2003s-28

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

Sílvia Gonçalves, Lutz Kilian

Série Scientifique
Scientific Series

Montréal
Mai 2003

CIRANO
Centre interuniversitaire de recherche en analyse des organisations

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de la Recherche, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de la Recherche, de la Science et de la Technologie, and grants and research mandates obtained by its research teams.

ISSN 1198-8177

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

Sílvia Gonçalves[†], *Lutz Kilian*[‡]

Résumé / ***Abstract***

La contribution de ce papier est double. Premièrement, nous dérivons les propriétés asymptotiques (convergence et normalité asymptotique) des estimateurs de moindre carrés ordinaires des paramètres autoregressifs dans le cadre de modèles autoregressifs d'ordre infini dont les innovations sont des différences de martingale possiblement hétéroscédastiques. Deuxièmement, nous démontrons la validité asymptotique d'une méthode de bootstrap dans ce contexte. Nos résultats justifient théoriquement l'utilisation de la loi asymptotique ou l'utilisation de la distribution de bootstrap comme méthodes d'inférence pour les paramètres autoregressifs ou les fonctions de ceux-ci.

Mots clés : autoregression d'ordre infini, hétéroscédasticité conditionnelle, wild bootstrap, bootstrap par couples.

The main contribution of this paper is twofold. First, we derive the consistency and asymptotic normality of the estimated 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. To the best of our knowledge, the asymptotic distribution of the least-squares estimator has not been derived under these conditions. Second, we show that a suitably constructed bootstrap estimator will have the same limit distribution as the OLS estimator. Our results provide theoretical justification for the use of either the conventional asymptotic approximation or the bootstrap approximation of the distribution of smooth functions of autoregressive parameters.

Keywords: *infinite autoregression, conditional heteroskedasticity, wild bootstrap, pairwise bootstrap.*

---

* Part of this research was conducted while the second author was serving as an advisor to the European Central Bank.

[†] CIREQ, CIRANO and Département de sciences économiques, Université de Montréal.

[‡] University of Michigan and CEPR.

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. 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).

In the literature on sieve approximations of autoregressions 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. In this paper, we relax this i.i.d. assumption. Instead we postulate that the innovations driving the linear AR(∞) process follow a martingale difference sequence subject to possible conditional heteroskedasticity of unknown form. Our analysis covers data generating processes that belong to the stationary linear AR(∞) class. This class includes finite-order stationary ARMA processes as a special case.

The main contribution of this paper is twofold. First, we derive the consistency and asymptotic normality of the estimated autoregressive sieve parameters under weak conditions on the form of conditional heteroskedasticity. To the best of our knowledge, the asymptotic distribution of the least-squares estimator has not been derived under these conditions. Our analysis shows that the asymptotic distribution of estimated autoregressive parameters derived under the assumption of an AR(∞) 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(∞) processes (see Eicker 1963; White 1980, and Nicholls and Pagan 1983).

The second main contribution of this paper is to extend the application of the bootstrap to AR(∞) processes with possible conditional heteroskedasticity. We show that suitably constructed bootstrap estimators will have the same limit distribution as the ordinary least-squares (OLS) estimator. 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(∞) with conditional heteroskedasticity.

As our results 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 (2003).

In this paper, 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 (2003), because such methods are more restrictive than the fixed-design WB and the pairwise bootstrap. Specifically, as shown by Gonçalves and Kilian (2003), 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 remainder of the paper is organized as follows. In section 2 we present the main theoretical results for the OLS estimator. In section 3, we develop the theoretical results for the corresponding bootstrap estimator. In section 4, we extend the results to smooth functions of autoregressive parameters. In section 5, we discuss implications of our results for a number of research areas. Details of the proofs are provided in the appendix.

2. Asymptotic Theory for the OLS 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 OLS estimator of the autoregressive slope parameters and of smooth functions of those parameters. For concreteness, we focus on univariate autoregressions. Multivariate generalizations of our results should be possible at the cost of more complicated notation.

