Econometric Reviews

Asymptotic and Bootstrap Inference for AR(∞) Processes with Conditional Heteroskedasticity

Sílvia Gonçalves [a]; Lutz Kilian [b]
[a] Départment de Sciences Économiques, CIREQ and CIRANO, Université de Montréal, Montreal, Québéc, Canada
[b] Department of Economics, University of Michigan, Ann Arbor, Michigan, USA and CEPR

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

ASYMPTOTIC AND BOOTSTRAP INFERENCE FOR AR(∞) PROCESSES WITH CONDITIONAL HETEROSKEDASTICITY

Sílvia Gonçalves □ *Départment de Sciences Économiques, CIREQ and CIRANO, Université de Montréal, Montreal, Québec, Canada*

Lutz Kilian □ *Department of Economics, University of Michigan, Ann Arbor, Michigan, USA and CEPR*

□ *The main contribution of this paper is a proof of the asymptotic validity of the application of the bootstrap to AR(∞) processes with unmodelled conditional heteroskedasticity. We first derive the asymptotic properties of the least-squares estimator of the autoregressive sieve parameters when the data are generated by a stationary linear process with martingale difference errors that are possibly subject to conditional heteroskedasticity of unknown form. These results are then used in establishing that a suitably constructed bootstrap estimator will have the same limit distribution as the least-squares estimator. Our results provide theoretical justification for the use of either the conventional asymptotic approximation based on robust standard errors or the bootstrap approximation of the distribution of autoregressive parameters. A simulation study suggests that the bootstrap approach tends to be more accurate in small samples.*

1. INTRODUCTION

Much applied work relies on linear autoregressions for the purpose of estimation and inference in time series analysis (see, e.g., Canova, 1995; Stock and Watson, 2001). Standard methods of inference for linear autoregressions are based on the presumption that the data generating process can be represented as a finite-order autoregression. This assumption is clearly unrealistic. It is more plausible to think of the autoregressive model as a rough approximation to the underlying data generating process (see, e.g., Braun and Mittnik, 1993; Ng and

Perron, 1995; Lütkepohl and Saikkonen, 1997; Inoue and Kilian, 2002). In response to this problem, there has been increasing interest in developing an alternative asymptotic theory for inference in linear autoregressions under the presumption that the data are generated by a possibly infinite-order autoregression. The thought experiment is that the researcher fits a sequence of finite-order autoregressions, the lag order of which is assumed to increase with the sample size. The fitted autoregression is thus viewed as an approximation to the possibly infinite-order autoregression, the quality of which improves with the sample size. Such methods are commonly referred to as sieve methods in the literature (see, e.g., Grenander, 1981; Geman and Hwang, 1982; Bühlmann, 1995, 1997).

The sieve approach has a long tradition in econometrics. For example, Ng and Perron (1995) study problems of lag order selection for sieve autoregressions by sequential *t*-tests. Diebold and Kilian (2001) and Galbraith (2003) consider inference about measures of predictability based on autoregressive sieves. Lütkepohl (1988a) derives the asymptotic distribution of the estimated dynamic multipliers. Lütkepohl and Poskitt (1991) extend these results to orthogonalized impulse response estimates and forecast error decompositions. Lütkepohl and Poskitt (1996) propose tests of Granger causality in the context of infinite-order autoregressions. Lütkepohl (1988b) investigates tests of structural instability for autoregressive sieves. Inoue and Kilian (2002) propose the use of the sieve approximation in evaluating the fit of dynamic macroeconomic models. The sieve approach has also proved useful in studying cointegrated processes, building on the framework developed by Saikkonen (1992) who proposed to approximate cointegrated linear systems with i.i.d. innovations via autoregressive sieves (also see Saikkonen and Lütkepohl, 1996; Lütkepohl and Saikkonen, 1997; Saikkonen and Luukkonen, 1997).

Sieve approximations are valid for all data generating processes that belong to the stationary linear $AR(\infty)$ class. This class includes finite-order stationary invertible ARMA processes as a special case. Compared to fitting parametric ARMA models, an important advantage of the autoregressive sieve approximation is that it retains the computational simplicity of the parametric finite lag-order autoregressive model, while allowing for model misspecification within the class of linear models. This is especially important in the vector case. The specification of an identified vector ARMA structure is much more complicated than setting up a pure VAR model and the estimation of vector ARMA models is subject to numerical difficulties. For these reasons, autoregressive models are the preferred models in most empirical studies in macroeconometrics.

In the literature on autoregressive sieve approximations of linear processes, it is typically postulated that the data generating process can be represented as an infinite-order autoregression with i.i.d. innovations.

Although this model is substantially less restrictive than the conventional finite-lag order autoregressive model, for many applications in finance and economics the i.i.d. error assumption appears too restrictive. In particular, the i.i.d. error assumption rules out conditional heteroskedasticity, of which there is evidence in many economic time series (see, e.g., Engle, 1982; Bollerslev, 1986; Weiss, 1988; Hodrick, 1992; Bekaert and Hodrick, 2001). Methods that relax the i.i.d. error assumption in fully parametric autoregressive models have been discussed for example in Kuersteiner (2001) and in Gonçalves and Kilian (2004). In this paper we relax the i.i.d. error assumption in the semiparametric autoregressive sieve model. We postulate instead that the innovations driving the linear $AR(\infty)$ process follow a martingale difference sequence (m.d.s.) subject to possible conditional heteroskedasticity of unknown form.

The main contribution of this paper is a proof of the asymptotic validity of the application of the bootstrap to $AR(\infty)$ processes with possible conditional heteroskedasticity. We also derive the consistency and asymptotic normality of the least-squares (LS) estimator of the autoregressive sieve parameters under weak conditions on the form of conditional heteroskedasticity. These results, while less innovative from a technical point of view, constitute important intermediate results used in establishing the validity of the bootstrap. In related work, Hannan and Deistler (1988, p. 335) state an asymptotic normality result for the LS estimator in the autoregressive sieve model with m.d.s. errors. Our derivation relies on stronger moment conditions, but goes beyond Hannan and Deistler (1988) in that we derive a consistent estimator of the limiting variance.

Our analysis shows that the asymptotic distribution of estimated autoregressive parameters derived under the assumption of an $AR(\infty)$ data generating process with i.i.d. errors does not apply when the errors are conditionally heteroskedastic. In particular, the form of the asymptotic covariance matrix of the estimated parameters is affected by conditional heteroskedasticity. In contrast, the asymptotic results derived in this paper enable applied users to conduct inference that is robust to conditional heteroskedasticity of unknown form. The use of the asymptotic normal approximation in practice requires a consistent estimator of the variance of the autoregressive parameter. We provide sufficient conditions for the consistency of a version of the Eicker–White heteroskedasticity-robust covariance matrix estimator in the context of sieve approximations to $AR(\infty)$ processes (see Eicker, 1963; White, 1980; Nicholls and Pagan, 1983).

These asymptotic results for the LS estimator provide the basis for inference on smooth functions of autoregressive parameters such as impulse responses and related statistics of interest in macroeconometrics and finance in the presence of unmodelled conditional heteroskedasticity.

They also are essential for establishing the asymptotic validity of the bootstrap for parameters of $AR(\infty)$ processes under these conditions. We show that suitably constructed bootstrap estimators will have the same limit distribution as the LS estimator. Providing such results is important, because the bootstrap method is already widely used in applied work involving autoregressions.

Our aim in this paper is not to establish that the sieve bootstrap approximation is superior to conventional asymptotic approximations. Rather we provide a broader set of conditions under which the bootstrap approach will be justified asymptotically, and we discuss the modifications required to make the bootstrap approach robust to conditional heteroskedasticity. We also provide simulation evidence that suggests that the bootstrap approximation tends to be more accurate in small samples than the conventional asymptotic approximation based on robust standard errors. We leave for future research the question of whether there are conditions under which the bootstrap approach will provide asymptotic refinements.

The remainder of the paper is organized as follows. In Section 2 we present the theoretical results for the LS estimator. In Section 3, we develop the theoretical results for the corresponding bootstrap estimator. Section 4 contains the results of a small simulation study designed to compare the finite-sample coverage accuracy of bootstrap confidence intervals for autoregressive coefficients and of intervals based on the asymptotic approximation. We conclude in Section 5. Details of the proofs are provided in the Appendix.

2. ASYMPTOTIC THEORY FOR THE LS ESTIMATOR

Our analysis in this section builds on work by Berk (1974), Bhansali (1978), and Lewis and Reinsel (1985). Berk (1974) in a seminal paper establishes the consistency and asymptotic normality of the spectral density estimator for linear processes with i.i.d. innovations. Based on Berk's results, Bhansali (1978) derives explicitly the limiting distribution of the estimated autoregressive coefficients. Lewis and Reinsel (1985) provide a multivariate extension of Bhansali's (1978) results in a form more suitable for econometric analysis. Here we generalize the analysis of Lewis and Reinsel (1985) by allowing for conditionally heteroskedastic martingale difference sequence errors. We use these modified results to study the asymptotic properties of the LS estimator of the autoregressive slope parameters and of the corresponding bootstrap estimator. For concreteness, we focus on univariate autoregressions. Multivariate generalizations of our results are possible at the cost of more complicated notation.

Let the time series $\{y_t, t \in \mathbb{Z}\}$ be generated from

$$y_t = \sum_{j=1}^{\infty} \phi_j y_{t-j} + \varepsilon_t, \qquad (2.1)$$

where $\Phi(z) \equiv 1 - \sum_{j=1}^{\infty} \phi_j z^j \neq 0$ for all $|z| \leq 1$, and $\sum_{j=1}^{\infty} |\phi_j| < \infty$. The AR($\infty$) data generating process (2.1) includes the class of stationary invertible ARMA(p, q) processes as a special case. For $i \in \mathbb{N}$, let $\kappa_\varepsilon(0, l_1, \ldots, l_{i-1})$ denote the ith order joint cumulant of $(\varepsilon_0, \varepsilon_{l_1}, \ldots, \varepsilon_{l_{i-1}})$ (see Brillinger, 1981, p. 19), where $l_1, \ldots, l_{i-1}$ are integers. We make the following assumption.

Assumption 1. (i) $\{\varepsilon_t\}$ is strictly stationary and ergodic such that $E(\varepsilon_t \mid \mathscr{F}_{t-1}) = 0$, a.s., 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) $E(\varepsilon_t^2) = \sigma^2 > 0$ and $E(\varepsilon_t^2 \varepsilon_{t-j}^2) > \underline{\alpha}$ for some $\underline{\alpha} > 0$ and all j; and (iii) $\sum_{l_1=-\infty}^{\infty} \cdots \sum_{l_{i-1}=-\infty}^{\infty} |\kappa_\varepsilon(0, l_1, \ldots, l_{i-1})| < \infty$, for $i = 2, \ldots, 8$.

Instead of assuming i.i.d. errors as in Lewis and Reinsel (1985), we postulate that $\{\varepsilon_t\}$ is a possibly conditionally heteroskedastic m.d.s. Assumption 1(iii) requires the absolute summability of joint cumulants of ε_t up to the eighth order, which restricts the dependence in the error process. It is implied by an α-mixing condition plus an eighth-order moment condition on ε_t (see, e.g., Remark A.1 of Künsch, 1989, or Andrews, 1991). Hong and Lee (2003) use a similar assumption in deriving tests for serial correlation that are robust to conditional heteroskedasticity of unknown form. Kuersteiner (2001) imposes a stronger version of this assumption (cf. his Assumption E1) to show consistency of a feasible version of his optimal IV estimator under conditional heteroskedasticity.

Under these assumptions, it follows that y_t has a causal infinite-order moving average representation

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

where $\psi_0 \equiv 1$, $\Psi(z) = 1/\Phi(z) = \sum_{j=0}^{\infty} \psi_j z^j$, $\sum_{j=0}^{\infty} |\psi_j| < \infty$ (see Bühlmann, 1995).

Let $\phi(k) = (\phi_1, \ldots, \phi_k)$ denote the first k autoregressive coefficients in the AR(∞) representation. Given a realization $\{y_1, \ldots, y_n\}$ of (2.1), we estimate an approximating AR(k) model by minimizing $(n-k)^{-1} \sum_{t=1+k}^{n} (y_t - \beta(k)' Y_{t-1,k})^2$, by choice of $\beta(k) = (\beta_1, \ldots, \beta_k)'$, where $Y_{t-1,k} = (y_{t-1}, \ldots, y_{t-k})'$. This yields the LS estimators

$$\hat{\phi}(k) = (\hat{\phi}_{1,k}, \ldots, \hat{\phi}_{k,k})' = \widehat{\Gamma}_k^{-1} \widehat{\Gamma}_{k,1},$$

where

$$\widehat{\Gamma}_k = (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y_{t-1,k}' \qquad \text{and} \qquad \widehat{\Gamma}_{k,1} = (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} y_t.$$

The population analogues of $\widehat{\Gamma}_k$ and $\widehat{\Gamma}_{k,1}$ are $\Gamma_k = E(Y_{t-1,k} Y_{t-1,k}')$ and $\Gamma_{k,1} = E(Y_{t-1,k} y_t)$, respectively. Our assumptions on k are identical to those employed by Lewis and Reinsel (1985).

Assumption 2. k is chosen as a function of n such that (i) $\frac{k^3}{n} \to 0$ as $k, n \to \infty$ and (ii) $n^{1/2} \sum_{j=k+1}^{\infty} |\phi_j| \to 0$ as $k, n \to \infty$.

Assumption 2 requires that $k \to \infty$ as $n \to \infty$. Assumption 2(i) stipulates that, nevertheless, k should not increase at a rate faster than $n^{1/3}$. Since $\sup_{j>k} |\phi_j| \le \sum_{j=k+1}^{\infty} |\phi_j|$, Assumption 2 (ii) implies that $\sup_{j>k} |\phi_j| = o(n^{-1/2})$ as $k \to \infty$. This assumption allows us to approximate the AR(∞) process by a sequence of finite-order AR models.

In order to state our main result in this section we need to introduce some notation. Let $\ell(k)$ be an arbitrary sequence of $k \times 1$ vectors satisfying $0 < M_1 \le \|\ell(k)\|^2 \le M_2 < \infty$, and let $v_k^2 = \ell(k)' \Gamma_k^{-1} B_k \Gamma_k^{-1} \ell(k)$, where Γ_k is as defined previously and $B_k = E(Y_{t-1,k} Y_{t-1,k}' \varepsilon_t^2)$. Finally, let $\Rightarrow$ denote convergence in distribution.

