, provided that $E^*|z_t|^r = O_P(1)$, or $E(E^*|z_t|^r) = O(1)$, for some $r > 1$, which holds under our moment conditions (in particular, the existence of $4r$ moments of ε_t suffices). A similar

argument applies to the last two terms of χ_1, where we note that $\hat{\phi} - \phi \xrightarrow{P} 0$. For (b), and given $\hat{\phi}^* - \hat{\phi} = o_{P^*}(1)$, it suffices that $n^{-1} \sum_{t=1}^n y_{t-1}^3 \hat{\varepsilon}_t^* = O_{P^*}(1)$, in probability, or that $E^* |n^{-1} \sum_{t=1}^n y_{t-1}^3 \hat{\varepsilon}_t^*| = O_P(1)$. This condition holds under Assumption A (first apply the triangle inequality, then use the definition of $\hat{\varepsilon}_t$, and finally apply repeatedly the Cauchy–Schwartz inequality to the sums involving products of y_{t-1} and/or ε_t.). For (c), by a reasoning similar to (b), it suffices that $n^{-1} \sum_{t=1}^n y_{t-1}^4 = O_P(1)$, which holds under our moment conditions. For the pairwise bootstrap, we proceed similarly, but rely on the (conditional) independence of (y_t^*, y_{t-1}^*) to obtain the results. In particular, for (a), following Theorem 3.3 we can define $\hat{\varepsilon}_t^* = \varepsilon_t^* - (\hat{\phi} - \phi) y_{t-1}^*$, with $\varepsilon_t^* = y_t^* - \phi y_{t-1}^*$, which implies $\hat{B}_1^* \equiv \chi_1 + \chi_2$, say. In particular, $\chi_1 = n^{-1} \sum_{t=1}^n z_{1t}^* + \zeta$, where $z_{1t}^* = y_{t-1}^{*2} \varepsilon_{t-1}^{*2} - n^{-1} \sum_{t=1}^n y_{t-1}^2 \varepsilon_t^2$ and $\zeta = n^{-1} \sum_{t=1}^n y_{t-1}^2 \varepsilon_t^2$. Under our conditions, $\zeta \xrightarrow{P} B$. Since z_{1t}^* is a uniformly square-integrable m.d.s. (conditional on the original data), Andrews' LLN implies that the first term of χ_1 is $o_{P^*}(1)$ in probability. Similarly, we can show that $\chi_2 = o_{P^*}(1)$ in probability. For the recursive-design WB, for part (a), note that we can write $\hat{B}_1^* = \chi_1 + \chi_2$, where $\chi_1 = \sum_{j=1}^{n-1} \hat{\psi}_{j-1}^2 (n^{-1} \sum_{t=j+1}^n \hat{\varepsilon}_{t-j}^{*2} \hat{\varepsilon}_t^{*2})$, and $\chi_2 = n^{-1} \sum_{t=1}^n \sum_{i,j=1,i\neq j}^{t-1} \hat{\psi}_{j-1} \hat{\psi}_{i-1} \hat{\varepsilon}_{t-i}^* \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^{*2}$. Now, using arguments analogous to those used in the proof of Lemmas A.4 and A.5 we can show that $\chi_1 \xrightarrow{P^*} \tilde{B}$, and $\chi_2 \xrightarrow{P^*} 0$, in probability. Similar arguments apply for (b) and (c).

Proof of Lemma A.1. The proof follows closely that of Lemma A.1 of Kuersteiner (2001). We reproduce his steps under our weaker Assumption A. In particular, we show that for all $\lambda \in \mathbb{R}^m$ such that $\lambda' \lambda = 1$ we have $n^{-1/2} \sum_{t=1}^n \lambda' W_t \Rightarrow N(0, \lambda' \Omega_m \lambda)$, where $W_t = (\varepsilon_t \varepsilon_{t-1}, \ldots, \varepsilon_t \varepsilon_{t-m})'$. Noting that $\{W_t, \mathscr{F}_t\}$ is a vector m.d.s., we check the m.d.s. CLT conditions (cf. Davidson, 1994, Theorem 24.3). Let $Z_t = \lambda' W_t$. We check: (i) $n^{-1} \sum_{t=1}^n [Z_t^2 - E(Z_t^2)] \xrightarrow{P} 0$, where $E(Z_t^2) = \lambda' E(W_t W_t') \lambda = \lambda' \Omega_m \lambda$; and (ii) $n^{-1/2} \max_{1 \leqslant t \leqslant n} |Z_t| \xrightarrow{P} 0$. To see (i), note that $n^{-1} \sum_{t=1}^n [Z_t^2 - E(Z_t^2)] = A_1 + A_2$, with

$$A_1 = n^{-1} \sum_{t=1}^n [Z_t^2 - E(Z_t^2|\mathscr{F}_{t-1})] \quad \text{and} \quad A_2 = n^{-1} \sum_{t=1}^n [E(Z_t^2|\mathscr{F}_{t-1}) - E(Z_t^2)].$$