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

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

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. Instead of the usual i.i.d. assumption, we postulate that $\{\varepsilon_t\}$ is a possibly conditionally heteroskedastic martingale difference sequence. We further restrict the temporal dependence of $\{\varepsilon_t\}$ by controlling the behavior of its high order cumulants. For $j \in \mathbb{N}$, let $\kappa_\varepsilon(0, l_1, \ldots, l_{j-1})$ denote the j^{th} order joint cumulant of $(\varepsilon_0, \varepsilon_{l_1}, \ldots, \varepsilon_{l_{j-1}})$ (see Brillinger, 1981, p. 19), where $l_1, \ldots, l_{j-1}$ are integers. We make the following assumption:

Assumption 1. (i) $\{\varepsilon_t\}$ is strictly stationary and ergodic such that $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0$, a.s., where $\mathcal{F}_{t-1} = \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$ is the σ-field generated by $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$; (ii) $E(\varepsilon_t^2) = \sigma^2 < \infty$; and (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 1 allows for conditional heteroskedasticity of unknown form in ε_t and can be verified for ARCH, GARCH, and stochastic volatility models, provided that $E(\varepsilon_t^4) < \infty$, which typically requires additional restrictions on the innovation distribution and on the parameters defining these processes (cf. Kuersteiner, 2001 and 2002).

Assumption 1 (iii) imposes a summability condition on the fourth-order joint cumulants of ε_t, $\kappa_\varepsilon(0, l_1, l_2, l_3)$. By stationarity of ε_t, these are a function only of l_1, l_2 and l_3, and not of t, i.e. $\kappa_\varepsilon(t, t+l_1, t+l_2, t+l_3) = \kappa_\varepsilon(0, l_1, l_2, l_3)$ for all t. The mean zero assumption on ε_t implies $\kappa_\varepsilon(0, l_1, l_2, l_3) = E(\varepsilon_t \varepsilon_{t+l_1} \varepsilon_{t+l_2} \varepsilon_{t+l_3}) - E(\varepsilon_t \varepsilon_{t+l_1}) E(\varepsilon_{t+l_2} \varepsilon_{t+l_3}) - E(\varepsilon_t \varepsilon_{t+l_2}) E(\varepsilon_{t+l_1} \varepsilon_{t+l_3}) - E(\varepsilon_t \varepsilon_{t+l_3}) E(\varepsilon_{t+l_1} \varepsilon_{t+l_2})$. If ε_t is i.i.d., then $\kappa_\varepsilon(0, l_1, l_2, l_3) = E(\varepsilon_t^4) - 3(E(\varepsilon_t^2))^2$ for $l_1 = l_2 = l_3 = 0$ and zero otherwise. With higher-order dependence of ε_t, the fourth-order cumulants have a more complicated structure. The summability assumption on $\kappa_\varepsilon(0, l_1, l_2, l_3)$ restricts the dependence in the error process. This assumption is standard in the time series literature (see Andrews, 1991, pp. 823-824, and the references therein); it is implied by an α-mixing plus a (fourth order) moment condition on ε_t (see, e.g., Lemma 1 of Andrews, 1991, or Remark A.1 of Künsch, 1989).

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 OLS estimators

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

where

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

The population analogues of $\hat{\Gamma}_k$ and $\hat{\Gamma}_{k,1}$ are $\Gamma_k = E\left(Y_{t-1,k} Y_{t-1,k}'\right)$ and $\Gamma_{k,1} = E(Y_{t-1,k} y_t)$, respectively. As in Lewis and Reinsel (1985), we impose the following conditions on k:

Assumption 2. k is chosen as a function of n such that (i) $\frac{k^2}{n} \to 0$ 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 $n^{1/2}$. Assumption 2 (ii) imposes a lower bound condition on k. In particular, k should be large enough as to ensure that the ϕ_j are effectively zero for all $j > k$, for some k. This assumption allows us to approximate the AR(∞) process by a finite-order AR model.