Theorem 2.1. *Let* $\{y_t\}$ *satisfy* (2.1) *and assume that Assumptions* 1 *and* 2 *hold. Then* $\ell(k)' \sqrt{n-k} (\hat{\phi}(k) - \phi(k))/v_k \Rightarrow N(0,1)$.

Theorem 2.1 extends (the univariate version of) Theorem 4 of Lewis and Reinsel (1985) to the m.d.s. case. In particular, we relax the i.i.d. assumption on ε_t and allow for conditionally heteroskedastic errors. The conditions imposed on k are the same as those of Lewis and Reinsel (1985). Compared to Lewis and Reinsel (1985), our results require a strengthening of the moment and cumulant conditions. These conditions are sufficient, but may not be necessary (see Hannan and Deistler, 1988, p. 335).

One implication of Theorem 2.1 is that the limiting joint distribution of any fixed set of autoregressive estimators is multivariate normal with an asymptotic covariance matrix that reflects the possible presence of conditional heteroskedasticity. Results of this type are central to inference on many statistics of interest such as impulse responses, variance decompositions, measures of predictability, and tests of Granger noncausality (see the references in the introduction).

According to Theorem 2.1, under our assumptions the asymptotic variance of $\ell(k)' \sqrt{n-k} (\hat{\phi}(k) - \phi(k))$ is $v_k^2 \equiv \ell(k)' \Gamma_k^{-1} B_k \Gamma_k^{-1} \ell(k)$, as

opposed to $\sigma^2 \ell(k)' \Gamma_k^{-1} \ell(k)$ in the i.i.d. case (cf. Lewis and Reinsel, 1985, Theorem 4). Thus, the presence of conditional heteroskedasticity invalidates the usual LS inference for $AR(\infty)$ processes. To characterize further the asymptotic covariance matrix of the estimated autoregressive coefficients of the sieve approximation to (2.1) it is useful to define $\alpha_{l_1, l_2} = E(\varepsilon_{t-l_1} \varepsilon_{t-l_2} \varepsilon_t^2)$, for $l_1, l_2 = 1, 2, \ldots$. We note that α_{l_1, l_2} is closely related to the fourth-order joint cumulants of ε_t. More specifically, for $l_1, l_2 \geq 1$, we have that $\alpha_{l_1, l_2} = \kappa_\varepsilon(0, -l_1, -l_2, 0)$ when $l_1 \neq l_2$, and $\alpha_{l_1, l_2} = \kappa_\varepsilon(0, -l_1, -l_2, 0) + \sigma^4$ when $l_1 = l_2$. In the i.i.d. case, α_{l_1, l_2} are equal to 0 when $l_1 \neq l_2$, and they are equal to σ^4 when $l_1 = l_2$. As we will see next, B_k depends on the fourth order cumulants, or the closely related α_{l_1, l_2}, whose form is affected by conditional heteroskedasticity. Let $b_{j,k} = (\psi_{j-1}, \ldots, \psi_{j-k})'$, with $\psi_j = 0$ for $j < 0$, and note that $Y_{t-1,k} = \sum_{j=1}^{\infty} b_{j,k} \varepsilon_{t-j}$. This implies

$$B_k = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} b_{j,k} b'_{i,k} E(\varepsilon_{t-j} \varepsilon_{t-i} \varepsilon_t^2) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} b_{j,k} b'_{i,k} \alpha_{i,j},$$

given the definition of $\alpha_{i,j}$. Under conditional homoskedasticity (or the stronger i.i.d. assumption), $\alpha_{i,j} = \sigma^4 1 \ (i = j)$, where $1(\cdot)$ is the indicator function. Thus, in this case, $B_k = \sigma^4 \sum_{j=1}^{\infty} b_{j,k} b'_{j,k} = \sigma^2 \Gamma_k$, implying that v_k^2 simplifies to $\sigma^2 \ell(k)' \Gamma_k^{-1} \ell(k)$, the asymptotic variance of the estimated autoregressive coefficients in the i.i.d. case.

In practice, $v_k^2 \equiv \ell(k)' \Gamma_k^{-1} B_k \Gamma_k^{-1} \ell(k)$ is unknown and needs to be consistently estimated for the normal approximation result of Theorem 2.1 to be useful in applications. Under our assumptions, a consistent estimator of Γ_k is given by $\widehat{\Gamma}_k = (n - k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y'_{t-1,k}$ (see Lemma A.1 in the Appendix). In the possible presence of conditional heteroskedasticity of unknown form, consistent estimation of B_k requires the use of a heteroskedasticity-robust estimator. Here we use a version of the Eicker–White estimator, specifically, $\widehat{B}_k = (n - k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y'_{t-1,k} \hat{\varepsilon}_{t,k}^2$, where $\hat{\varepsilon}_{t,k} = y_t - Y'_{t-1,k} \hat{\phi}(k)$ is the LS residual of the autoregressive sieve. Our next result shows that $\hat{v}_k^2 = \ell(k)' \widehat{\Gamma}_k^{-1} \widehat{B}_k \widehat{\Gamma}_k^{-1} \ell(k)$ is a consistent estimator of v_k^2 under the same assumptions on ε_t as in Theorem 2.1, but with a slightly tighter upper bound on the rate of growth of k. In particular, we now require $k^4/n \to 0$ instead of the weaker condition $k^3/n \to 0$ needed for asymptotic normality.

Theorem 2.2. *Under the assumptions of Theorem* 2.1, *if in addition k satisfies $k \to \infty$ as $n \to \infty$ such that $k^4/n \to 0$, then $\hat{v}_k^2 - v_k^2 = o_P(1)$.*

Given Theorems 2.1 and 2.2, the *t*-statistic

$$t_k = \ell(k)'\sqrt{n-k}(\hat{\phi}(k) - \phi(k))/\hat{v}_k$$

has an asymptotic standard normal distribution. For inference on elements or linear combinations of $\phi(k)$, we may compute critical values from this asymptotic distribution. Alternatively, we can bootstrap the *t*-statistic and use bootstrap critical values instead. In Section 4 we will present evidence that the bootstrap approach tends to have better finite sample properties. The asymptotic properties of the bootstrap estimator are derived in the next section.

3. ASYMPTOTIC VALIDITY OF THE BOOTSTRAP

In this section we study the theoretical properties of bootstrap methods for AR(∞) processes subject to conditional heteroskedasticity of unknown form in the error term. In related work, bootstrap methods for inference on univariate infinite-order autoregressions with i.i.d. innovations have been studied by Kreiss (1997), Bühlmann (1997), and Choi and Hall (2000), among others. Extensions to the multivariate case are discussed in Paparoditis (1996) and in Inoue and Kilian (2002). The sieve bootstrap considered by these papers resamples randomly the residuals of an estimated truncated autoregression, the order of which is assumed to grow with the sample size at an appropriate rate. The bootstrap data are generated recursively from the fitted autoregressive model, given the resampled residuals and appropriate initial conditions. Given that the residuals are conditionally i.i.d. by construction, this sieve bootstrap method is not valid for AR(∞) models with conditional heteroskedasticity.

As our results below show, this problem may be solved by considering a fixed-design bootstrap method that applies the wild bootstrap (WB) to the regression residuals of the autoregressive sieve. Bootstrap observations on the dependent variable are generated by adding the WB residuals to the fitted values of the autoregressive sieve. These pseudo-observations are then regressed on the original regressor matrix. Thus, the fixed-design WB treats the regressors as fixed in repeated sampling, even though the regressors are lagged dependent variables. The fixed-design WB was originally suggested by Kreiss (1997), building on work by Wu (1986), Mammen (1993), and Liu (1988) who studied the WB in the cross-sectional context. A similar "fixed-regressor bootstrap" has been proposed by Hansen (2000) in the context of testing for structural change in regression models. Here we prove the asymptotic validity of the fixed-design WB for inference on AR(∞) processes with martingale difference errors that are possibly subject to conditional heteroskedasticity, which to the best of our knowledge has not been done elsewhere.

We also study the validity of an alternative bootstrap proposal that involves resampling pairs (or tuples) of the dependent and the explanatory variables. This pairwise bootstrap was originally suggested by Freedman (1981) in the cross-sectional context. Both bootstrap proposals have been studied in the context of finite-order autoregressions by Gonçalves and Kilian (2004).

In contrast to this earlier literature, here we establish the asymptotic validity of these two bootstrap proposals for sieve autoregressions under weak conditions on the form of conditional heteroskedasticity. We do not pursue more conventional recursive-design bootstrap methods, such as the recursive-design WB discussed in Kreiss (1997) and in Gonçalves and Kilian (2004), because such methods are more restrictive than the fixed-design WB and the pairwise bootstrap. Specifically, as shown by Gonçalves and Kilian (2004), the recursive-design method requires more stringent assumptions on the form of conditional heteroskedasticity than the two methods discussed in this paper. These restrictions run counter to the aim of imposing as little parametric structure as possible in bootstrap inference for linear stationary processes. In addition, the standard results of Paparoditis (1996) and Inoue and Kilian (2002) require exponential decay of the coefficients of the moving average representation of the underlying process. The results for the fixed-design bootstrap and the pairwise bootstrap, in contrast, only require a polynomial rate of decay.

The fixed-design WB consists of the following steps.

Step 1. Estimate an approximating $AR(k)$ model by LS and obtain LS residuals

$$\hat{\varepsilon}_{t,k} = y_t - Y'_{t-1,k}\hat{\phi}(k) \qquad \text{for } t = 1+k,\ldots,n,$$

where $\hat{\phi}(k) = (\hat{\phi}_{1,k},\ldots,\hat{\phi}_{k,k})'$ is the vector of LS estimators.

Step 2. Generate WB residuals according to

$$\hat{\varepsilon}^*_{t,k} = \hat{\varepsilon}_{t,k}\eta_t, \qquad \text{for } t = 1+k,\ldots,n,$$

with $\eta_t \sim \text{i.i.d.}(0,1)$ and $E^*|\eta_t|^4 \le \Delta < \infty$. One possible choice is $\eta_t \sim \text{i.i.d.}$ $N(0,1)$. Other choices have been discussed by Liu (1988) and Mammen (1993), among others.

Step 3. Given $\hat{\phi}(k)$ and $\hat{\varepsilon}^*_{t,k}$, generate bootstrap data for the dependent variable y^*_t according to

$$y^*_t = Y'_{t-1,k}\hat{\phi}(k) + \hat{\varepsilon}^*_{t,k}, \qquad \text{for } t = 1+k,\ldots,n.$$

Step 4. Compute $\hat{\phi}^*_{fwb}(k) = (\hat{\phi}^*_{fwb,1,k},\ldots,\hat{\phi}^*_{fwb,k,k})'$ by regressing y^*_t on $Y_{t-1,k}$.

According to the previous algorithm,

$$\hat{\phi}_{fwb}^*(k) = \widehat{\Gamma}_{fwb,k}^{*-1} \widehat{\Gamma}_{fwb,k,1}^*,$$

where $\widehat{\Gamma}_{fwb,k}^* = \widehat{\Gamma}_k$ and $\widehat{\Gamma}_{fwb,k,1}^* = (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} y_t^*$, with y_t^* as described in Step 3. Note that for the fixed-design WB $Y_{t-1,k}^* = Y_{t-1,k}$.

The pairwise bootstrap consists of the following steps.

Step 1. For given k, let $\mathscr{Z} = \{(y_t, Y_{t-1,k}') : t = 1+k, \ldots, n\}$ be the set of all "pairs" (or tuples) of data.

Step 2. Generate a bootstrap sample $\mathscr{Z}^* = \{(y_t^*, Y_{t-1,k}^{*\prime}) : t = 1+k, \ldots, n\}$ by resampling with replacement the "pairs" of data from $\mathscr{Z}$.

Step 3. Compute $\hat{\phi}_{pb}^*(k) = (\hat{\phi}_{pb,1,k}^*, \ldots, \hat{\phi}_{pb,k,k}^*)'$ by regressing y_t^* on $Y_{t-1,k}^*$.

Accordingly, let

$$\hat{\phi}_{pb}^*(k) = \widehat{\Gamma}_{pb,k}^{*-1} \widehat{\Gamma}_{pb,k,1}^*,$$

where $\widehat{\Gamma}_{pb,k}^* = (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}^* Y_{t-1,k}^{*\prime}$ and $\widehat{\Gamma}_{pb,k,1}^* = (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}^* y_t^*$.

It is useful to differentiate the pairwise bootstrap from the blocks-of-blocks (BOB) bootstrap, as discussed in Gonçalves and White (2004). Let r denote the first-stage block size and s the second-stage block size of the BOB bootstrap. The pairwise bootstrap for AR(∞) processes emerges as a special case of the BOB bootstrap with $r = p + 1$ and $s = 1$. Note that under our assumptions choosing $s > 1$ given $r = p + 1$ would be inefficient.

We focus on bootstrapping the studentized statistic t_k. The fixed-design WB analog of t_k is given by

$$t_{fwb,k}^* = \ell(k)' \sqrt{n-k} (\hat{\phi}_{fwb}^*(k) - \hat{\phi}(k)) / \hat{v}_{fwb,k}^*,$$

where $\hat{v}_{fwb,k}^{*2}$ is a bootstrap variance estimator consistent for v_k^2 (see Lemma A.7). In particular, $\hat{v}_{fwb,k}^{*2} = \ell(k)' \widehat{\Gamma}_{fwb,k}^{*-1} \widehat{B}_{fwb,k}^* \widehat{\Gamma}_{fwb,k}^{*-1} \ell(k)$, where $\widehat{\Gamma}_{fwb,k}^* = \widehat{\Gamma}_k$ and

$$\widehat{B}_{fwb,k}^* = (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}^* Y_{t-1,k}^{*\prime} \tilde{\varepsilon}_{t,k}^{*2} = \sum_{t=1+k}^{n} Y_{t-1,k} Y_{t-1,k}' \tilde{\varepsilon}_{t,k}^{*2},$$

where $\tilde{\varepsilon}_{t,k}^* = y_t^* - Y_{t-1,k}' \hat{\phi}^*(k)$ is the fixed-design wild bootstrap residual.