First consider A_1. Since $\{Z_t^2 - E(Z_t^2|\mathscr{F}_{t-1}), \mathscr{F}_t\}$ is a m.d.s., we have that $Z_t^2 - E(Z_t^2|\mathscr{F}_{t-1})$ is an L_1-mixingale with mixingale constants $c_t = E|Z_t^2 - E(Z_t^2|\mathscr{F}_{t-1})|$: $E|E(Z_t^2 - E(Z_t^2|\mathscr{F}_{t-1})|\mathscr{F}_{t-k})| \leqslant c_t \xi_k$, $k = 0, 1, \ldots$, with $\xi_k = 1$ for $k = 0$ and $\xi_k = 0$ otherwise. Thus, we apply Andrews' LLN for L_1-mixingales (Andrews, 1988) to show $A_1 \xrightarrow{P} 0$. It suffices that for some $r > 1$, $E|Z_t^2|^r \leqslant K < \infty$ and $n^{-1} \sum_{t=1}^n c_t < \infty$. Now, $E|Z_t|^{2r} = E|\sum_{i=1}^m \lambda_i \varepsilon_t \varepsilon_{t-i}|^{2r} \leqslant (\sum_{i=1}^m |\lambda_i| \|\varepsilon_t \varepsilon_{t-i}\|_{2r})^{2r} < K$ by repeated application of Minkowski and Cauchy–Schwartz, given Assumption A(vi). The second condition on $\{c_t\}$ follows similarly. Next we consider A_2. We have that

$$A_2 = \lambda' n^{-1} \sum_{t=1}^n (E(W_t W_t'|\mathscr{F}_{t-1}) - E(W_t W_t')) \lambda$$

$$= \lambda' \left[n^{-1} \sum_{t=1}^n \varepsilon_{t-i} \varepsilon_{t-j} E(\varepsilon_t^2|\mathscr{F}_{t-1}) - \sigma^4 \tau_{i,j} \right]_{i,j=1,\ldots,p} \lambda \xrightarrow{P} 0,$$

given Assumption A(v). This proves (i). To prove (ii), note that by Markov's inequality, for any $\delta > 0$ and for some $r > 1$,

$$P\left(\frac{1}{\sqrt{n}} \max_{1 \leqslant t \leqslant n} |Z_t| > \delta\right) \leqslant \sum_{t=1}^{n} P(|Z_t| > n^{1/2}\delta) \leqslant \delta^{-2r} n^{-r} \sum_{t=1}^{n} \mathrm{E}|Z_t|^{2r}$$

$$\leqslant K\delta^{-2r} n^{1-r} \to 0. \qquad \square$$

Proof of Lemma A.2. First we consider (i) with $j = 0$, without loss of generality. By definition, $\hat{\varepsilon}_t^* \equiv \hat{\varepsilon}_t \eta_t$, and thus

$$n^{-1}\sum_{t=1}^{n} \hat{\varepsilon}_t^{*2} - \sigma^2 = \left[n^{-1}\sum_{t=1}^{n} \hat{\varepsilon}_t^2(\eta_t^2 - 1)\right] + \left[n^{-1}\sum_{t=1}^{n} \hat{\varepsilon}_t^2 - \sigma^2\right] \equiv F_1^* + F_2,$$

with the obvious definitions. Under our assumptions $F_2 = \mathrm{o}_P(1)$. So it suffices to show that $P^*[|F_1^*| > \delta] = \mathrm{o}_P(1)$, for any $\delta > 0$, or, by Chebyshev's inequality, that $\mathrm{E}^*((F_1^*)^2) = \mathrm{o}_P(1)$. Let $z_t^* \equiv \hat{\varepsilon}_t^2(\eta_t^2 - 1)$ and note that $\mathrm{E}^*(z_t^* z_s^*) = 0$ for $t \neq s$, $\mathrm{E}^*(z_t^{*2}) = \hat{\varepsilon}_t^4 \mathrm{E}^*(\eta_t^4 - 2\eta_t^2 + 1) = \hat{\varepsilon}_t^4(\mathrm{E}^*(\eta_t^4) - 1)$. Thus,

$$\mathrm{E}^*[(F_1^*)^2] = \mathrm{E}^*\left(n^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n} z_t^* z_s^*\right) = n^{-1}\left(n^{-1}\sum_{t=1}^{n} \hat{\varepsilon}_t^4(\mathrm{E}^*(\eta_t^4) - 1)\right)$$

$$\leqslant n^{-1}K\left(n^{-1}\sum_{t=1}^{n} \hat{\varepsilon}_t^4\right) = \mathrm{o}_P(1),$$

where the last inequality holds by $\mathrm{E}^*(\eta_t^4) \leqslant \Delta < \infty$ and $n^{-1}\sum_{t=1}^{n} \hat{\varepsilon}_t^4 = \mathrm{O}_P(1)$, given that $\mathrm{E}|\varepsilon_t|^4 < K < \infty$ and that $\hat{\phi} \to \phi$ in probability. For (ii), by a similar reasoning, it suffices to note that

$$\mathrm{E}^*\left[\left(n^{-1}\sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^*\right)^2\right] = n^{-2}\sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^2 \hat{\varepsilon}_t^2 \mathrm{E}^*(\eta_t^2 \eta_{t-j}^2) = n^{-2}\sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^2 \hat{\varepsilon}_t^2 = \mathrm{o}_P(1).$$