Our first result is as follows:

Theorem 2.1. *Let $\{y_t\}$ satisfy (2.1) and assume Assumptions 1 and 2 hold. Then* $\left\|\hat{\phi}(k) - \phi(k)\right\| = O_P\left(\frac{k^{1/2}}{n^{1/2}}\right)$.

Here and throughout the paper, for an arbitrary vector a we let $\|a\| = (a'a)^{1/2}$ denote the Euclidean norm of a; similarly, for any matrix A, we let $\|A\|$ denote the matrix norm defined by $\|A\|^2 = tr(A'A)$.

Theorem 2.1 implies the consistency of $\hat{\phi}(k)$ for $\phi(k)$, since $\frac{k^{1/2}}{n^{1/2}} \to 0$ under Assumption 2. It is a univariate extension of Theorem 2.1 of Paparoditis (1996) to the conditionally heteroskedastic AR(∞) case. For the multivariate context, Lewis and Reinsel's (1985) Theorem 1 contains the weaker result that $\left\|\hat{\phi}(k) - \phi(k)\right\| = o_P(1)$ when ε_t is i.i.d. and k is chosen such that $k^2/n \to 0$ and $k^{1/2} \sum_{j=1+k}^{\infty} |\phi_j| \to 0$. The latter condition imposes a lower bound on k that is less restrictive than our Assumption 2 (ii). This lower bound condition is satisfied by any $\{y_t\}$ generated by (2.1), provided that $|\phi_j| \to 0$ at a geometric rate as $j \to \infty$, independently of the rate at which k increases with n. In particular, all stationary and invertible ARMA(p, q) processes will satisfy this condition, if $k \to \infty$. Here, we impose the stronger Assumption 2 (ii) to ensure consistency of $\hat{\phi}(k)$ for $\phi(k)$ at rate $\frac{k^{1/2}}{n^{1/2}}$, which will prove useful for establishing our bootstrap results in the next section. Assumption 2 (ii) is standard in the literature (cf. Berk, 1974, Theorem 2, Lewis and Reinsel, 1985, Theorem 4). It is also used below to obtain the limiting distribution of (a linear combination of) $\sqrt{n-k}\left(\hat{\phi}(k) - \phi(k)\right)$. As Ng and Perron (1995) remark, Assumption 2 (ii) amounts to requiring that k grow at least at a polynomial rate with n, as opposed to the rate $\log(n)$, for example. Ng and Perron (1995) discuss implications of this condition for the choice of k based on alternative lag-order selection criteria.

To derive the asymptotic distribution of the estimated autoregressive parameters we impose an additional summability condition on the eighth-order cumulants of ε_t, thus further restricting the dependence in the error process. We also impose a stronger upper bound condition on k, similar to Lewis and Reinsel (1985). Thus, we strengthen Assumption 1 (iii) and Assumption 2 (i) as follows:

Assumption 1. (iii$'$) $\sum_{l_1=-\infty}^{\infty} \cdots \sum_{l_7=-\infty}^{\infty} |\kappa_\varepsilon (0, l_1, \ldots, l_7)| < \infty$.

Assumption 2. (i$'$) k is chosen as a function of n such that $\frac{k^3}{n} \to 0$ as $k, n \to \infty$.

Theorem 2.2. *Let $\{y_t\}$ satisfy (2.1) and assume Assumptions 1 and 2 hold, strengthened by Assumptions 1 (iii$'$) and 2 (i$'$). Let $\ell(k)$ be an arbitrary sequence of $k \times 1$ vectors satisfying $0 < M_1 \leq \|\ell(k)\|^2 \leq M_2 < \infty$, and let $v_k^2 = \ell(k)' \Gamma_k^{-1} B_k \Gamma_k^{-1} \ell(k)$, with $\Gamma_k = E\left(Y_{t-1,k} Y_{t-1,k}'\right)$ and $B_k = E\left(Y_{t-1,k} Y_{t-1,k}' \varepsilon_t^2\right)$. If $v_k^2 > 0$ for all k, then*

$$\ell(k)' \sqrt{n-k} \left(\hat{\phi}(k) - \phi(k)\right) / v_k \Rightarrow N(0,1),$$

where $\Rightarrow$ denotes convergence in distribution.