The pairwise bootstrap analog of t_k is given by

$$t_{pb,k}^* = \ell(k)' \sqrt{n-k} (\hat{\phi}_{pb}^*(k) - \hat{\phi}(k)) / \hat{v}_{pb,k}^*,$$

with $\hat{v}^{*2}_{pb,k} = \ell(k)' \widehat{\Gamma}^{*-1}_{pb,k} \widehat{B}^*_{pb,k} \widehat{\Gamma}^{*-1}_{pb,k} \ell(k)$, where

$$\widehat{B}^*_{pb,k} = (n-k)^{-1} \sum_{t=1+k}^{n} Y^*_{t-1,k} Y^{*\prime}_{t-1,k} \tilde{\varepsilon}^{*2}_{t,k}$$

and $\tilde{\varepsilon}^*_{t,k} = y^*_t - Y^{*\prime}_{t-1,k} \hat{\phi}^*(k)$ is the pairwise bootstrap LS residual. Lemma A.7 shows that $\hat{v}^{*2}_{pb,k}$ is consistent for v^2_k.

Next, we show that the conditional distributions of $t^*_{fwb,k}$ and $t^*_{pb,k}$ can be used to approximate the true but unknown finite-sample distribution of t_k. Our main result is as follows.

Theorem 3.1. *Under the assumptions of Theorem* 2.1, *if in addition k satisfies* $k \to \infty$ *as* $n \to \infty$ *such that* $k^4/n \to 0$, *then for the fixed-design WB and for the pairwise bootstrap and for any* $\delta > 0$ *it follows that*

$$P\left\{ \sup_{x \in \mathbf{R}} |P^*[t^*_k \leq x] - P[t_k \leq x]| > \delta \right\} \to 0,$$

where t^*_k denotes either $t^*_{fwb,k}$ or $t^*_{pb,k}$, and P^* is the probability measure induced by the corresponding bootstrap scheme.

The assumptions underlying the bootstrap approximation in Theorem 3.1 are the same as those needed to apply the asymptotic normal approximation based on the consistent variance estimator of Theorem 2.2. Note that for both the fixed-design WB and the pairwise bootstrap, the asymptotic bootstrap population variance is $v^{*2}_k \equiv \ell(k)' \Gamma^{-1}_k \widehat{B}_k \Gamma^{-1}_k \ell(k)$ (cf. Lemma 6). Thus, v^{*2}_k depends on the same heteroskedasticity-robust covariance matrix estimator of B_k as the estimator $\hat{v}^2_k$ in Theorem 2.2. In both cases, the same upper bound on the rate of increase of k is needed to ensure consistency for v^2_k.

Although Theorem 3.1 focuses on the bootstrap t statistics, the bootstrap is also asymptotically valid for the estimated autoregressive parameters. Indeed, to prove the asymptotic validity of the bootstrap t statistics we first prove the asymptotic validity of the bootstrap for the autoregressive estimated parameters (cf. the proof of Theorem 3.1). Theorem 3.1 follows from this result and the consistency of each bootstrap variance estimator $\hat{v}^{*2}_{fwb,k}$ and $\hat{v}^{*2}_{pb,k}$ for v^2_k (see Lemma A.7).

4. SIMULATION EVIDENCE

In this section we study the finite-sample accuracy of the asymptotic and bootstrap approximations introduced in the previous section.

TABLE 1 Coverage rates of nominal 90% symmetric percentile-t intervals for ϕ_i, $i = 1, 2, 3, 4$ in the approximating AR(k) model: ARMA(1, 1)-NID DGP

DGP:			$y_t = 0.9y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \varepsilon_t \sim NID(0,1)$							
			$n = 120$				$n = 240$			
θ	k	Method	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_1	ϕ_2	ϕ_3	ϕ_4
0.3	6	Robust Gaussian	86.42	88.45	88.82	88.85	88.73	89.47	89.29	89.01
		Fixed WB	87.57	89.26	89.95	89.88	89.10	89.99	89.81	89.72
		Pairwise bootstrap	88.50	90.42	90.71	90.82	89.62	90.45	90.21	89.95
	8	Robust Gaussian	85.89	88.35	88.75	88.72	88.52	89.37	89.20	89.26
		Fixed WB	87.10	89.36	89.86	89.75	88.95	89.86	89.76	89.73
		Pairwise bootstrap	88.46	90.76	90.92	90.89	89.60	90.40	90.27	90.33
	10	Robust Gaussian	85.56	87.96	88.27	88.27	88.42	89.05	89.12	89.26
		Fixed WB	86.89	89.19	89.49	89.44	89.00	89.67	89.71	89.70
		Pairwise bootstrap	88.72	90.76	91.01	91.12	89.55	90.20	90.32	90.65
	12	Robust Gaussian	85.05	87.63	87.91	88.17	88.30	89.02	89.20	89.14
		Fixed WB	86.43	89.09	89.22	89.17	88.87	89.43	89.62	89.69
		Pairwise bootstrap	88.72	90.72	91.24	91.11	89.78	90.35	90.39	90.73
0.6	6	Robust Gaussian	86.13	87.17	89.14	90.33	88.49	89.06	89.65	89.94
		Fixed WB	87.15	88.79	90.29	91.05	89.94	89.65	90.20	90.31
		Pairwise bootstrap	88.09	89.69	91.00	91.89	89.48	90.04	90.53	90.77
	8	Robust Gaussian	85.80	87.94	88.48	89.04	88.42	89.12	89.32	89.61
		Fixed WB	86.85	88.97	89.56	90.24	88.90	89.63	89.84	89.98
		Pairwise bootstrap	88.08	90.28	90.60	91.16	89.55	90.36	90.42	90.76
	10	Robust Gaussian	85.25	87.56	88.19	88.65	88.34	89.02	89.10	89.15
		Fixed WB	86.60	88.92	89.21	89.70	88.81	89.50	89.93	89.58
		Pairwise bootstrap	88.39	90.48	90.68	91.25	89.41	90.13	90.52	90.37
	12	Robust Gaussian	84.92	87.05	87.95	88.27	88.12	88.83	89.21	88.89
		Fixed WB	86.18	88.57	88.79	89.28	88.73	89.37	89.78	89.74
		Pairwise bootstrap	88.41	90.55	91.31	90.99	89.65	90.25	90.66	90.45
−0.3	6	Robust Gaussian	86.80	88.24	88.52	88.53	88.84	89.42	89.28	89.05
		Fixed WB	87.96	89.32	89.58	89.57	89.31	89.92	89.91	89.58
		Pairwise bootstrap	88.91	90.22	90.46	90.50	89.85	90.27	90.08	89.94
	8	Robust Gaussian	86.30	88.07	88.33	87.89	88.68	89.11	89.29	88.73
		Fixed WB	87.70	89.06	89.35	89.07	89.21	89.61	89.76	89.33
		Pairwise bootstrap	89.07	90.10	90.56	90.06	89.73	90.20	90.43	89.73
	10	Robust Gaussian	86.13	87.72	87.91	88.04	88.70	89.08	89.09	88.67
		Fixed WB	87.42	88.99	89.18	88.98	89.23	89.69	89.55	89.30
		Pairwise bootstrap	89.14	90.42	90.83	90.61	89.86	90.26	90.28	89.99
	12	Robust Gaussian	85.43	87.15	87.59	87.55	88.48	89.08	88.90	88.69
		Fixed WB	86.79	88.47	88.90	88.70	88.96	89.74	89.52	89.19
		Pairwise bootstrap	88.99	90.18	90.72	90.62	89.84	90.43	90.26	90.00

Source: Based on 20,000 Monte Carlo draws with 1,000 bootstrap replications each.

The data generating process for $\{y_t\}$ is a stationary ARMA(1,1) model of the form

$$y_t = \rho y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \qquad \text{for } t = 1 + k, \ldots, n$$

TABLE 2 Coverage rates of nominal 90% symmetric percentile-t intervals for $\phi_i, i = 1, 2, 3, 4$ in the approximating AR(k) model: ARMA(1, 1)-N-ARCH DGP

DGP:			$y_t = 0.9y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \varepsilon_t = h_t^{1/2}v_t, h_t = \omega + 0.5\varepsilon_{t-1}^2, v_t \sim N(0, 1)$							
			$n = 120$				$n = 240$			
θ	k	Method	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_1	ϕ_2	ϕ_3	ϕ_4
0.3	6	Robust Gaussian	83.96	85.80	87.72	88.46	86.92	87.15	88.28	88.53
		Fixed WB	87.83	89.04	89.62	90.03	89.11	89.81	89.96	89.98
		Pairwise bootstrap	89.04	90.29	90.64	91.12	89.76	90.38	90.37	90.50
	8	Robust Gaussian	83.17	85.72	87.42	87.96	86.81	87.21	88.16	88.45
		Fixed WB	87.17	89.17	89.76	89.83	89.13	89.64	89.67	89.57
		Pairwise bootstrap	88.68	90.45	91.00	91.33	89.97	90.44	90.39	90.56
	10	Robust Gaussian	82.64	85.43	87.47	87.92	86.50	87.00	87.94	88.38
		Fixed WB	86.79	88.86	89.48	89.54	88.98	89.43	89.60	89.57
		Pairwise bootstrap	88.94	90.49	91.21	91.46	89.86	90.38	90.22	90.54
	12	Robust Gaussian	82.10	84.96	87.11	87.61	86.18	86.68	87.90	88.34
		Fixed WB	86.31	88.53	89.42	89.24	88.77	89.23	89.54	89.66
		Pairwise bootstrap	88.90	90.63	91.55	91.70	89.87	90.46	90.52	90.73
0.6	6	Robust Gaussian	83.75	85.96	88.04	89.45	86.91	87.14	88.40	89.80
		Fixed WB	87.38	88.88	90.20	91.23	89.20	89.61	90.14	91.13
		Pairwise bootstrap	88.65	90.04	91.09	92.28	89.73	90.18	90.79	91.60
	8	Robust Gaussian	83.01	85.38	87.16	88.40	86.61	87.19	87.84	88.65
		Fixed WB	87.02	88.88	89.71	90.14	88.97	89.51	89.69	90.10
		Pairwise bootstrap	88.63	90.38	91.09	91.48	89.76	90.37	90.45	90.62
	10	Robust Gaussian	82.53	84.88	86.80	88.01	86.49	87.02	87.55	88.35
		Fixed WB	86.59	88.72	89.31	89.57	88.93	89.52	89.48	89.75
		Pairwise bootstrap	88.94	90.55	91.26	91.60	89.86	90.36	90.31	90.58
	12	Robust Gaussian	81.92	84.43	86.33	87.65	86.14	86.79	87.51	88.00
		Fixed WB	86.21	88.14	89.03	89.67	88.75	89.33	89.44	89.50
		Pairwise bootstrap	88.85	90.52	91.33	91.60	89.73	90.21	90.45	90.46
−0.3	6	Robust Gaussian	84.28	86.21	87.63	88.13	87.01	87.47	88.69	88.81
		Fixed WB	88.15	89.06	89.52	89.74	89.37	89.53	90.10	89.98
		Pairwise bootstrap	89.18	90.09	90.80	90.76	89.97	90.27	90.93	90.62
	8	Robust Gaussian	83.42	85.83	87.17	87.53	86.82	87.24	88.22	88.68
		Fixed WB	87.31	88.80	89.53	89.29	89.20	89.33	89.71	90.01
		Pairwise bootstrap	88.82	90.30	91.09	90.84	90.01	90.39	90.61	90.64
	10	Robust Gaussian	83.16	85.42	87.04	87.58	86.52	87.02	88.00	88.66
		Fixed WB	87.10	88.65	89.53	89.04	89.10	89.26	89.61	89.73
		Pairwise bootstrap	89.11	90.43	91.30	91.10	89.96	90.30	90.60	90.80
	12	Robust Gaussian	82.35	85.05	86.58	87.28	86.26	87.10	87.93	88.66
		Fixed WB	86.83	88.24	88.87	88.74	88.87	89.29	89.61	89.77
		Pairwise bootstrap	89.17	90.51	91.20	91.06	89.86	90.47	90.74	90.70

Source: See Table 1.

with $\rho = 0.9$ and $\theta \in \{0.3, 0.6, -0.3, -0.6\}$. These parameter settings capture typical situations in applied work. We model the error term ε_t alternatively as an ARCH model or a GARCH model. For a motivation of the specific models of conditional heteroskedasticity used see Gonçalves and Kilian (2004). The sample size is $n \in \{120, 240\}$, corresponding to 10 and 20 years of monthly data, respectively. We compare the coverage

S. Gonçalves and L. Kilian

TABLE 3 Coverage rates of nominal 90% symmetric percentile-t intervals for $\phi_i, i = 1, 2, 3, 4$ in the approximating $AR(k)$ model: ARMA$(1, 1)$-N-GARCH DGP