For (iii), note that

$$n^{-1}\sum_{t=\max(i,j)+1}^{n} \hat{\varepsilon}_{t-i}^* \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^{*2} - \sigma^4 \tau_{i,j} 1(i=j)$$

$$= n^{-1}\sum_{t=\max(i,j)+1}^{n} \hat{\varepsilon}_{t-i} \hat{\varepsilon}_{t-j} \hat{\varepsilon}_t^2 (\eta_t^2 \eta_{t-i} \eta_{t-j} - 1(i=j))$$

$$+ n^{-1}\sum_{t=\max(i,j)+1}^{n} (\hat{\varepsilon}_{t-i} \hat{\varepsilon}_{t-j} \hat{\varepsilon}_t^2 - \sigma^4 \tau_{ij})1(i=j) \equiv G_1^* + G_2.$$

Under our assumptions, for any fixed i, j,

$$n^{-1}\sum_{t=\max(i,j)+1}^{n} \hat{\varepsilon}_{t-i} \hat{\varepsilon}_{t-j} \hat{\varepsilon}_t^2 = n^{-1}\sum_{t=\max(i,j)+1}^{n} \varepsilon_{t-i} \varepsilon_{t-j} \varepsilon_t^2 + R_n,$$

where the remainder R_n involves products of elements of $\hat{\phi} - \phi$, which are $o_P(1)$ under our assumptions, with averages of products of elements of Y_{t-1-j} and ε_t, up to the fourth order, which are bounded in probability, given that $E|\varepsilon_t|^4 < \Delta < \infty$. Thus, $R_n = o_P(1)$, and since $n^{-1} \sum_{t=\max(i,j)+1}^{n} \varepsilon_{t-i} \varepsilon_{t-j} \varepsilon_t^2 \to \sigma^4 \tau_{i,j}$ (cf. proof of Lemma A.1), it follows that $G_2 = o_p(1)$. So, if we let $z_t^{*(i,j)} = \hat{\varepsilon}_{t-i} \hat{\varepsilon}_{t-j} \hat{\varepsilon}_t^2 (\eta_{t-i} \eta_{t-j} \eta_t^2 - 1(i=j))$, it suffices that $P^*(|G_1^*| > \delta) = o_P(1)$ for any $\delta > 0$. But

$$P^*(|G_1^*| > \delta) \leqslant \frac{1}{\delta^2 n^2} E^* \left[\sum_{t=\max(i,j)+1}^{n} \sum_{s=\max(i,j)+1}^{n} E^*(z_t^{*(i,j)} z_s^{*(i,j)}) \right]$$

$$= \frac{1}{\delta^2 n^2} \sum_{t=\max(i,j)+1}^{n} \hat{\varepsilon}_{t-i}^2 \hat{\varepsilon}_{t-j}^2 \hat{\varepsilon}_t^4 E^*[(\eta_{t-i} \eta_{t-j} \eta_t^2 - 1(i=j))^2]$$

$$\leqslant \frac{K}{\delta^2 n} \left(n^{-1} \sum_{t=\max(i,j)+1}^{n} \hat{\varepsilon}_{t-i}^2 \hat{\varepsilon}_{t-j}^2 \hat{\varepsilon}_t^4 \right),$$

where the equality holds because $E^*(z_t^{*(i,j)} z_s^{*(i,j)}) = 0$ for $s \neq t$ by the properties of $\{\eta_t\}$, and the second inequality uses the fact that $E^*|\eta_t|^4 < \Delta < \infty$. Under Assumption A strengthened by A'(vi'), we can show that $n^{-1} \sum_{t=\max(i,j)+1}^{n} \hat{\varepsilon}_{t-i}^2 \hat{\varepsilon}_{t-j}^2 \hat{\varepsilon}_t^4 = O_P(1)$, which implies that $P^*(|G_1^*| > \delta) = o_P(1)$. In fact, given that $\hat{\varepsilon}_t = \varepsilon_t - (\hat{\phi} - \phi)' Y_{t-1}$, it follows that $n^{-1} \sum_{t=\max(i,j)+1}^{n} \hat{\varepsilon}_{t-i}^2 \hat{\varepsilon}_{t-j}^2 \hat{\varepsilon}_t^4 = n^{-1} \sum_{t=\max(i,j)+1}^{n} \varepsilon_{t-i}^2 \varepsilon_{t-j}^2 \varepsilon_t^4 + o_P(1)$. In particular, the remainder contains terms involving products of elements of $\hat{\phi} - \phi$ (which are $o_P(1)$) with terms involving averages of cross products of elements of Y_{t-1-j} and ε_t, up to the eighth order, which are $O_P(1)$, given $E|\varepsilon_t|^8 \leqslant \Delta < \infty$. The latter assumption also ensures that $n^{-1} \sum_{t=\max(i,j)+1}^{n} \varepsilon_{t-i}^2 \varepsilon_{t-j}^2 \varepsilon_t^4 = O_P(1)$, by an application of the Markov and Cauchy–Schwartz inequalities. $\square$