Theorem 2.2 extends (the univariate version of) Theorem 4 of Lewis and Reinsel (1985) to the martingale difference sequence case, given the same lower and upper bounds on the rate of growth of k.

As in Berk (1974) and Lewis and Reinsel (1985), we show in the Appendix (cf. Lemma A.3) that $\ell(k)' \sqrt{n-k} \left(\hat{\phi}(k) - \phi(k)\right)$ is asymptotically equivalent to $\ell(k)' \Gamma_k^{-1} \left((n-k)^{-1/2} \sum_{t=1+k}^{n} Y_{t-1,k} \varepsilon_t\right)$. We then apply a CLT for martingale difference sequences to $\ell(k)' \Gamma_k^{-1} \left((n-k)^{-1/2} \sum_{t=1+k}^{n} Y_{t-1,k} \varepsilon_t\right)$. In particular, we prove that $S_k \equiv (n-k)^{-1} \sum_{t=1+k}^{n} \left(Y_{t-1,k} Y_{t-1,k}' \varepsilon_t^2 - E\left(Y_{t-1,k} Y_{t-1,k}' \varepsilon_t^2\right)\right)$ vanishes in probability by applying a mean square convergence argument (i.e. we show $E\|S_k\|^2 \to 0$). This explains the need to introduce Assumption 1 (iii$'$).

According to Theorem 2.2, under our assumptions the asymptotic variance of $\ell(k)' \sqrt{n-k} \left(\hat{\phi}(k) - \phi(k)\right)$ 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 OLS inference for AR(∞) 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 $\alpha_{l_1,l_2} = E\left(\varepsilon_{t-l_1} \varepsilon_{t-l_2} \varepsilon_t^2\right)$, 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 the 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

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

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 2.2 to be useful in applications. Under our assumptions, a consistent estimator of Γ_k is given by $\hat{\Gamma}_k = (n-k)^{-1}\sum_{t=1+k}^{n} Y_{t-1,k}Y'_{t-1,k}$ (see Lemma A.2 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, $\hat{B}_k = (n-k)^{-1}\sum_{t=1+k}^{n} Y_{t-1,k}Y'_{t-1,k}\hat{\varepsilon}_{t,k}^2$, where $\hat{\varepsilon}_{t,k} = y_t - Y'_{t-1,k}\hat{\phi}(k)$ is the OLS residual of the autoregressive sieve. Our next result shows that $\hat{v}_k^2 = \ell(k)'\hat{\Gamma}_k^{-1}\hat{B}_k\hat{\Gamma}_k^{-1}\ell(k)$ is a consistent estimator of v_k^2 under the same assumptions on ε_t as in Theorem 2.2, 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 2.3. *Under the assumptions of Theorem 2.2, if in addition k satisfies $k \to \infty$ as $n \to \infty$ such that $k^4/n \to 0$, then* $\left|\hat{v}_k^2 - v_k^2\right| = o_P(1)$.

3. Asymptotic Theory for the Bootstrap OLS Estimator

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.

3.1. Fixed-design Wild Bootstrap for Sieve Autoregressions

The fixed-design WB consists of the following steps:

Step 1 Estimate an approximating AR(k) model by OLS and obtain OLS residuals

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

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

Step 2 Generate WB residuals according to

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

with $\eta_t \sim$ i.i.d.$(0, 1)$ and $E^* |\eta_t|^4 \leq \Delta < \infty$. One possible choice is $\eta_t \sim$ 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

$$y_t^* = Y_{t-1,k}'\hat{\phi}(k) + \hat{\varepsilon}_{t,k}^*, \quad \text{for } t = 1 + k, \ldots, n.$$

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

According to the previous algorithm, $\hat{\phi}_{fwb}^*(k) = \hat{\Gamma}_k^{-1}\hat{\Gamma}_{fwb,k,1}^*$, where $\hat{\Gamma}_k$ is defined as before and $\hat{\Gamma}_{fwb,k,1}^* = (n-k)^{-1}\sum_{t=1+k}^{n} Y_{t-1,k}y_t^*$, with y_t^* as described in Step 3. Next, we show that the conditional distribution of $\ell(k)'\sqrt{n-k}\left(\hat{\phi}_{fwb}^*(k) - \hat{\phi}(k)\right)$ can be used to approximate the true but unknown finite-sample distribution of $\ell(k)'\sqrt{n-k}\left(\hat{\phi}(k) - \phi(k)\right)$.