DGP:			$y_t = 0.9y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \varepsilon_t = h_t^{1/2}v_t, h_t = \omega + 0.05\varepsilon_{t-1}^2 + 0.94h_{t-1}, v_t \sim N(0,1)$							
			$n = 120$				$n = 240$			
θ	k	Method	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_1	ϕ_2	ϕ_3	ϕ_4
0.3	6	Robust Gaussian	86.29	88.06	88.42	88.62	88.08	89.03	89.32	88.95
		Fixed WB	87.72	89.50	89.82	89.83	89.08	89.89	90.01	89.94
		Pairwise bootstrap	88.57	90.49	90.50	90.70	89.44	90.33	90.34	90.11
	8	Robust Gaussian	85.78	87.99	88.37	88.47	87.86	89.08	88.92	89.21
		Fixed WB	87.30	89.42	89.61	89.79	88.89	89.90	89.64	89.82
		Pairwise bootstrap	88.60	90.83	90.77	90.87	89.40	90.43	90.31	90.26
	10	Robust Gaussian	85.58	87.60	87.97	88.29	87.70	88.90	88.73	89.11
		Fixed WB	87.05	89.18	89.50	89.69	88.60	89.80	89.48	89.97
		Pairwise bootstrap	88.92	90.61	90.99	91.10	89.25	90.48	90.21	90.57
	12	Robust Gaussian	84.84	86.93	87.72	88.04	87.80	88.89	88.59	89.09
		Fixed WB	86.53	88.75	89.22	89.27	88.67	89.55	89.61	89.99
		Pairwise bootstrap	88.76	90.84	91.09	91.17	89.37	90.44	90.48	90.89
0.6	6	Robust Gaussian	85.95	83.37	88.77	89.91	87.99	88.85	89.68	89.89
		Fixed WB	87.27	88.96	89.88	90.93	88.85	89.71	90.34	90.69
		Pairwise bootstrap	88.20	89.81	90.79	91.60	89.08	90.01	90.66	91.13
	8	Robust Gaussian	85.52	87.70	87.85	88.68	87.89	88.89	88.87	89.12
		Fixed WB	86.99	89.29	89.41	89.75	88.70	89.80	89.63	90.02
		Pairwise bootstrap	88.33	90.44	90.42	90.98	89.32	90.24	90.31	90.46
	10	Robust Gaussian	85.08	87.09	87.74	88.27	87.69	88.68	88.71	88.85
		Fixed WB	86.71	88.91	89.15	89.54	88.59	89.45	89.54	89.57
		Pairwise bootstrap	88.60	90.46	90.97	91.06	89.26	89.94	90.18	90.45
	12	Robust Gaussian	84.62	86.75	87.34	88.06	87.67	88.16	88.77	88.55
		Fixed WB	86.32	88.30	88.92	89.32	88.64	88.92	89.57	89.61
		Pairwise bootstrap	88.54	90.25	91.00	91.16	89.25	90.01	90.39	90.49
−0.3	6	Robust Gaussian	86.65	88.07	88.33	88.68	88.40	89.15	88.89	88.87
		Fixed WB	88.04	89.25	89.64	89.75	89.17	89.93	89.61	89.78
		Pairwise bootstrap	88.74	90.20	90.66	90.60	89.71	90.32	90.15	90.19
	8	Robust Gaussian	86.30	87.77	87.98	88.02	88.26	88.99	88.74	88.75
		Fixed WB	87.87	89.15	89.54	89.50	89.05	89.78	89.47	89.50
		Pairwise bootstrap	89.18	90.31	90.65	90.56	89.53	90.23	90.31	90.24
	10	Robust Gaussian	86.09	87.48	87.61	87.79	88.04	88.78	88.38	88.82
		Fixed WB	87.54	88.82	88.98	89.40	88.97	89.56	89.28	89.69
		Pairwise bootstrap	89.20	90.54	90.73	91.05	89.52	90.46	90.21	90.48
	12	Robust Gaussian	85.25	87.01	87.56	87.72	87.89	88.44	88.38	88.50
		Fixed WB	86.87	88.59	89.05	89.00	88.86	89.54	89.35	89.44
		Pairwise bootstrap	89.18	90.54	90.91	90.88	89.61	90.21	90.28	90.31

Source: See Table 1.

accuracy of alternative nominal 90% confidence intervals for the first four slope parameters of the approximating autoregressive model

$$y_t = \sum_{i=1}^{k} \phi_i y_{t-i} + \varepsilon_{t,k}, \qquad \text{for } t = 1 + k, \dots, n,$$

where $k \in \{6, 8, 10, 12\}$. The intervals considered include the asymptotic normal interval based on robust standard errors as well as the corresponding symmetric bootstrap percentile-t intervals based on the fixed-design bootstrap algorithm and based on the pairwise bootstrap algorithm. The simulation results are fairly robust to changes in the lag order, provided the lag order of the approximating model is not too short. The fixed-design WB method is 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 $1 - p$, suggested by Mammen (1993).

Table 1 shows results for an ARMA(1,1) data generating process with NID(0,1) innovations. Tables 2 and 3 show the corresponding results for ARMA(1,1) models with ARCH(1) and GARCH(1,1) errors. To conserve space we only show some of the results. The other simulation results are very similar. The bootstrap intervals perform consistently well under all designs. The asymptotic interval rarely is more accurate than the two bootstrap intervals. For $n = 120$, the pairwise bootstrap interval can be more accurate than the asymptotic interval by up to 7 percentage points; for $n = 240$ the gain still may approach 4 percentage points. There is no clear ranking of the two bootstrap methods, but on average the pairwise bootstrap interval is more accurate for $n = 120$.

5. CONCLUSION

Our main contribution in this paper has been the derivation of the asymptotic distribution of the bootstrap least-squares estimator in the sieve model with conditional heteroskedasticity of unknown form. In addition, we developed robust asymptotic methods of inference for this model. Our results are immediately applicable to problems of testing lag-order restrictions for autoregressions and of computing dynamic multipliers or measures of persistence. They may also be generalized to allow the construction of variance decompositions and of orthogonalized impulse responses as well as tests of Granger noncausality and tests of structural stability.

Our analysis in this paper has assumed that the unconditional error variance is constant. A more general class of volatility models allows for time-varying unconditional variances. The latter class of models could be handled by a subsampling approach, as discussed in the context of the finite-lag order autoregressive model by Politis et al. (1999, Chapter 12.2). This does not mean that subsampling is generally preferred, however. If the unconditional variance is constant, as we have assumed, consistent with the leading examples of volatility models used in applied work, then subsampling will be inefficient relative to bootstrapping.

An interesting extension of this paper would be a study of the relative accuracy of the first order asymptotic approximation and of the

bootstrap approximation, when closed form solutions for the variance are available. It also would be useful to clarify the conditions under which the bootstrap approximation will provide asymptotic refinements. We defer these important questions to future research.

A. APPENDIX

Throughout this Appendix, C denotes a generic constant independent of n. Given a matrix A, let $\|A\|$ denote the matrix norm defined by $\|A\|^2 = tr(A'A)$, and let $\|A\|_1 = \sup_{x \neq 0}\{\|Ax\|/\|x\|\}$. The following inequalities relating $\|\cdot\|$ and $\|\cdot\|_1$ are often used below: $\|A\|_1^2 \leq \|A\|^2$, $\|AB\|^2 \leq \|A\|_1^2\|B\|^2$, and $\|AB\|^2 \leq \|B\|_1^2\|A\|^2$.

For any bootstrap statistic T_n^* we write $T_n^* = o_{P*}(1)$ in probability when for any $\delta > 0$, $P^*(|T_n^*| > \delta) = o_P(1)$. We write $T_n^* = O_{P*}(n^\delta)$ in probability when for all $\delta > 0$ there exists a $M_\delta < \infty$ such that $\lim_{n \to \infty} P[P^*(|n^{-\lambda}T_n^*| > M_\delta) > \delta] = 0$. 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. For a more detailed exposition of the in-probability bootstrap asymptotics used in this paper see Giné and Zinn (1990). $E^*(\cdot)$ and $Var^*(\cdot)$ denote the expectation and variance with respect to the bootstrap data conditional on the original data.

To conserve space, we state the bootstrap results jointly for the fixed-design WB and for the pairwise bootstrap, dropping the subscript *fwb* or *pb*, whenever this distinction is not needed. For instance, $\hat{\phi}^*(k)$ denotes either $\hat{\phi}_{fwb}^*(k)$ or $\hat{\phi}_{pb}^*(k)$ and we use $(y_t^*, Y_{t-1,k}^*)$ to denote bootstrap data in general. Note that for the fixed-design WB $Y_{t-1,k}^* = Y_{t-1,k}$. Similarly, we let v_k^{*2} denote either $v_{fwb,k}^{*2}$ or $v_{pb,k}^{*2}$.

A.1. Auxiliary Lemmas

Some of the lemmas used to prove our main results require assumptions that are weaker than Assumptions 1 and 2 in the main text. Since these lemmas are of independent interest, we state the weaker version of these assumptions below.

Assumption 1.(iii$'$) $\sum_{l_1=-\infty}^{\infty} \sum_{l_2=-\infty}^{\infty} \sum_{l_3=-\infty}^{\infty} |\kappa_\varepsilon(0, l_1, l_2, l_3)| < \infty$.
Assumption 2.(i$'$) $\frac{k^2}{n} \to 0$ as $k, n \to \infty$

Lemma A.1 is an extension of Berk's (1974) Lemma 3 for AR(∞) processes with i.i.d. errors to the case of AR(∞) processes with m.d.s. errors satisfying Assumption 1(i), (ii), and (iii$'$).

Lemma A.1. *Let* $\{y_t\}$ *be generated from* (2.1) *and assume Assumptions* $1(i)$, (ii), *and* (iii') *holds. Then, if* $k, n \to \infty$ *such that* $k^2/n \to 0$,

$$\left\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 = o_P(1). \tag{A.1}$$

If instead $k^3/n \to 0$,

$$k^{1/2}\left\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 = o_P(1). \tag{A.2}$$

The next result states the consistency of $\hat{\phi}(k)$ for $\phi(k)$ provided $k/n \to 0$. It is an univariate extension of Theorem 2.1 of Paparoditis (1996) to the conditionally heteroskedastic AR(∞) case. It is used in proving our main results further below.

Lemma A.2. *Let* $\{y_t\}$ *satisfy* (2.1) *and assume that Assumptions* $1(i)$, (ii), *and* (iii') *and* $2(i')$ *and* (ii) *hold. Then* $\|\hat{\phi}(k) - \phi(k)\| = O_P\left(\frac{k^{1/2}}{n^{1/2}}\right)$.

The next lemma is useful for deriving the asymptotic distribution of the estimated autoregressive parameters. For the univariate case it is the m.d.s. extension of Lewis and Reinsel's (1985), Theorem 2.

Lemma A.3. *Let* $\{y_t\}$ *satisfy* (2.1) *and assume that Assumptions* $1(i)$, (ii), *and* (iii') *and Assumption* 2 *hold. Then*

$$\ell(k)'\sqrt{n-k}(\hat{\phi}(k) - \phi(k)) - \ell(k)'\sqrt{n-k}\Gamma_k^{-1}\left((n-k)^{-1}\sum_{t=1+k}^{n} Y_{t-1,k}\varepsilon_t\right) = o_P(1).$$

The following result is the pairwise bootstrap analogue of Lemma A.1. Note that for the fixed-design WB, $\widehat{\Gamma}_{fwb,k}^* = \widehat{\Gamma}_k$ and therefore the fixed-design version of this result is not needed.

Lemma A.4. *Under the conditions of Lemma* A.1, *if* $k, n \to \infty$ *such that* $k^2/n \to 0$, $\|\widehat{\Gamma}_{pb,k}^{*-1} - \widehat{\Gamma}_k^{-1}\|_1 = o_{P*}(1)$, *in probability. If instead* $k^3/n \to 0$, *then* $k^{1/2}\|\widehat{\Gamma}_{pb,k}^{*-1} - \widehat{\Gamma}_k^{-1}\|_1 = o_{P*}(1)$, *in probability.*

Lemmas A.5–A.7 below are the bootstrap (fixed-design WB and pairwise bootstrap) analogs of Lemmas A.2, A.3, and Theorem 2.2, respectively.

Lemma A.5. *Under the assumptions of Lemma* A.2, *for the fixed-design WB and for the pairwise bootstrap, it follows that* $\|\hat{\phi}^*(k) - \hat{\phi}(k)\| = O_{P*}\left(\frac{k^{1/2}}{n^{1/2}}\right)$, *in probability.*

Lemma A.6. *Under the assumptions of Lemma* A.3, *for the fixed-design WB and for the pairwise bootstrap, it follows that*

$$\ell(k)'\sqrt{n-k}(\hat{\phi}^*(k) - \hat{\phi}(k))/v_k = (n-k)^{-1/2} \sum_{t=1+k}^{n} \ell(k)'\Gamma_k^{-1}v_k^{-1}Y_{t-1,k}^*\hat{\varepsilon}_{t,k}^* + r^*,$$

where $r^* = o_{P^*}(1)$ *in probability.*

Lemma A.7. *Under the assumptions of Theorem* 2.2, *if* $k^4/n \to 0$, $\hat{v}_k^{*2} - v_k^2 = o_{P^*}(1)$, *in probability.*

Proof of Lemma A.1. We follow the proof of Berk's (1974) Lemma 3. Since $\|\Gamma_k^{-1}\|_1$ is bounded, we can write

$$\left\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 \le \frac{C^2\|\widehat{\Gamma}_k - \Gamma_k\|_1}{1 - C\|\widehat{\Gamma}_k - \Gamma_k\|_1}. \tag{A.3}$$

Because $E\|\widehat{\Gamma}_k - \Gamma_k\|_1^2 \le E\|\widehat{\Gamma}_k - \Gamma_k\|^2$, it suffices to show that $E\|\widehat{\Gamma}_k - \Gamma_k\|^2 = O\left(\frac{k^2}{n-k}\right) \to 0$, if $k^2/n \to 0$, which implies that $\|\widehat{\Gamma}_k - \Gamma_k\|_1 = o_P(1)$, and thus from (A.3), $\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1 = o_P(1)$. If instead $k^3/n \to 0$, then $E((k^{1/2}\|\widehat{\Gamma}_k - \Gamma_k\|_1)^2) \le kE(\|\widehat{\Gamma}_k - \Gamma_k\|^2) \le Ck^3/(n-k) \to 0$, showing that $k^{1/2}\|\widehat{\Gamma}_k - \Gamma_k\|_1 = o_P(1)$ and consequently $k^{1/2}\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1 = o_P(1)$. To show that $E\|\widehat{\Gamma}_k - \Gamma_k\|^2 = O(\frac{k^2}{n-k})$, we use routine calculations (see, e.g., Hannan, 1970, p. 209) to write $E((n-k)^{-1}\sum_{t=1+k}^{n}(y_{t-i}y_{t-j} - E(y_{t-i}y_{t-j})))^2$ as a function of fourth-order cumulants of $\{y_t\}$ and then use Assumption 1(iii′) and Theorem 2.8.1 of Brillinger (1981) to bound this expression.   □