Proof of Lemma A.3. Let $\mathscr{F}_t^* = \sigma(\eta_t, \eta_{t-1}, \ldots, \eta_1)$, and define $W_t^* = (\hat{\varepsilon}_t^* \hat{\varepsilon}_{t-1}^*, \ldots, \hat{\varepsilon}_t^* \hat{\varepsilon}_{t-m}^*)'$. Conditional on the original sample, we have $E^*(W_t^* | \mathscr{F}_{t-1}^*) = E^*(\hat{\varepsilon}_t^* | \mathscr{F}_{t-1}^*)(\hat{\varepsilon}_{t-1}^*, \ldots, \hat{\varepsilon}_{t-m}^*)' = 0$ since $E^*(\hat{\varepsilon}_t^* | \mathscr{F}_{t-1}^*) = E^*(\hat{\varepsilon}_t \eta_t | \mathscr{F}_{t-1}^*) = \hat{\varepsilon}_t E^*(\eta_t | \mathscr{F}_{t-1}^*) = 0$, where $E^*(\eta_t | \mathscr{F}_{t-1}^*) = E^*(\eta_t) = 0$, by the independence and mean zero properties of $\{\eta_t\}$. Thus, $\{W_t^*, \mathscr{F}_t^*\}$ is a vector m.d.s. We now apply Theorem A.1 to $Z_t^* = \lambda' W_t^*$ for arbitrary $\lambda \in \mathbb{R}^m$, $\lambda' \lambda = 1$. First, note that $\bar{\sigma}_n^{*2} \equiv Var^*(n^{-1/2} \sum_{t=m+1}^{n} Z_t^*) = \lambda' n^{-1} \sum_{t=m+1}^{n} E^*(W_t^* W_t^{*'}) \lambda \equiv \lambda' \Omega_{n,m}^* \lambda$, where by direct evaluation and using the independence and zero mean properties of $\{\eta_t\}$,

$$\Omega_{n,m}^* = diag \left(n^{-1} \sum_{t=m+1}^{n} \hat{\varepsilon}_t^2 \hat{\varepsilon}_{t-1}^2, \ldots, n^{-1} \sum_{t=m+1}^{n} \hat{\varepsilon}_t^2 \hat{\varepsilon}_{t-m}^2 \right).$$

Under our assumptions, we can show that $n^{-1} \sum_{t=m+1}^{n} \hat{\varepsilon}_t^2 \hat{\varepsilon}_{t-i}^2 \xrightarrow{P} \sigma^4 \tau_{i,i}$, $i = 1, \ldots, m$, which implies $\Omega_{n,m}^* \xrightarrow{P} \tilde{\Omega}_m \equiv \sigma^4 diag(\tau_{1,1}, \ldots, \tau_{m,m})$. Thus, to verify the first condition of

the CLT it suffices that

$$\lambda' \left[n^{-1} \sum_{t=m+1}^{n} W_t^* W_t^{*\prime} - \tilde{\Omega}_m \right] \lambda \equiv \lambda' V_n^* \lambda \overset{P^*}{\to} 0, \text{ in probability.}$$

A typical element (k,l) of the middle matrix V_n^* is given by

$$(V_n^*)_{k,l} \equiv n^{-1} \sum_{t=m+1}^{n} \hat{\varepsilon}_{t-k}^* \hat{\varepsilon}_{t-l}^* \hat{\varepsilon}_t^{*2} - \sigma^4 \tau_{k,l} 1 \ (k=l),$$

where by Lemma A.2 (iii), under Assumption A strengthened by A′(vi′), we have that $(V_n^*)_{k,l} \overset{P^*}{\to} 0$ in probability. Lastly, condition 2 holds if for some $r > 1$, $n^{-r} \sum_{t=m+1}^{n} \mathrm{E}^* |\lambda' W_t^*|^{2r} = o_P(1)$. We take $r = 2$. By the c_r-inequality, we have

$$n^{-r} \sum_{t=m+1}^{n} \mathrm{E}^* |\lambda' W_t^*|^{2r} = n^{-r} \sum_{t=m+1}^{n} \mathrm{E}^* \left| \sum_{i=1}^{m} \lambda_i \hat{\varepsilon}_t^* \hat{\varepsilon}_{t-i}^* \right|^{2r}$$