Our main result is as follows:

Theorem 3.1. *Under the assumptions of Theorem 2.2, if in addition k satisfies $k \to \infty$ as $n \to \infty$ such that $k^4/n \to 0$, then*

$$\ell(k)' \sqrt{n-k}\left(\hat{\phi}^*_{fwb}(k) - \hat{\phi}(k)\right)/v_k \Rightarrow N(0,1),$$

under P^* with probability approaching one, and for any $\delta > 0$

$$P\left\{\sup_{x \in \mathbf{R}} \left| P^*\left[\ell(k)' \sqrt{n-k}\left(\hat{\phi}^*_{fwb}(k) - \hat{\phi}(k)\right) \le x\right] - P\left[\ell(k)' \sqrt{n-k}\left(\hat{\phi}(k) - \phi(k)\right) \le x\right]\right| > \delta\right\} \to 0,$$

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

Remark 1. *Under the conditions of Theorem 3.1, the (conditional) asymptotic distribution of the fixed-design WB OLS estimators is identical to the asymptotic distribution of the OLS estimators evaluated on the original data. Thus, Theorem 3.1 establishes the first-order asymptotic validity of the fixed-design WB for the autoregressive parameters of AR(∞) processes subject to possible conditional heteroskedasticity of unknown form.*

Remark 2. *The assumptions underlying the bootstrap approximation in Theorem 3.1 are the same as those needed to apply the asymptotic normal approximation based on Theorem 2.3. Note that the bootstrap population variance, $v_k^{*2} \equiv \ell(k)' \Gamma_k^{-1} \hat{B}_k \Gamma_k^{-1} \ell(k)'$, depends on the same heteroskedasticity-robust covariance matrix estimator of B_k as the estimator $\hat{v}_k^2$ in Theorem 2.3. In both cases, the same upper bound on the rate of increase of k is needed to ensure consistency for v_k^2.*

3.2. Pairwise Bootstrap for Sieve Autoregressions

The pairwise bootstrap consists of the following steps:

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

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

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

Accordingly, let $\hat{\phi}^*_{pb}(k) = \hat{\Gamma}^{*-1}_{pb,k}\hat{\Gamma}^*_{pb,k,1}$, where $\hat{\Gamma}^*_{pb,k} = (n-k)^{-1}\sum_{t=1+k}^{n} Y^*_{t-1,k}Y^{*\prime}_{t-1,k}$ and $\hat{\Gamma}^*_{pb,k,1} = (n-k)^{-1}\sum_{t=1+k}^{n} Y^*_{t-1,k}y^*_t$, and define $\varepsilon^*_{t,k} = y^*_t - Y^{*\prime}_{t-1,k}\phi(k)$, $\hat{\varepsilon}^*_{t,k} = y^*_t - Y^{*\prime}_{t-1,k}\hat{\phi}(k)$ and $\hat{\varepsilon}_{t,k} = y_t - Y'_{t-1,k}\hat{\phi}(k)$. Our main result is as follows:

Theorem 3.2. *Under the assumptions of Theorem 2.2, if in addition k satisfies $k \to \infty$ as $n \to \infty$ such that $k^4/n \to 0$, then*

$$\ell(k)'\sqrt{n-k}\left(\hat{\phi}^*_{pb}(k) - \hat{\phi}(k)\right)/v_k \Rightarrow N(0,1),$$

under P^* with probability approaching one, and for any $\delta > 0$

$$P\left\{\sup_{x\in\mathbf{R}}\left|P^*\left[\ell(k)'\sqrt{n-k}\left(\hat{\phi}^*_{pb}(k) - \hat{\phi}(k)\right) \le x\right] - P\left[\ell(k)'\sqrt{n-k}\left(\hat{\phi}(k) - \phi(k)\right) \le x\right]\right| > \delta\right\} \to 0,$$

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

It is useful to differentiate the pairwise bootstrap from the blocks-of-blocks (BOB) bootstrap, as discussed in Gonçalves and White (2002). 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.