Proof of Lemma A.2. We follow Lewis and Reinsel's (1985) proof of their Theorem 1. Let $\varepsilon_{t,k} = y_t - Y_{t-1,k}'\phi(k)$. We have that

$$\left\|\hat{\phi}(k) - \phi(k)\right\| \le \left\|\widehat{\Gamma}_k^{-1}\right\|_1 \|U_{1n}\| + \left\|\widehat{\Gamma}_k^{-1}\right\|_1 \|U_{2n}\|, \tag{A.4}$$

with $U_{1n} = (n-k)^{-1}\sum_{t=k+1}^{n} Y_{t-1,k}(\varepsilon_t - \varepsilon_{t,k})$, and $U_{2n} = (n-k)^{-1}\sum_{t=k+1}^{n} Y_{t-1,k}\varepsilon_t$. We can show that (a) $\|\widehat{\Gamma}_k^{-1}\|_1 = O_P(1)$, (b) $\|U_{1n}\| = o_P(\frac{k^{1/2}}{n^{1/2}})$, and (c) $\|U_{2n}\| = O_P(\frac{k^{1/2}}{n^{1/2}})$. (a) and (b) follow as in Lewis and Reinsel (1985, cf. Eq. (2.9)), given (A.1). For (c), we can write

$$E(\|U_{2n}\|^2) = (n-k)^{-2} \sum_{j=1}^{k} \sum_{t=1+k}^{n} \sum_{s=1+k}^{n} E(y_{t-j}y_{s-j}\varepsilon_t\varepsilon_s)$$

$$= (n-k)^{-2} \sum_{j=1}^{k} \sum_{t=1+k}^{n} E(y_{t-j}^2\varepsilon_t^2), \tag{A.5}$$

since $E(y_{t-j}y_{s-j}\varepsilon_t\varepsilon_s) = 0$ for $t \neq s$. It follows that

$$E(y_{t-j}^2\varepsilon_t^2) = E\left(\left(\sum_{l=0}^{\infty}\psi_l\varepsilon_{t-j-l}\right)^2\varepsilon_t^2\right) = \sum_{l_1=0}^{\infty}\sum_{l_2=0}^{\infty}\psi_{l_1}\psi_{l_2}E(\varepsilon_{t-j-l_1}\varepsilon_{t-j-l_2}\varepsilon_t^2)$$

$$= \sum_{l_1=0}^{\infty}\sum_{l_2=0}^{\infty}\psi_{l_1}\psi_{l_2}\alpha_{l_1+j,l_2+j} \leq C\left(\sum_{l=0}^{\infty}\psi_l\right)^2 < \infty,$$

given that α_{l_1+j,l_2+j} are uniformly bounded under Assumption 1(iii) and that $\sum_{l=0}^{\infty}|\psi_l| < \infty$. Thus, $E(\|U_{2n}\|^2) \leq C\frac{k}{n-k}$, implying $\|U_{2n}\| = O_P\left(\frac{k^{1/2}}{n^{1/2}}\right)$ by the Markov inequality. □

Proof of Lemma A.3. The proof follows exactly the proof of Theorem 2 of Lewis and Reinsel (1985, p. 399), given in particular our Lemma A.1, Eq. (A.2), and the proof of our Lemma A.2. □

Proof of Lemma A.4. Following the same argument as in the proof of Lemma A.1, for the first result it is enough to show that $\|\widehat{\Gamma}_{pb,k}^* - \widehat{\Gamma}_k\|_1 = o_{P^*}(1)$ in probability, or by the Markov inequality, that $E^*(\|\widehat{\Gamma}_{pb,k}^* - \widehat{\Gamma}_k\|^2) = o_P(1)$. By definition of the Euclidean matrix norm,

$$E^*\left(\left\|\widehat{\Gamma}_{pb,k}^* - \widehat{\Gamma}_k\right\|^2\right)$$

$$= tr\left((n-k)^{-2}\sum_{t=1+k}^{n}\sum_{s=1+k}^{n}E^*\left[\left(Y_{t-1,k}^*Y_{t-1,k}^{*\prime} - \widehat{\Gamma}_k\right)\left(Y_{s-1,k}^*Y_{s-1,k}^{*\prime} - \widehat{\Gamma}_k\right)\right]\right)$$

$$= tr\left((n-k)^{-2}\sum_{t=1+k}^{n}\left(Y_{t-1,k}Y_{t-1,k}' - \widehat{\Gamma}_k\right)\left(Y_{t-1,k}Y_{t-1,k}' - \widehat{\Gamma}_k\right)\right)$$

$$= (n-k)^{-1}tr\left((n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}Y_{t-1,k}'Y_{t-1,k}Y_{t-1,k}' - \widehat{\Gamma}_k\cdot\widehat{\Gamma}_k\right),$$

where the second equality uses the fact that $E^*[(Y_{t-1,k}^*Y_{t-1,k}^{*\prime} - \widehat{\Gamma}_k)(Y_{s-1,k}^*Y_{s-1,k}^{*\prime} - \widehat{\Gamma}_k)] = 0$ when $t \neq s$. Since $\|(n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}Y_{t-1,k}'Y_{t-1,k}Y_{t-1,k}'\| \leq (n-k)^{-1}\sum_{t=1+k}^{n}\|Y_{t-1,k}\|^4 = O_P(k^2)$ (given that $\sup_t E|y_t|^4 \leq C < \infty$) and $\widehat{\Gamma}_k = O_P(1)$, it follows that $E^*(\|\widehat{\Gamma}_{pb,k}^* - \widehat{\Gamma}_k\|^2) = O_P\left(\frac{k^2}{n-k}\right) + O_P\left(\frac{1}{n-k}\right) = o_P(1)$ given that $\frac{k^2}{n} \to 0$. The second result follows similarly given that $\frac{k^3}{n} \to 0$. □

Proof of Lemma A.5. We only present the proof for the fixed-design WB, since the proof for the pairwise bootstrap follows using similar

arguments. We can write $\hat{\phi}^*(k) - \hat{\phi}(k) = \widehat{\Gamma}_k^{-1}(n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}\hat{\varepsilon}_{t,k}^*$, which implies $\|\hat{\phi}^*(k) - \hat{\phi}(k)\| \leq \|\widehat{\Gamma}_k^{-1}\|_1\|(n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}\hat{\varepsilon}_{t,k}^*\|$. Since $\|\widehat{\Gamma}_k^{-1}\|_1 = O_P(1)$, as we argued before, and as we show next, $\|(n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}\hat{\varepsilon}_{t,k}^*\| = O_{P^*}\left(\frac{k^{1/2}}{n^{1/2}}\right)$, it follows that $\|\hat{\phi}^*(k) - \hat{\phi}(k)\| = O_{P^*}\left(\frac{k^{1/2}}{n^{1/2}}\right)$ in probability. To show that $\|(n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}\hat{\varepsilon}_{t,k}^*\| = O_{P^*}\left(\frac{k^{1/2}}{n^{1/2}}\right)$ in probability, it suffices to show that

$$E^*\left(\left\|(n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}\hat{\varepsilon}_{t,k}^*\right\|^2\right) = O_P\left(\frac{k}{(n-k)}\right), \tag{A.6}$$

by the Markov inequality. We can write

$$E^*\left(\left\|(n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}\hat{\varepsilon}_{t,k}^*\right\|^2\right) = E^*\left\{(n-k)^{-2}\sum_{t=1+k}^{n}\sum_{s=1+k}^{n}Y_{t-1,k}'Y_{s-1,k}\hat{\varepsilon}_{t,k}^*\hat{\varepsilon}_{s,k}^*\right\}$$

$$= (n-k)^{-2}\sum_{t=1+k}^{n}\sum_{s=1+k}^{n}Y_{t-1,k}'Y_{s-1,k}E^*(\hat{\varepsilon}_{t,k}^*\hat{\varepsilon}_{s,k}^*)$$

$$= (n-k)^{-2}\sum_{t=1+k}^{n}Y_{t-1,k}'Y_{t-1,k}\hat{\varepsilon}_{t,k}^2 \equiv (n-k)^{-1}\chi_1,$$

where the last inequality follows because $E^*(\hat{\varepsilon}_{t,k}^*\hat{\varepsilon}_{s,k}^*) = 0$ if $t \neq s$ and $E^*(\hat{\varepsilon}_{t,k}^*\hat{\varepsilon}_{s,k}^*) = \hat{\varepsilon}_{t,k}^2$ otherwise. Next, we show that $\chi_1 = O_P(k)$, which in turn implies (A.6). Applying the triangle inequality first and then the Cauchy–Schwartz inequality, we have that

$$|\chi_1| \leq (n-k)^{-1}\sum_{t=1+k}^{n}|Y_{t-1,k}'Y_{t-1,k}||\hat{\varepsilon}_{t,k}^2|$$

$$\leq \left((n-k)^{-1}\sum_{t=1+k}^{n}|Y_{t-1,k}'Y_{t-1,k}|^2\right)^{1/2}\left((n-k)^{-1}\sum_{t=1+k}^{n}\hat{\varepsilon}_{t,k}^4\right)^{1/2}$$

$$= \left((n-k)^{-1}\sum_{t=1+k}^{n}\|Y_{t-1,k}\|^4\right)^{1/2}\left((n-k)^{-1}\sum_{t=1+k}^{n}\hat{\varepsilon}_{t,k}^4\right)^{1/2} \equiv A_1 \cdot A_2.$$

Because $\sup_t E|y_t|^4 \leq C < \infty$, $E\|Y_{t-1,k}\|^4 = O(k^2)$, which implies $A_1 = O_P(k)$. Since $A_2 = O_P(1)$, as we show next, this proves the result. To show that $A_2 = O_P(1)$, note that $\hat{\varepsilon}_{t,k} = \varepsilon_t - \sum_{j=1+k}^{\infty}\phi_j y_{t-j} - (\hat{\phi}(k) - \phi(k))'Y_{t-1,k}$.

By the c_r-inequality (Davidson, 1994, p. 140), we have that

$$(n - k)^{-1} \sum_{t=1+k}^{n} \hat{\varepsilon}_{t,k}^4$$

$$\leq C(n - k)^{-1} \sum_{t=1+k}^{n} \left(\varepsilon_t^4 + \left| \sum_{j=1+k}^{\infty} \phi_j y_{t-j} \right|^4 + \left| (\hat{\phi}(k) - \phi(k))' Y_{t-1,k} \right|^4 \right)$$

$$\equiv B_1 + B_2 + B_3.$$

$B_1 = O_P(1)$ since $E|\varepsilon|^4 \leq \Delta < \infty$ for all t. Next consider B_2. We have that

$$E|B_2| = C(n - k)^{-1} \sum_{t=1+k}^{n} \left[E \left(\left| \sum_{j=1+k}^{\infty} \phi_j y_{t-j} \right|^4 \right) \right]^{\frac{1}{4} \times 4}$$

$$\leq C(n - k)^{-1} \sum_{t=1+k}^{n} \left(\sum_{j=1+k}^{\infty} |\phi_j| \left(E|y_{t-j}|^4 \right)^{1/4} \right)^4 \leq C \left(\sum_{j=1+k}^{\infty} |\phi_j| \right)^4,$$

where the first inequality follows by Minkowski's inequality and the last inequality holds by $E|y_{t-j}|^4 \leq \Delta < \infty$ for all t, j. Thus, by the Markov inequality, it follows that $B_2 = O_P\left(\left(\sum_{j=1+k}^{\infty} |\phi_j| \right)^4 \right) = o_P(1)$ given that $\sum_{j=1}^{\infty} |\phi_j| < \infty$ and $k \rightarrow \infty$. Finally, consider B_3. By the triangle inequality for vector norms, we have that

$$B_3 \leq \|\hat{\phi}(k) - \phi(k)\|^4 (n - k)^{-1} \sum_{t=1+k}^{n} \|Y_{t-1,k}\|^4 = O_P\left(\frac{k^2}{n^2} \right) O_P(k^2)$$

$$= O_P\left(\left(\frac{k^2}{n} \right)^2 \right) = o_P(1),$$

given Lemma A.2, the fact that $\|Y_{t-1,k}\|^4 = O_P(k^2)$ and $k^2/n \rightarrow 0$. $\qquad\square$

Proof of Lemma A.6. We start with the fixed-design WB. Adding and subtracting appropriately yields

$$|r^*| \leq Ck^{1/2} \left\| \widehat{\Gamma}_k^{-1} - \Gamma_k^{-1} \right\|_1 (n - k)^{1/2} k^{-1/2} \left\| (n - k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} \hat{\varepsilon}_{t,k}^* \right\|,$$

given that $\|\ell(k)\|$ and $|v_k^{-1}|$ are bounded, with r^* defined as

$$r^* \equiv \ell(k)'\big(\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\big)v_k^{-1}(n-k)^{-1/2} \sum_{t=1+k}^{n} Y_{t-1,k}\hat{\varepsilon}_{t,k}^*.$$

Since by (A.6), $E^*\big(\big\|(n-k)^{-1}\sum_{t=1+k}^{n} Y_{t-1,k}\hat{\varepsilon}_{t,k}^*\big\|^2\big) = O_P\big(\frac{k}{(n-k)}\big)$, we have that $E^*|r^*| \leq Ck^{1/2}\big\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\big\|_1 O_P(1) = o_P(1)$, by Lemma A.1, see (A.2). By the Markov inequality, for any $\delta > 0$, we have $P^*(|r^*| > \delta) \leq \frac{1}{\delta}E^*(|r^*|) = o_P(1)$ and the desired result follows. For the pairwise bootstrap, simple algebra shows that $r^* = A_1 + A_2$, where

$$A_1 = \ell(k)'\sqrt{n-k}\big(\widehat{\Gamma}_{pb,k}^{*-1} - \widehat{\Gamma}_k^{-1}\big)v_k^{-1}(n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}^*\hat{\varepsilon}_{t,k}^*$$

$$A_2 = \ell(k)'\sqrt{n-k}\big(\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\big)v_k^{-1}(n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}^*\hat{\varepsilon}_{t,k}^*.$$

Consider A_2 first. We have that

$$|A_2| \leq C\|\ell(k)\|k^{1/2}\big\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\big\|k^{-1/2}(n-k)^{1/2}\bigg\|(n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}^*\hat{\varepsilon}_{t,k}^*\bigg\|. \tag{A.7}$$

Next we will show that

$$E^*\bigg(\bigg\|(n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}^*\hat{\varepsilon}_{t,k}^*\bigg\|^2\bigg) = O_P\bigg(\frac{k}{n-k}\bigg), \tag{A.8}$$

which, combined with Lemma A.1 and (A.7), shows that $A_2 = o_{P^*}(1)$ in probability. To prove (A.8), note that

$$E^*\bigg(\bigg\|(n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}^*\hat{\varepsilon}_{t,k}^*\bigg\|^2\bigg) = (n-k)^{-2} \sum_{t=1+k}^{n}\sum_{s=1+k}^{n} E^*\big(Y_{t-1,k}^{*\prime}\hat{\varepsilon}_{t,k}^* Y_{s-1,k}^*\hat{\varepsilon}_{s,k}^*\big).$$