$$\leqslant m^{2r-1} \sum_{i=1}^{m} |\lambda_i|^{2r} n^{-r} \sum_{t=m+1}^{n} \mathrm{E}^* |\hat{\varepsilon}_t^* \hat{\varepsilon}_{t-i}^*|^{2r}$$

$$\leqslant n^{-(r-1)} m^{2r-1} \sum_{i=1}^{m} |\lambda_i|^{2r} n^{-1} \sum_{t=m+1}^{n} |\hat{\varepsilon}_t \hat{\varepsilon}_{t-i}|^{2r}$$

$$\times \mathrm{E}^* |\eta_t|^{2r} \mathrm{E}^* |\eta_{t-i}|^{2r} = o_P(1),$$

given in particular that $n^{-1} \sum_{t=m+1}^{n} |\hat{\varepsilon}_t \hat{\varepsilon}_{t-i}|^{2r} = O_P(1)$. $\square$

Proof of Lemma A.4. We can write $y_t^* = \sum_{j=0}^{t-1} \hat{\psi}_j \hat{\varepsilon}_{t-j}^*$, $t = 1,\ldots,n$, where $\{\hat{\psi}_j\}$ is defined by $\hat{\psi}_j = \sum_{i=1}^{\min(j,p)} \hat{\phi}_i \hat{\psi}_{j-1}$, with $\hat{\psi}_0 = 1$ and $\hat{\psi}_j = 0$ for $j < 0$. It follows that $Y_{t-1}^* = \sum_{j=1}^{t-1} \hat{b}_j \hat{\varepsilon}_{t-j}^*$, for $t = 2,\ldots,n$, where $\hat{b}_j = (\hat{\psi}_{j-1},\ldots,\hat{\psi}_{j-p})'$. Note that for $t = 1$, $Y_{t-1}^* = Y_0^* = 0$, given the zero initial conditions. Hence,

$$n^{-1} \sum_{t=1}^{n} Y_{t-1}^* Y_{t-1}^{*\prime} = T_{1n}^* + T_{2n}^*, \quad \text{with } T_{1n}^* = \sum_{j=1}^{n-1} \hat{b}_j \hat{b}_j' \left(n^{-1} \sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^{*2} \right), \quad \text{and}$$

$$T_{2n}^* = \sum_{k=1}^{n-2} \sum_{j=1}^{n-k-1} (\hat{b}_j \hat{b}_{j+k}' + \hat{b}_{j+k} \hat{b}_j') \left(n^{-1} \sum_{t=1+k}^{n-j} \hat{\varepsilon}_{t-k}^* \hat{\varepsilon}_t^* \right).$$

Next, we show: (a) $T_{1n}^* \xrightarrow{P^*} A \equiv \sigma^2 \sum_{j=1}^{\infty} b_j b_j'$, and (b) $T_{2n}^* \xrightarrow{P^*} 0$, in probability. To prove (a), consider for fixed $m \in \mathbb{N}$,

$$T_{1n}^* = T_{1n}^{*m} + R_{1n}^{*m}, \quad \text{with } T_{1n}^{*m} = \sum_{j=1}^{m-1} \hat{b}_j \hat{b}_j' \left(n^{-1} \sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^{*2} \right), \quad \text{and}$$

$$R_{1n}^{*m} = \sum_{j=m}^{n-1} \hat{b}_j \hat{b}_j' \left(n^{-1} \sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^{*2} \right).$$

By Lemma A.2(i), for each $j = 1, \ldots, m$, m fixed, $n^{-1} \sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^{*2} \xrightarrow{P^*} \sigma^2$, in probability; also, under Assumption A, $\hat{\psi}_j \xrightarrow{P} \psi_j$, implying $\hat{b}_j \xrightarrow{P} b_j$. Thus, by Slutsky's theorem, $T_{1n}^{*m} \xrightarrow{P^*} \sum_{j=1}^{m-1} b_j b_j' \sigma^2 \equiv A_m$, in probability. Since $\{\psi_j\}$ is absolutely summable, it follows that $A_m \to A$ as $m \to \infty$. Thus, $T_{1n}^{*m} \xrightarrow{P^*} A$, in probability. Choose $\lambda \in \mathbb{R}^p$ arbitrarily such that $\lambda' \lambda = 1$. By BD's Proposition 6.3.9, it now suffices to show that, for any $\delta > 0$, $\lim_{m \to \infty} \lim \sup_{n \to \infty} P^*(|\lambda' R_{1n}^{*m} \lambda| > \delta) = 0$, in probability, or $\lim_{m \to \infty} \lim \sup_{n \to \infty} \mathrm{E}^* (|\lambda' R_{1n}^{*m} \lambda|) = 0$, in probability, by Markov's inequality. Using the triangle inequality and the properties of $\{\eta_t\}$, we get

$$\mathrm{E}^*(|\lambda' R_{1n}^{*m} \lambda|) \leqslant \sum_{j=m}^{n-1} |\lambda' \hat{b}_j \hat{b}_j' \lambda| \mathrm{E}^* \left(n^{-1} \sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^{*2} \right) = \sum_{j=m}^{n-1} |\lambda' \hat{b}_j \hat{b}_j' \lambda| n^{-1} \sum_{t=j+1}^{n} \hat{\varepsilon}_{t-j}^2$$

$$\leqslant \left(n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2 \right) \left(\sum_{j=m}^{n-1} |\lambda' \hat{b}_j \hat{b}_j' \lambda| \right).$$