By the properties of the pairwise bootstrap, conditional on the data, $Y_{t-1,k}^{*\prime}\hat{\varepsilon}_{t,k}^*$ is independent of $Y_{s-1,k}^*\hat{\varepsilon}_{s,k}^*$ when $t \neq s$, which implies that

$$E^*\big(Y_{t-1,k}^{*\prime}\hat{\varepsilon}_{t,k}^* Y_{s-1,k}^*\hat{\varepsilon}_{s,k}^*\big) = E^*\big(Y_{t-1,k}^{*\prime}\hat{\varepsilon}_{t,k}^*\big)E^*\big(Y_{s-1,k}^*\hat{\varepsilon}_{s,k}^*\big)$$

$$= \bigg\|(n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}\hat{\varepsilon}_{t,k}\bigg\|^2 = 0,$$

where the last equality holds by the FOC of the optimization problem that defines $\hat{\phi}(k)$. For $t = s$, instead we have

$$E^*\big(Y^{*\prime}_{t-1,k}\hat{\varepsilon}^*_{t,k}Y^*_{s-1,k}\hat{\varepsilon}^*_{s,k}\big) = (n-k)^{-1}\sum_{t=1+k}^{n}Y'_{t-1,k}Y_{t-1,k}\hat{\varepsilon}^2_{t,k}.$$

Thus, the LHS of (A.8) equals $(n-k)^{-2}\sum_{t=1+k}^{n}Y'_{t-1,k}Y_{t-1,k}\hat{\varepsilon}^2_{t,k}$, which is $O_P\big(\frac{k}{n-k}\big)$, as we showed earlier. Next, we show that $A_1 = o_{P^*}(1)$ in probability. We can write

$$|A_1| \leq C\|\ell(k)\|\big\|\widehat{\Gamma}^{*-1}_{pb,k} - \widehat{\Gamma}^{-1}_k\big\|_1(n-k)^{1/2}\bigg\|(n-k)^{-1}\sum_{t=1+k}^{n}Y^*_{t-1,k}\hat{\varepsilon}^*_{t,k}\bigg\|$$

$$\leq C\big\|\widehat{\Gamma}^{*-1}_{pb,k} - \widehat{\Gamma}^{-1}_k\big\|_1(n-k)^{1/2}O_{P^*}\bigg(\frac{k^{1/2}}{(n-k)^{1/2}}\bigg),$$

conditional on the data, given (A.8) and the Markov inequality. So, it suffices to show that $k^{1/2}\big\|\widehat{\Gamma}^{*-1}_{pb,k} - \widehat{\Gamma}^{-1}_k\big\|_1 = o_{P^*}(1)$ in probability, which follows by Lemma A.4.

Proof of Lemma A.7. We follow the proof of Theorem 2.2. For both bootstrap schemes, it suffices that (a) $\big\|\widehat{\Gamma}^{*-1}_k - \Gamma^{-1}_k\big\| = o_{P^*}(1)$ in probability, and that (b) $\|\widehat{B}^*_k - B_k\| = o_{P^*}(1)$, in probability. For the fixed-design WB, $\widehat{\Gamma}^*_k = \widehat{\Gamma}_k$ (since $Y^*_{t-1,k} = Y_{t-1,k}$) and (a) corresponds to showing that $\|\widehat{\Gamma}^{-1}_k - \Gamma^{-1}_k\| = o_P(1)$, which follows by Lemma A.1. For (b), note that $\tilde{\varepsilon}^*_{t,k} = \hat{\varepsilon}^*_{t,k} - Y'_{t-1,k}(\hat{\phi}^*(k) - \hat{\phi}(k))$, where $\hat{\varepsilon}^*_{t,k} = \hat{\varepsilon}_{t,k}\eta_t$. We can write $\|\widehat{B}^*_k - B_k\| \leq \|A^*_1\| + \|A^*_2\| + \|A^*_3\|$, where

$$A^*_1 = (n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}Y'_{t-1,k}\big(\tilde{\varepsilon}^{*2}_{t,k} - \hat{\varepsilon}^{*2}_{t,k}\big),$$

$$A^*_2 = (n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}Y'_{t-1,k}\big(\hat{\varepsilon}^{*2}_{t,k} - \hat{\varepsilon}^2_{t,k}\big),$$

and $A^*_3 = \widehat{B}_k - B_k$. By Theorem 2.2, $\|A^*_3\| = o_P(1)$. Next, we will show that $\|A^*_1\| = O_{P^*}((\frac{k^4}{n})^{1/2}) = o_{P^*}(1)$ in probability, if $\frac{k^4}{n} \to 0$, and $A^*_2 = o_{P^*}(1)$ in probability. Following the proof of Theorem 2.2, we can write $\|A^*_1\| \leq A^*_{11} + A^*_{12}$, where A^*_{11} and A^*_{12} are exactly as A_{11} and A_{12} except that we replace $\hat{\varepsilon}_{t,k}$ with $\tilde{\varepsilon}^*_{t,k}$ and $\varepsilon_{t,k}$ with $\hat{\varepsilon}^*_{t,k}$. For A^*_{11}, we proceed as for A_{11} but replace $\tilde{\varepsilon}^*_{t,k} - \hat{\varepsilon}^*_{t,k}$ with $-Y'_{t-1,k}(\hat{\phi}^*(k) - \hat{\phi}(k))$, yielding

$$A^*_{11} \leq \big\|\hat{\phi}^*(k) - \hat{\phi}(k)\big\|\bigg((n-k)^{-1}\sum_{t=1+k}^{n}\|Y_{t-1,k}\|^6\bigg)^{1/2}\bigg((n-k)^{-1}\sum_{t=1+k}^{n}|\tilde{\varepsilon}^*_{t,k}|^2\bigg)^{1/2}.$$

By Lemma A.5, $\|\hat{\phi}^*(k) - \hat{\phi}(k)\| = O_{P^*}\left(\frac{k^{1/2}}{n^{1/2}}\right)$ in probability. In addition, $E\|Y_{t-1,k}\|^6 = O(k^3)$ (since $E|y_t|^6 \le \Delta < \infty$) and we can show that $(n-k)^{-1}$ $\sum_{t=1+k}^n |\tilde{\varepsilon}_{t,k}^*|^2 = O_{P^*}(1)$ in probability (write $\tilde{\varepsilon}_{t,k}^* = \hat{\varepsilon}_{t,k} + \hat{\varepsilon}_{t,k}(\eta_t - 1) - (\hat{\phi}(k) - \phi(k))'Y_{t-1,k}$, apply the c_r-inequality and show that each term is $O_{P^*}(1)$ in probability). Since $k^4/n \to 0$, it follows that $A_{11}^* = o_{P^*}(1)$ in probability. A_{12}^* can be handled similarly.

To show that $A_2^* = o_{P^*}(1)$ in probability, we show that $E^*(A_2^*) \xrightarrow{P} 0$ and $Var^*(A_2^*) \xrightarrow{P} 0$. Note that $A_2^* = (n-k)^{-1} \sum_{t=1+k}^n Y_{t-1,k}Y_{t-1,k}'\hat{\varepsilon}_{t,k}^2(\eta_t^2 - 1)$ and since $E^*(\eta_t^2) = 1$, $E^*(A_2^*) = 0$. Similarly, using the fact that $E^*((\eta_t^2 - 1)(\eta_s^2 - 1)) = 0$ for $t \ne s$, we can write $Var^*(A_2^*) = (n-k)^{-2} \sum_{t=1+k}^n Y_{t-1,k}Y_{t-1,k}'Y_{t-1,k}Y_{t-1,k}'\hat{\varepsilon}_{t,k}^4\Delta_\eta$, where $\Delta_\eta = E^*[(\eta^2 - 1)^2] < \infty$. By the Cauchy–Schwartz inequality and noting that $\|Y_{t-1,k}\|^8 = O_P(k^4)$, it follows that

$$|Var^*(A_2^*)| \le (n-k)^{-1}\left((n-k)^{-1}\sum_{t=1+k}^n \|Y_{t-1,k}\|^8\right)^{1/2}\left((n-k)^{-1}\sum_{t=1+k}^n |\hat{\varepsilon}_{t,k}|^8\right)^{1/2}$$

$$\Delta_\eta = O_P\left(\frac{k^2}{n}\right),$$

which is $o_P(1)$ if $k^2/n \to 0$.

The proof for the pairwise bootstrap is similar, so we omit the details. Lemmas A.1 and A.4 imply (a). For (b), using $\tilde{\varepsilon}_{t,k}^* = \hat{\varepsilon}_{t,k}^* - Y_{t-1,k}^{*'}(\hat{\phi}^*(k) - \hat{\phi}(k))$, where $\hat{\varepsilon}_{t,k}^* = y_t^* - Y_{t-1,k}^*\hat{\phi}(k)$, we can write $\widehat{B}_k^* - B_k = B_1^* + B_2^* + B_3^* + B_4^*$, where $B_1^* = (n-k)^{-1}\sum_{t=1+k}^n Y_{t-1,k}^*Y_{t-1,k}^{*'}\hat{\varepsilon}_{t,k}^{*2} - \widehat{B}_k$, $B_2^* = \widehat{B}_k - B_k$, $B_3^* = -2(n-k)^{-1}\sum_{t=1+k}^n Y_{t-1,k}^*Y_{t-1,k}^{*'}Y_{t-1,k}^{*'}\hat{\varepsilon}_{t,k}^*(\hat{\phi}^*(k) - \hat{\phi}(k))$, and $B_4^* = (n-k)^{-1}\sum_{t=1+k}^n Y_{t-1,k}^*Y_{t-1,k}^{*'}(Y_{t-1,k}^{*'}(\hat{\phi}^*(k) - \hat{\phi}(k)))^2$. Theorem 2.2 implies that $B_2^* = o_P(1)$. By Lemma A.5 and using arguments similar to those used above, we can show that $B_1^* = o_{P^*}(1)$, $B_3^* = O_{P^*}\left(\frac{k^{4/2}}{n^{1/2}}\right)$, $B_4^* = O_{P^*}\left(\frac{k^3}{n}\right)$. Thus, $\widehat{B}_k^* - B_k = o_{P^*}(1)$ in probability if $\frac{k^4}{n} \to 0$. $\square$

A.2. Proofs of the Theorems

Proof of Theorem 2.1. Given Lemma A.3 and the Asymptotic Equivalence Lemma (cf. White, 2000, Lemma 4.7), the proof proceeds in two steps. First, we show that v_k^2 is bounded above and bounded away from zero. Second, we show that $(n-k)^{-1/2}\sum_{t=1+k}^n \ell(k)'\Gamma_k^{-1}Y_{t-1,k}\varepsilon_t/v_k \Rightarrow N(0,1)$.

Let $P_k' = [b_{1,k}, b_{2,k}, \ldots]$ be a $k \times \infty$-dimensional matrix, where $b_{j,k} = (\psi_{j-1}, \ldots, \psi_{j-k})'$ with $\psi_j = 0$ for $j < 0$ and $\psi_0 = 1$, and let $\Omega = [\alpha_{i,j}]$ be the infinite-dimensional matrix associated with Ω_m as defined by equation (7) of Kuersteiner (2001, p. 368). It follows that $\Gamma_k = \sigma^2 P_k'P_k$ and

$B_k = P_k'\Omega P_k$, implying that

$$v_k^2 = \ell(k)'(\sigma^2 P_k'P_k)^{-1}P_k'\Omega P_k(\sigma^2 P_k'P_k)^{-1}\ell(k) = z'\Omega z,$$

where we let $z' \equiv \ell(k)'(\sigma^2 P_k'P_k)^{-1}P_k'$. For any sequence of real numbers $x = \{x_1, x_2, \dots\}$, define the norm $\|x\|_2 = \left(\sum_{i=1}^\infty |x_i|^2\right)^{1/2}$ and let l^2 denote the space of all sequences that are bounded under this norm. Note that $z'z = \|z\|_2^2 = \sigma^{-2}\ell(k)'\Gamma_k^{-1}\ell(k)$ is bounded above and bounded away from zero (uniformly in k), so $z \in l^2$. Lemma 4.1 of Kuersteiner (2001) shows that Ω^{-1} exists, i.e., $\min_{x'x=1} x'\Omega x > 0$. Under our Assumption 1, which is identical to Kuersteiner's Assumption A1, it follows thus that $v_k^2 \geq (z'z)\min_{x'x=1} x'\Omega x > 0$. The boundedness of v_k^2 also follows by Lemma 4.1 of Kuersteiner and an application of the Cauchy–Schwartz inequality, given that $z \in l^2$ and $\Omega z \in l^2$.

Let $z_{nt} = \ell(k)'\Gamma_k^{-1}Y_{t-1,k}\varepsilon_t/v_k$. To prove that $(n-k)^{-1/2}\sum_{t=1+k}^n z_{nt} \Rightarrow N(0,1)$ we apply a CLT for m.d.s. (cf. Davidson, 1994, p. 383) since $E(z_{nt} \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \ell(k)'\Gamma_k^{-1}Y_{t-1,k}v_k^{-1}E(\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = 0$ under Assumption 1. Hence we need to show that (a) $(n-k)^{-1}\sum_{t=1+k}^n z_{nt}^2 - 1 \overset{P}{\to} 0$, and (b) $\max_{1+k\leq t\leq n}(n-k)^{-1/2}|z_{nt}| \overset{P}{\to} 0$. Lewis and Reinsel (1985) also verify similar conditions for establishing their asymptotic normality result (cf. their Theorem 4). However, the arguments of Lewis and Reinsel (1985) are based on the independence between y_{t-j} and ε_t and do not apply in our context. Here, the simplification afforded by the independence between y_{t-j} and ε_t is not available; we overcome the extra difficulty by exploiting well known properties of joint cumulants. This explains in particular our strengthening of Assumption 1(iii$'$).

We start with (a). Note that $Var(z_{nt}) = \ell(k)'\Gamma_k^{-1}B_k\Gamma_k^{-1}\ell(k)/v_k^2 = 1$. We can write

$$\frac{1}{n-k}\sum_{t=1+k}^n z_{nt}^2 - 1$$

$$= v_k^{-2}\left\{\ell(k)'\Gamma_k^{-1}\left(\frac{1}{n-k}\sum_{t=1+k}^n Y_{t-1,k}Y_{t-1,k}'\varepsilon_t^2 - E\big(Y_{t-1,k}Y_{t-1,k}'\varepsilon_t^2\big)\right)\Gamma_k^{-1}\ell(k)\right\}$$

$$\equiv v_k^{-2}\{\ell(k)'\Gamma_k^{-1}S_k\Gamma_k^{-1}\ell(k)\},$$

with the obvious definition of S_k. Because $v_k^2 > 0$, v_k^{-2} is bounded, and it suffices to show $\ell(k)'\Gamma_k^{-1}S_k\Gamma_k^{-1}\ell(k) \overset{P}{\to} 0$. We have that

$$\left|\ell(k)'\Gamma_k^{-1}S_k\Gamma_k^{-1}\ell(k)\right| \leq \|\ell(k)\|\left\|\Gamma_k^{-1}\right\|_1\|S_k\|_1\left\|\Gamma_k^{-1}\right\|_1\|\ell(k)\| \leq M_2 C_2^2\|S_k\|_1,$$

since $\|\ell(k)\|^2 \leq M_2$ and $\|\Gamma_k^{-1}\|_1 \leq C_2$ uniformly in k. Next we show that $\|S_k\|_1 = o_P(1)$. By the Markov inequality it suffices to show that $E\|S_k\|_1^2 \to 0$

or, that $E\|S_k\|^2 \to 0$, since $\|S_k\|_1^2 \leq \|S_k\|^2$. We have

$$E\|S_k\|_1^2 \leq E\|S_k\|^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} E\big([S_k]_{i,j}^2\big) = \frac{1}{n-k} \sum_{i=1}^{k} \sum_{j=1}^{k} (n-k) E\big([S_k]_{i,j}^2\big),$$

where $[S_k]_{i,j}$ denotes element (i,j) of S_k. Below we use Assumption 1(iii) to bound

$$(n-k) E([S_k]_{i,j}^2) = (n-k) E\left[\left((n-k)^{-1} \sum_{t=1+k}^{n} \big(y_{t-i} y_{t-j} \varepsilon_t^2 - E\big(y_{t-i} y_{t-j} \varepsilon_t^2\big)\big) \right)^2 \right] \tag{A.9}$$

by a constant C, independent of i, j or n, implying that $E\|S_k\|_1^2 \leq C \frac{k^2}{n-k} \to 0$ if $k^2/n \to 0$. To prove (b), note that for any $\delta > 0$ and for some $r > 1$,

$$P\left(\max_{1+k \leq t \leq n} |z_{nt}| > \sqrt{n-k}\,\delta \right) \leq \sum_{t=1+k}^{n} P\big(|z_{nt}| > \sqrt{n-k}\,\delta \big) \leq \sum_{t=1+k}^{n} \frac{E|z_{nt}|^r}{(n-k)^{r/2} \delta^r}. \tag{A.10}$$

Letting $v_{t,k} = \ell(k)' \Gamma_k^{-1} Y_{t-1,k}$, we can write $z_{nt} = v_k^{-1} v_{t,k} \varepsilon_t$, and by the Cauchy–Schwartz inequality it follows that $E|z_{nt}|^r = E|v_{t,k} \varepsilon_t / v_k|^r \leq |v_k^{-1}|^r (E|v_{t,k}|^{2r})^{1/2} (E|\varepsilon_t|^{2r})^{1/2} \leq C(E|v_{t,k}|^{2r})^{1/2}$, since $v_k^{-1} = O(1)$ and $E|\varepsilon_t|^{2r} = O(1)$, for $r \leq 4$. We now prove that $E|v_{t,k}|^{2r} = O(k^r)$. We can write $|v_{t,k}| = |\ell(k)' \Gamma_k^{-1} Y_{t-1,k}| \leq \|\ell(k)\| \|\Gamma_k^{-1}\|_1 \|Y_{t-1,k}\|$, so that $|v_{t,k}|^{2r} \leq M_2^r C_2^r |\sum_{j=1}^{k} y_{t-j}^2|^r$. By an application of the Minkowski inequality, and the fact that $E|y_{t-j}|^{2r} \leq \Delta < \infty$ for all $j = 1, \ldots, k$ and some $r \leq 4$, it follows that $E|v_{t,k}|^{2r} \leq M_2^r C_2^r \Delta k^r$, implying $E|z_{nt}|^r \leq C k^{r/2}$. Using this bound in (A.10) with $r = 3$ implies that the LHS of (A.10) is $O\big(\frac{k^{3/2}}{(n-k)^{1/2}}\big)$, which is $o(1)$ provided $k^3/n \to 0$, as we assume.

To conclude the proof, we show that (A.9) is bounded uniformly in $i, j = 1, \ldots, k$ and n. Define $\psi_j = 0$ for $j < 0$. Using the MA(∞) representation of y_t, we have that (A.9) is equal to

$$(n-k)^{-1} \sum_{t=1+k}^{n} \sum_{s=1+k}^{n} Cov(y_{t-i} y_{t-j} \varepsilon_t^2, y_{s-i} y_{s-j} \varepsilon_s^2)$$

$$= \sum_{l_1, \ldots, l_4 = -\infty}^{\infty} \psi_{l_1} \psi_{l_2} \psi_{l_3} \psi_{l_4} (n-k)^{-1}$$

$$\times \sum_{t=1+k}^{n} \sum_{s=1+k}^{n} Cov\big(\varepsilon_{t-i-l_1} \varepsilon_{t-j-l_2} \varepsilon_t^2, \varepsilon_{s-i-l_3} \varepsilon_{s-j-l_4} \varepsilon_s^2 \big).$$

Next, we show that

$$(n-k)^{-1} \sum_{t=1+k}^{n} \sum_{s=1+k}^{n} Cov\big(\varepsilon_{t-i-l_1}\varepsilon_{t-j-l_2}\varepsilon_t^2, \varepsilon_{s-i-l_3}\varepsilon_{s-j-l_4}\varepsilon_s^2\big) \leq C, \qquad \text{(A.11)}$$

uniformly in $i, j, l_1, \ldots, l_4$, and n, which proves the result given the absolute summability of $\{\psi_j\}$. By an application of Theorem 2.3.2 of Brillinger (1981, p. 21) we can write $Cov(\varepsilon_{t-i-l_1}\varepsilon_{t-j-l_2}\varepsilon_t^2, \varepsilon_{s-i-l_3}\varepsilon_{s-j-l_4}\varepsilon_s^2)$ as the sum of products of cumulants of ε_t of order eight and lower (see also McCullagh's 1987 equation (3.3), p. 39). In particular, if we let $Y_1 = \varepsilon_{t-i-l_1}\varepsilon_{t-j-l_2}\varepsilon_t^2$ and $Y_2 = \varepsilon_{s-i-l_3}\varepsilon_{s-j-l_4}\varepsilon_s^2$, then

$$Cum(Y_1, Y_2) = \sum_{v} Cum(X_{ij} : i,j \in v_1) \ldots Cum(X_{ij} : i,j \in v_p) \qquad \text{(A.12)}$$

where the sum extends over all indecomposable partitions $v = v_1 \cup \cdots \cup v_p$ of the following table:

$$X = \begin{bmatrix} \varepsilon_{t-i-l_1} & \varepsilon_{t-j-l_2} & \varepsilon_t & \varepsilon_t \\ \varepsilon_{s-i-l_3} & \varepsilon_{s-j-l_4} & \varepsilon_s & \varepsilon_s \end{bmatrix}.$$

We let $Cum(\cdot, \ldots, \cdot)$ denote the joint cumulant of the set of random variables involved. By the mean zero property of ε_t only partitions with a number of sets smaller or equal to 4 (i.e., with $p \leq 4$) contribute to the sum in (A.12). For a list of the indecomposable partitions that omit the unit parts (i.e., the blocks with only one element), see Table 2 of McCullagh (1987, p. 259).

Consider $p = 1$, i.e., consider $v = \{\varepsilon_{t-i-l_1}, \varepsilon_{t-j-l_2}, \varepsilon_t, \varepsilon_t, \varepsilon_{s-i-l_3}, \varepsilon_{s-j-l_4}, \varepsilon_s, \varepsilon_s\}$. This term contributes towards the sum with the 8th order joint cumulant $Cum(\varepsilon_{t-i-l_1}, \varepsilon_{t-j-l_2}, \varepsilon_t, \varepsilon_t, \varepsilon_{s-i-l_3}, \varepsilon_{s-j-l_4}, \varepsilon_s, \varepsilon_s)$, which by stationarity can be written as $\kappa_\varepsilon(t-s-i-l_1, t-s-j-l_2, t-s, t-s, -i-l_3, -j-l_4, 0, 0) \equiv \kappa_\varepsilon(\tau-i-l_1, \tau-j-l_2, \tau, \tau, -i-l_3, -j-l_4, 0, 0)$, if we set $\tau = t - s$. Thus, by a change of variables, the contribution of this term to (A.11) is

$$\sum_{\tau=-n+(1+k)}^{n-(1+k)} \left(1 - \frac{|\tau|}{n-k}\right) \kappa_\varepsilon\big(\tau-i-l_1, \tau-j-l_2, \tau, \tau, -i-l_3, -j-l_4, 0, 0\big)$$

$$\leq \sum_{\tau=-\infty}^{\infty} \big|\kappa_\varepsilon\big(\tau-i-l_1, \tau-j-l_2, \tau, \tau, -i-l_3, -j-l_4, 0, 0\big)\big|$$

$$\leq \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \cdots \sum_{\tau_7=-\infty}^{\infty} \big|\kappa_\varepsilon(\tau_1, \tau_2, \ldots, \tau_7, 0)\big| < \infty,$$

by Assumption 1(iii). For $p = 2$ the mean zero property of ε_t implies that only partitions $v = v_1 \cup v_2$ with cardinalities $(\#v_1, \#v_2) \in \{(4,4), (2,6), (3,5)\}$ contribute to (A.12) with a nonzero value, i.e., products of cumulants of orders 2 to 6 enter this term. Here $\#v_i$ is used to denote the number of elements contained in each set v_i. Because the sum is taken over indecomposable partitions there is at least one element of each row of X in at least one set of each partition. This implies that we can express some of the cumulants entering the product as a function of $t - s$. The summability condition Assumption 1(iii$'$) then ensures the boundedness of the contribution of these terms to the sum in (A.11). The same reasoning can be applied for $p = 3$, where $(\#v_1, \#v_2, \#v_3) \in \{(2,2,4), (3,3,2)\}$, and for $p = 4$, where $(\#v_1, \#v_2, \#v_3, \#v_4) \in \{(2,2,2,2)\}$.

Proof of Theorem 2.2.   Adding and subtracting appropriately, we can write

$$\hat{v}_k^2 - v_k^2 = \ell(k)' \widehat{\Gamma}_k^{-1} \big(\widehat{B}_k - B_k\big) \Gamma_k^{-1} \ell(k) + \ell(k)' \big(\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\big) B_k \Gamma_k^{-1} \ell(k)$$
$$+ \ell(k)' \widehat{\Gamma}_k^{-1} \widehat{B}_k \big(\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\big) \ell(k).$$

Since $\|\ell(k)\|$, $\|\Gamma_k^{-1}\|_1$, and $\|B_k\|_1$ are bounded, and $\|\widehat{\Gamma}_k^{-1}\|_1$ and $\|\widehat{B}_k\|_1$ are bounded in probability, it suffices that $\|\widehat{\Gamma}_k^{-1} - \Gamma_k^{-1}\| = o_P(1)$ (which follows by Lemma A.1), and that $\|\widehat{B}_k - B_k\| = o_P(1)$, which we prove next. We can write $\|\widehat{B}_k - B_k\| \le A_1 + A_2 + A_3$, where

$$A_1 = \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y_{t-1,k}' \big(\hat{\varepsilon}_{t,k}^2 - \varepsilon_{t,k}^2\big) \right\|,$$

$$A_2 = \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y_{t-1,k}' \big(\varepsilon_{t,k}^2 - \varepsilon_t^2\big) \right\|,$$

$$A_3 = \left\| (n-k)^{-1} \sum_{t=1+k}^{n} \big(Y_{t-1,k} Y_{t-1,k}' \varepsilon_t^2 - E\big(Y_{t-1,k} Y_{t-1,k}' \varepsilon_t^2\big)\big) \right\|.$$

$A_3 = O_P\big(\frac{k}{(n-k)^{1/2}}\big)$ under our conditions (see proof of Theorem 2.1; A_3 here corresponds to S_k there). Next, we will show that $A_1 = O_P\big(\big(\frac{k^4}{n}\big)^{1/2}\big)$, which is $o_P(1)$ if $\frac{k^4}{n} \to 0$, and $A_2 = O_P\big(k \sum_{j=1+k}^{\infty} |\phi_j|\big)$, which is $o_P(1)$ if $n^{1/2} \sum_{t=1+k}^{n} |\phi_j| \to 0$ and $k^2/n \to 0$. Consider A_1. Write

$$A_1 = \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y_{t-1,k}' (\hat{\varepsilon}_{t,k} - \varepsilon_{t,k})(\hat{\varepsilon}_{t,k} + \varepsilon_{t,k}) \right\|$$

$$\leq \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y'_{t-1,k} \hat{\varepsilon}_{t,k} (\hat{\varepsilon}_{t,k} - \varepsilon_{t,k}) \right\|$$

$$+ \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y'_{t-1,k} \varepsilon_{t,k} (\hat{\varepsilon}_{t,k} - \varepsilon_{t,k}) \right\| \equiv A_{11} + A_{12}.$$

We will consider only A_{11}. The analysis of A_{12} follows by similar arguments. Replacing $\hat{\varepsilon}_{t,k} - \varepsilon_{t,k}$ with $-Y'_{t-1,k}(\hat{\phi}(k) - \phi(k))$ and applying the triangle inequality and the Cauchy–Schwartz inequality yields

$$A_{11} \leq \left\| \hat{\phi}(k) - \phi(k) \right\| (n-k)^{-1} \sum_{t=1+k}^{n} \| Y_{t-1,k} \|^3 |\hat{\varepsilon}_{t,k}|$$

$$\leq \left\| \hat{\phi}(k) - \phi(k) \right\| \left((n-k)^{-1} \sum_{t=1+k}^{n} \| Y_{t-1,k} \|^6 \right)^{1/2} \left((n-k)^{-1} \sum_{t=1+k}^{n} |\hat{\varepsilon}_{t,k}|^2 \right)^{1/2}$$

$$= O_P\left(\frac{k^{1/2}}{n^{1/2}} \right) O_P(k^{3/2}) O_P(1) = O_P\left(\frac{k^{4/2}}{n^{1/2}} \right),$$

where the second equality holds by Lemma A.2, the fact that $E\| Y_{t-1,k} \|^6 = O(k^3)$ (since $E|y_t|^6 \leq \Delta < \infty$) and $(n-k)^{-1} \sum_{t=1+k}^{n} |\hat{\varepsilon}_{t,k}|^2 = O_P(1)$ (cf. proof of Lemma A.6). Since $k^4/n \to 0$, it follows that $A_{11} = o_P(1)$.