Given that $\hat{\varepsilon}_t = \varepsilon_t - (\hat{\phi} - \phi)' Y_{t-1}$, and that $\hat{\phi} - \phi \xrightarrow{P} 0$, we can show $n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2 = \mathrm{O}_P(1)$. Thus,

$$\mathrm{E}^*(|\lambda' R_{1n}^{*m} \lambda|) \leqslant \mathrm{O}_P(1) \sum_{j=m}^{n-1} |\lambda' \hat{b}_j \hat{b}_j' \lambda| \leqslant \mathrm{O}_P(1) \sum_{k=1}^{p} \sum_{l=1}^{p} |\lambda_k \lambda_l| \sum_{j=m}^{\infty} |\hat{\psi}_{j-k} \hat{\psi}_{j-l}|.$$

Under our assumptions, $\sum_{j=1}^{p} |\hat{\phi}_j - \phi_j| = \mathrm{o}_P(1)$, so there exists n_1 such that $\sup_{n \geqslant n1} \sum_{j=1}^{\infty} |\hat{\psi}_j| < \infty$ in probability (cf. Bühlmann, 1995, Lemma 2.2.). This implies $\sup_{n \geqslant n_1} \sum_{j=m}^{\infty} |\hat{\psi}_{j-k} \hat{\psi}_{j-l}| = \mathrm{o}_P(1)$ as $m \to \infty$, which completes the proof that $T_{1n}^* \xrightarrow{P^*} A$, in probability. Finally, to show (b), consider first for fixed $m \in \mathbb{N}$, $T_{2n}^{*m} = \sum_{k=1}^{m-2} \sum_{j=1}^{m-k-1} (\hat{b}_j \hat{b}_{j+k}' + \hat{b}_{j+k} \hat{b}_j') (n^{-1} \sum_{t=1+k}^{n-j} \hat{\varepsilon}_{t-k}^* \hat{\varepsilon}_t^*)$. For fixed j and k, it follows by Lemma A.2(ii) that $n^{-1} \sum_{t=1+k}^{n-j} \hat{\varepsilon}_{t-k}^* \hat{\varepsilon}_t^* \xrightarrow{P^*} 0$, in probability. Since $\hat{b}_j \hat{b}_{j+k}' + \hat{b}_{j+k} \hat{b}_j' \xrightarrow{P} b_j b_{j+k}' + b_{j+k} b_j'$, we have that $T_{2n}^{*m} \xrightarrow{P^*} 0$, in probability. To complete the proof of (b) we need to show

that each of the following

$$R_{2,1n}^{*m} = \sum_{k=m-1}^{n-1} \sum_{j=1}^{n-k-1} (\hat{b}_j \hat{b}_{j+k}' + \hat{b}_{j+k} \hat{b}_j') \left(n^{-1} \sum_{t=1+k}^{n-j} \hat{\varepsilon}_{t-k}^* \hat{\varepsilon}_t^* \right), \quad \text{and}$$

$$R_{2,2n}^{*m} = \sum_{k=1}^{m-2} \sum_{j=m-k}^{n-k-1} (\hat{b}_j \hat{b}_{j+k}' + \hat{b}_{j+k} \hat{b}_j') \left(n^{-1} \sum_{t=1+k}^{n-j} \hat{\varepsilon}_{t-k}^* \hat{\varepsilon}_t^* \right),$$

satisfies the condition $\lim_{m\to\infty} \limsup_{n\to\infty} P^*(|\lambda' R_{2,in}^{*m} \lambda| > \delta) = 0$ in probability, for $i = 1, 2$, where λ and δ are as above. This can be verified analogously. □

Proof of Lemma A.5. As in the proof of Lemma A.4, we have $Y_{t-1}^* = \sum_{j=1}^{t-1} \hat{b}_j \hat{\varepsilon}_{t-j}^*$, where $\hat{b}_j = (\hat{\psi}_{j-1}, \ldots, \hat{\psi}_{j-p})'$, with $\hat{\psi}_0 = 1$ and $\hat{\psi}_j = 0$ for $j < 0$. Noting that $Y_0^* = 1$,

$$n^{-1/2} \sum_{t=1}^n Y_{t-1}^* \hat{\varepsilon}_t^* = n^{-1/2} \sum_{t=2}^n \sum_{j=1}^{t-1} \hat{b}_j \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^* = \sum_{j=1}^{n-1} \hat{b}_j n^{-1/2} \sum_{t=j+1}^n \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^* \equiv \chi_n^*.$$