Next, take A_2. Since $\varepsilon_{t,k} = \varepsilon_t - \sum_{j=1+k}^{\infty} \phi_j y_{t-j}$ and $\varepsilon_{t,k}^2 - \varepsilon_t^2 = (\varepsilon_{t,k} - \varepsilon_t)(\varepsilon_{t,k} + \varepsilon_t)$, we can write

$$A_2 \leq \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y'_{t-1,k} \varepsilon_{t,k} \left(- \sum_{j=1+k}^{\infty} \phi_j y_{t-j} \right) \right\|$$

$$+ \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y'_{t-1,k} \varepsilon_t \left(- \sum_{j=1+k}^{\infty} \phi_j y_{t-j} \right) \right\| \equiv A_{21} + A_{22}.$$

Consider A_{22}. The analysis of A_{21} is analogous. An application of the triangle inequality and the Cauchy–Schwartz inequality yields

$$A_{22} \leq (n-k)^{-1} \sum_{t=1+k}^{n} \| Y_{t-1,k} \|^2 |\varepsilon_t| \sum_{j=1+k}^{\infty} |\phi_j| |y_{t-j}|$$

$$\leq \sum_{j=1+k}^{\infty} |\phi_j| (n-k)^{-1} \sum_{t=1+k}^{n} \| Y_{t-1,k} \|^2 |\varepsilon_t y_{t-j}|$$

$$\leq \sum_{j=1+k}^{\infty} |\phi_j| \left((n-k)^{-1} \sum_{t=1+k}^{n} \|Y_{t-1,k}\|^4 \right)^{1/2} \left((n-k)^{-1} \sum_{t=1+k}^{n} |\varepsilon_t y_{t-j}|^2 \right)^{1/2}$$

$$= O_P \left(k \sum_{j=1+k}^{\infty} |\phi_j| \right),$$

given that $E\|Y_{t-1,k}\|^4 = O_P(k^2)$ and that $E|\varepsilon_t y_{t-j}|^2 \leq \Delta < \infty$ for all t, j. But $k \sum_{j=1+k}^{\infty} |\phi_j| = \frac{k}{n^{1/2}} n^{1/2} \sum_{j=1+k}^{\infty} |\phi_j| \to 0$ under Assumption 2(ii) and $k/n^{1/2} \to 0$, which implies $A_{22} = o_P(1)$.

Proof of Theorem 3.1. Given Lemmas A.6 and A.7, it suffices to show that with probability approaching one $(n-k)^{-1/2} \sum_{t=1+k}^{n} w_{nt}^* \Rightarrow^{d_{P^*}} N(0,1)$, where $w_{nt}^* = \ell(k)' \Gamma_k^{-1} v_k^{-1} Y_{t-1,k}^* \hat{\varepsilon}_{t,k}^*$. We start with the fixed-design WB. Because conditional on the original data w_{nt}^* is an independent (not identically distributed) array of random variables, we will apply Lyapunov's theorem (Durrett, 1996, p. 121) conditional on the data. Note that $E^*\left((n-k)^{-1/2} \sum_{t=1+k}^{n} w_{nt}^* \right) = 0$ and $\bar{\sigma}_n^{*2} \equiv Var^*\left((n-k)^{-1/2} \sum_{t=1+k}^{n} w_{nt}^* \right) = \frac{v_k^{*2}}{v_k^2}$, with $v_k^2 \equiv \ell(k)' \Gamma_k^{-1} B_k \Gamma_k^{-1'} \ell(k)$ and $v_k^{*2} \equiv \ell(k)' \Gamma_k^{-1} \widehat{B}_k \Gamma_k^{-1'} \ell(k)$, $\widehat{B}_k = (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y_{t-1,k}' \hat{\varepsilon}_{t,k}^2$. The proof consists of two steps: *Step 1*. Show $\bar{\sigma}_n^{*2} \xrightarrow{P} 1$, or equivalently, $v_k^{*2} - v_k^2 \xrightarrow{P} 0$, given that v_k^2 is bounded away from zero. *Step 2*. Verify Lyapunov's condition, i.e., for some $r > 1$, $(n-k)^{-r} \sum_{t=1+k}^{n} E^* |w_{nt}^*|^{2r} \xrightarrow{P} 0$.

Proof of Step 1. Note that $|v_k^{*2} - v_k^2| \leq \|\ell(k)\|^2 \|\Gamma_k^{-1}\|_1^2 \|\widehat{B}_k - B_k\| \leq C\|\widehat{B}_k - B_k\|$, given that $\|\ell(k)\|$ and $\|\Gamma_k^{-1}\|_1$ are bounded, and that by Theorem 2.2, $\|\widehat{B}_k - B_k\| = o_P(1)$.

Proof of Step 2. We will show that Lyapunov's condition holds with $r = \frac{3}{2}$. Let $v_{t,k} = \ell(k)' \Gamma_k^{-1} Y_{t-1,k}$. Then $w_{nt}^* = v_k^{-1} v_{t,k} \hat{\varepsilon}_{t,k}^*$. We have that

$$(n-k)^{-r} \sum_{t=1+k}^{n} E^* |w_{nt}^*|^{2r} = (n-k)^{-r} \sum_{t=1+k}^{n} |v_k|^{-2r} |v_{t,k}|^{2r} |\hat{\varepsilon}_{t,k}|^{2r} E^* |\eta_t|^{2r}$$

$$\leq C(n-k)^{-r} \sum_{t=1+k}^{n} |v_{t,k}|^{2r} |\hat{\varepsilon}_{t,k}|^{2r}$$

$$\leq C(n-k)^{1-r} \left((n-k)^{-1} \sum_{t=1+k}^{n} |v_{t,k}|^{4r} \right)^{1/2}$$

$$\times \left((n-k)^{-1} \sum_{t=1+k}^{n} |\hat{\varepsilon}_{t,k}|^{4r} \right)^{1/2}. \qquad (A.13)$$

By an argument similar to that used in Lemma A.6, we can show that $(n-k)^{-1}\sum_{t=1+k}^{n}|\hat{\varepsilon}_{t,k}|^{4r} \le (n-k)^{-1}\sum_{t=1+k}^{n}|\varepsilon_{t}|^{4r} + O_{P}\big(\big(\sum_{j=1+k}^{\infty}|\phi_{j}|\big)^{4r}\big) + O_{P}\big(\frac{k^{4r}}{n^{4r/2}}\big)$, provided $E|y_{t}|^{4r} \le \Delta < \infty$ for all t. Thus, with $r=\frac{3}{2}$ it follows that $(n-k)^{-1}\sum_{t=1+k}^{n}|\hat{\varepsilon}_{t,k}|^{4r}= O_{P}(1)$. Similarly, we can show that $(n-k)^{-1}\sum_{t=1+k}^{n}|v_{t,k}|^{4r}= O_{P}(k^{2r}) = O_{P}(k^{3})$, with $r=\frac{3}{2}$. Hence, the LHS of (A.13) is $O_{P}\big(\big(\frac{k^{3}}{n-k}\big)^{1/2}\big) = o_{P}(1)$ if $k^{3}/n \to 0$.

The proof for the pairwise bootstrap follows similarly. In particular, to show $(n-k)^{-1/2}\sum_{t=1+k}^{n}w_{nt}^{*} \Rightarrow^{d_{P^{*}}} N(0,1)$ in probability, where $w_{nt}^{*} = \ell(k)'v_{k}^{-1}\Gamma_{k}^{-1}Y_{t-1,k}^{*}\hat{\varepsilon}_{t,k}^{*}$, we note that w_{nt}^{*} is independent (conditional on the original data) with $E^{*}(w_{nt}^{*}) = 0$ and $Var^{*}((n-k)^{-1/2}\sum_{t=1+k}^{n}w_{nt}^{*}) = \frac{v_{k}^{*2}}{v_{k}^{2}}$, where $v_{k}^{*2} = \ell(k)'\Gamma_{k}^{-1}\widehat{B}_{k}\Gamma_{k}^{-1}\ell(k)$, as for the fixed-design WB. Thus, $\frac{v_{k}^{*2}}{v_{k}^{2}} \to 1$ in probability, and we only need to check Lyapunov's condition. Using the properties of the pairwise bootstrap yields, for some $r > 1$,

$$
\begin{aligned}
E^{*}\big(|w_{nt}^{*}|^{2r}\big) &\le \|\ell(k)\|^{2r}\|\Gamma_{k}^{-1}\|_{1}^{2r}|v_{k}^{-1}|^{2r}E^{*}\big(\big\|Y_{t-1,k}^{*}\hat{\varepsilon}_{t,k}^{*}\big\|^{2r}\big) \\
&\le C(n-k)^{-1}\sum_{t=1+k}^{n}\|Y_{t-1,k}\hat{\varepsilon}_{t,k}\|^{2r} \\
&\le C\left((n-k)^{-1}\sum_{t=1+k}^{n}\|Y_{t-1,k}\|^{4r}\right)^{1/2}\left((n-k)^{-1}\sum_{t=1+k}^{n}|\hat{\varepsilon}_{t,k}|^{4r}\right)^{1/2} \\
&= O_{P}(k^{2r/2})O_{P}(1) = O_{P}(k^{r}),
\end{aligned}
$$

provided $\sup_{t} E|\varepsilon_{t}|^{4r} < C < \infty$. Choosing $r=\frac{3}{2}$ verifies Lyapunov's condition provided $k^{3}/n \to 0$.

ACKNOWLEDGMENTS

REFERENCES

Andrews, D. W. K. (1991). Heteroskedastic and autocorrelation consistent covariance matrix estimation. *Econometrica* 59:817–858.

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

Berk, K. (1974). Consistent autoregressive spectral estimates. *Annals of Statistics* 2:489–502.

Bhansali, R. (1978). Linear prediction by autoregressive model fitting in the time domain. *Annals of Statistics* 6:224–231.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307–327.

Braun, P. A., Mittnik, S. (1993). Misspecifications in vector autoregressions and their effects on impulse responses and variance decompositions. *Journal of Econometrics* 59:319–341.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory.* San Francisco: Siam.

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

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* 3:123–148.

Canova, F. (1995). Vector autoregressions: specification, estimation, inference and forecasting. In: Pesaran, H., Wickens, M., eds. *The Handbook of Applied Econometrics.* Vol. 1. London: Basil Blackwell.

Choi, E., Hall, P. (2000). Bootstrap confidence regions computed on autoregressions of arbitrary order. *Journal of the Royal Statistical Society, Series B* 62:461–477.

Davidson, J. (1994). *Stochastic Limit Theory.* New York: Oxford University Press.

Diebold, F. X., Kilian, L. (2001). Measuring predictability: theory and macroeconomic applications. *Journal of Applied Econometrics* 16:657–669.

Durrett, R. (1996). *Probability: Theory and Examples.* 2nd ed. Belmont, California: Duxury Press.

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.

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

Galbraith, J. W. (2003). Content horizons for univariate time-series forecasts. *International Journal of Forecasting* 19:43–55.

Geman, S., Hwang, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Annals of Statistics* 10:401–414.

Giné, E., Zinn, J. (1990). Bootstrapping general empirical measures. *Annals of Probability* 18:851–869.

Gonçalves, S., Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 123:89–120.

Gonçalves, S., White, H. (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics* 119:199–220.

Grenander, U. (1981). *Abstract Inference.* New York: Wiley.

Hannan, E. J. (1970). *Multiple Time Series.* New York: Wiley.

Hannan, E. J., Deistler, M. (1988). *The Statistical Theory of Linear Systems.* New York: Wiley.

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

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

Hong, Y., Lee, Y.-J. (2003). Consistent testing for serial correlation of unknown form under general conditional heteroskedasticity. Manuscript, Department of Economics, Cornell University.

Inoue, A., Kilian, L. (2002). Bootstrapping smooth functions of slope parameters and innovation variances in VAR(∞) models. *International Economic Review* 43:309–332.

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

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17:1217–1241.

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

Lewis, R., Reinsel, G. (1985). Prediction of multivariate time series of autoregressive model fitting. *Journal of Multivariate Analysis* 16:393–411.

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

Lütkepohl, H. (1988a). Asymptotic distribution of the moving-average coefficients of an estimated vector autoregressive process. *Econometric Theory* 4:77–85.

Lütkepohl, H. (1988b). Prediction tests for structural stability. *Journal of Econometrics* 39:267–296.

Lütkepohl, H., Poskitt, D. S. (1991). Estimating orthogonal impulse responses via vector autoregressive models. *Econometric Theory* 7:487–496.

Lütkepohl, H., Poskitt, D. S. (1996). Testing for causation using infinite order vector autoregressive processes. *Econometric Theory* 12:61–87.

Lütkepohl, H., Saikkonen, P. (1997). Impulse response analysis in infinite-order cointegrated vector autoregressive processes. *Journal of Econometrics* 81:127–157.

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

McCullagh, P. (1987). *Tensor Methods in Statistics*. London: Chapman and Hall.

Ng, S., Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90:268–281.

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

Paparoditis, E. (1996). Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *Journal of Multivariate Analysis* 57:277–296.

Politis, D. N., Romano, J. P., Wolf, M. (1999). *Subsampling*. New York: Springer-Verlag.

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory* 8:1–27.

Saikkonen, P., Lütkepohl, H. (1996). Infinite-order cointegrated vector autoregressive processes. *Econometric Theory* 12:814–844.

Saikkonen, P., Luukkonen, R. (1997). Testing cointegration in infinite order vector autoregressive processes. *Journal of Econometrics* 81:93–126.

Stock, J. H., Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives* 15(Fall):101–115.

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. (2000). *Asymptotic Theory for Econometricians*. Revised Edition. San Diego: Academic Press.

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