For fixed $m \in \mathbb{N}$, let $\chi_{n,m}^* \equiv \sum_{j=1}^{m-1} \hat{b}_j n^{-1/2} \sum_{t=j+1}^n \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^*$. Next we show: (a) for m fixed, $\chi_{n,m}^* \Rightarrow^{d_{P^*}} N(0, \tilde{B}_m)$, as $n \to \infty$, where $\tilde{B}_m = \sum_{j=1}^m b_j b_j' \sigma^4 \tau_{j,j}$; (b) $\tilde{B}_m \to \tilde{B}$ as $m \to \infty$, and (c) $\lim_{m\to\infty} \limsup_{n\to\infty} P^*(|\chi_n^* - \chi_{n,m}^*| > \delta) = 0$ for any $\delta > 0$. For (a), write

$$\chi_{n,m}^* = \sum_{j=1}^{m-1} b_j n^{-1/2} \sum_{t=j+1}^n \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^* + \sum_{j=1}^{m-1} (\hat{b}_j - b_j) n^{-1/2} \sum_{t=j+1}^n \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^* \equiv Q_1^* + Q_2^*.$$

By Lemma A.3, under Assumption A strengthened by A(vi'), $Q_1^* \Rightarrow^{d_{P^*}} N(0, \tilde{B}_{m-1})$, in probability, where $\tilde{B}_{m-1} = \sum_{j=1}^{m-1} b_j b_j' \sigma^4 \tau_{j,j}$. Next, note that $Q_2^* \xrightarrow{P^*} 0$ in probability, since $\hat{b}_j - b_j \xrightarrow{P} 0$ and $n^{-1/2} \sum_{t=j+1}^n \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^* = O_{P^*}(1)$ for each $j = 1, \ldots, m-1$. The asymptotic equivalence lemma now implies (a). (b) follows by dominated convergence given the summability of $\{\psi_j\}$ and the uniform boundedness of $\sigma^4 \tau_{j,j}$. To prove (c), note that it suffices to show that $\lim_{m\to\infty} \limsup_{n\to\infty} E^*(|\chi_n^* - \chi_{n,m}^*|^2) = o_P(1)$, by Chebyshev's inequality. Equivalently, we consider for any $\lambda \in \mathbb{R}^p$, such that $\lambda' \lambda = 1$,

$$E^*(|\lambda'(\chi_n^* - \chi_{n,m}^*)|^2) = E^* \left(\sum_{j=m}^{n-1} \sum_{i=m}^{n-1} \lambda' \hat{b}_j \hat{b}_i' \lambda Z_{nj}^* Z_{ni}^* \right),$$

where $Z_{nj}^* \equiv n^{-1/2} \sum_{t=j+1}^n \hat{\varepsilon}_{t-j}^* \hat{\varepsilon}_t^*$. Since $E^*(Z_{nj}^* Z_{ni}^*) = 0$ for $i \neq j$ and $E^*(Z_{nj}^{*2}) = n^{-1} \sum_{t=j+1}^n \hat{\varepsilon}_{t-j}^2 \hat{\varepsilon}_t^2$, it follows that

$$E^*(|\lambda'(\chi_n^* - \chi_{n,m}^*)|^2) = \sum_{j=m}^{n-1} \lambda' \hat{b}_j \hat{b}_j' \lambda \left(n^{-1} \sum_{t=j+1}^n \hat{\varepsilon}_{t-j}^2 \hat{\varepsilon}_t^2 \right)$$

$$\leqslant \left(n^{-1} \sum_{t=1}^n \hat{\varepsilon}_t^4 \right) \left(\sum_{j=m}^{n-1} \lambda' \hat{b}_j \hat{b}_j' \lambda \right),$$

where the last inequality holds by an application of the Cauchy–Schwartz inequality. Using the definition of $\hat{\varepsilon}_t$, i.e., $\hat{\varepsilon}_t = \varepsilon_t \eta_t - (\hat{\phi} - \phi)' Y_{t-1}$, and the fact that $\hat{\phi} - \phi \xrightarrow{P} 0$, we can show that $n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^4 = O_P(1)$. The proof of (c) now follows exactly the argument used in Lemma A.4 when dealing with R_{1n}^{*m}. $\square$

References

Andrews, D.W.K., 1988. Laws of large numbers for dependent nonidentically distributed random variables. Econometric Theory 4, 458–467.

Bekaert, G., Hodrick, R.J., 2001. Expectations hypothesis tests. Journal of Finance 56, 1357–1394.

Bekaert, G., Hodrick, R.J., Marshall, D.A., 1997. On biases in tests of the expectations hypothesis in the term structure of interest rates. Journal of Financial Economics 44, 309–348.

Berkowitz, J., Kilian, L., 2000. Recent developments in bootstrapping time series. Econometric Reviews 19, 1–48.

Berkowitz, J, Birgean, I., Kilian L., 2000. On the finite-sample accuracy of nonparametric resampling algorithms for economic time series. In: Fomby, T.B., Hill, C. (Eds.), Advances in Econometrics: Applying Kernel and Nonparametric Estimation to Economic Topics, Vol. 14. JAI Press, CT, 77–107.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 37, 307–327.

Bollerslev, T., 1990. Modeling the coherence on short-run nominal exchange rates: a multivariate GARCH model. Review of Economics and Statistics 72, 498–505.

Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. A capital-asset pricing model with time-varying covariances. Journal of Political Economy 96, 116–131.

Bose, A., 1988. Edgeworth correction by bootstrap in autoregression. Annals of Statistics 16, 1709–1722.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, 2nd Edition. Springer, New York.

Bühlmann, P., 1995. Moving-average representation for autoregressive approximations. Stochastic Processes and Their Applications 60, 331–342.

Chesher, A., Jewitt, I., 1987. The bias of a heteroskedasticity consistent covariance matrix estimator. Econometrica 55, 1217–1222.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, New York.

Davidson, R., Flachaire, E., 2001. The wild bootstrap, tamed at last. Working Paper, Darp58, STICERD, London School of Economics.

Deo, S.R., 2000. Spectral tests of the martingale hypothesis under conditional heteroskedasticity. Journal of Econometrics 99, 291–315.

Durrett, R., 1996. Probability: Theory and Examples. Duxury Press, California.

Eicker, F., 1963. Asymptotic normality and consistency of the least squares estimators for families of linear regressions. Annals of Mathematical Statistics 34, 447–456.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Engle, R.F., 1990. Discussion: stock market volatility and the crash of '87. Review of Financial Studies 3, 103–106.

Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.

Freedman, D.A., 1981. Bootstrapping regression models. Annals of Statistics 9, 1218–1228.

Glosten, L.R., Jaganathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of nominal excess returns on stocks. Journal of Finance 48, 1779–1801.

Godfrey, L.G., Orme, C.D., 2001. Significance levels of heteroskedasticity-robust tests for specification and misspecification: some results on the use of the wild bootstraps. Economics Letters, forthcoming.

Goetzmann, W.N., Jorion, P., 1993. Testing the predictive power of dividend yields. Journal of Finance 48, 663–679.

Goetzmann, W.N., Jorion, P., 1995. A longer look at dividend yields. Journal of Business 68, 483–508.

Gonçalves, S., Kilian, L., 2003. Asymptotic and bootstrap inference for $AR(\infty)$ processes with conditional heteroskedasticity. CIRANO Working Paper 2003-s28.

Hansen, B.E., 1999. The grid bootstrap and the autoregressive model. Review of Economics and Statistics 81, 594–607.

Hansen, B.E., 2000. Testing for structural change in conditional models. Journal of Econometrics 97, 93–115.

Härdle, W., Horowitz, J., Kreiss, J.-P., 2001. Bootstrap methods for time series. Manuscript, Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin.

He, C., Teräsvirta, T., 1999. Properties of moments of a family of GARCH processes. Journal of Econometrics 92, 173–192.

Hodrick, R.J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. Review of Financial Studies 5, 357–386.

Inoue, A., Kilian, L., 2002. Bootstrapping autoregressions with possible unit roots. Econometrica 70, 377–391.

Kreiss, J.P., 1997. Asymptotic properties of residual bootstrap for autoregressions. Manuscript, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.

Kuersteiner, G.M., 2001. Optimal instrumental variables estimation for ARMA models. Journal of Econometrics 104, 359–405.

Ledoit, O., Santa-Clara, P., Wolf, M., 2001. Multivariate GARCH modeling with an application to international stock markets. Working Paper No. 578, Department of Economics, Universitat Pompeu Fabra.

Liu, R.Y., 1988. Bootstrap procedure under some non-i.i.d. models. Annals of Statistics 16, 1696–1708.

Lütkepohl, H., 1990. Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models. Review of Economics and Statistics 72, 116–125.

MacKinnon, J.G., White, H., 1985. Some heteroskedasticity consistent covariance matrix estimators with improved finite-sample properties. Journal of Econometrics 29, 305–325.

Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models. Annals of Statistics 21, 255–285.

Milhøj, A., 1985. The moment structure of ARCH processes. Scandinavian Journal of Statistics 12, 281–292.

Nicholls, D.F., Pagan, A.R., 1983. Heteroskedasticity in models with lagged dependent variables. Econometrica 51, 1233–1242.

Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., et al. (Ed.), Time Series Models in Econometrics, Finance and Other Fields. Chapman & Hall, London.

Weiss, A.A., 1988. ARMA models with ARCH errors. Journal of Time Series Analysis 5, 129–143.

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

White, H., 1999. Asymptotic Theory for Econometricians, 2nd Edition. Academic Press, London.

Wu, C.F.J., 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics 14, 1261–1295.

Wolf, M., 2000. Stock returns and dividend yields revisited: a new way to look at an old problem. Journal of Business and Economic Statistics 18, 18–30.