4. Extensions to Smooth Functions of Autoregressive Parameters

Typically, in applied work interest focuses not on the autoregressive coefficients themselves, but on smooth functions of autoregressive parameters. A leading example are coefficients of impulse responses, which can be written as nonlinear functions of the autoregressive coefficients. Other examples include the half-life of deviations from the unconditional mean or cumulative impulse response coefficients. In this section we establish the validity of the two bootstrap methods analyzed in the previous section for smooth functions of $\phi(k)$. Extensions to other functions of interest in applied work are discussed in Inoue and Kilian (2002). Specifically, we let $g(\phi(k))$ denote the parameter of interest and $g\left(\hat{\phi}(k)\right)$ its estimator. In addition, we postulate that g is a smooth function from $\mathbb{R}^k$ to $\mathbb{R}$ satisfying Assumption 3 below, where $\nabla g(x) = \partial g/\partial x$ for any $x \in \mathbb{R}^k$.

Assumption 3. (i) $0 < M_1 \leq \|\nabla g(\phi(k))\|^2 \leq M_2 < \infty$; (ii) ∇g satisfies a Lipschitz condition, i.e. there exists $M < \infty$ such that $\|\nabla g(x) - \nabla g(y)\| \leq M \|x - y\|$ for all $x, y \in \mathbb{R}^k$.

With the additional Assumption 3, the following corollary to Theorems 3.1 and 3.2 is true. To conserve space, we let $\hat{\phi}^*(k)$ denote the OLS estimator obtained with either of the two bootstrap schemes studied in the previous section.

Corollary 4.1. *Let $\{y_t\}$ satisfy (2.1) and assume that Assumptions 1 and 2 hold, strengthened by Assumptions 1 (iii′) and 2 (i′). If in addition Assumption 3 holds and k satisfies $k \to \infty$ as $n \to \infty$ such that $k^4/n \to 0$, then for any $\delta > 0$*

$$P\left\{\sup_{x\in\mathbf{R}}\left|P^*\left[\sqrt{n-k}\left(g\left(\hat{\phi}^*(k)\right) - g\left(\hat{\phi}(k)\right)\right) \leq x\right] - P\left[\sqrt{n-k}\left(g\left(\hat{\phi}(k)\right) - g(\phi(k))\right) \leq x\right]\right| > \delta\right\} \to 0,$$

where $\hat{\phi}^*(k)$ denotes either the fixed-design WB or the pairwise bootstrap OLS estimator.

5. Discussion

A large number of papers in the time series literature has relied on the work of Berk (1974), Bhansali (1978) and Lewis and Reinsel (1985) in establishing theoretical results based on autoregressive sieve approximations. Our analysis is likely to be useful in extending these results to the case of models with martingale difference errors subject to possible conditional heteroskedasticity of unknown form. Of particular relevance are problems of inference both in the univariate and in the multivariate context.

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. All these studies are based on asymptotic distributions derived under the i.i.d. error assumption that is invalidated by the possible presence of conditional heteroskedasticity. Our asymptotic results provide the basis for developing robust methods of inference for these problems. Alternatively, appropriate inference may be conducted based on the robust bootstrap methods discussed in this paper. Using asymptotic approximations based on the delta method requires the evaluation of the gradient of the function of interest, which can be cumbersome. An attractive feature of the bootstrap approach in this context is that it dispenses with the need to derive closed-form formulae for the standard errors of each statistic of interest. 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.

Although we focused on stationary linear processes our results will also be useful in the context of 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).

A. Appendix

Throughout this Appendix, the scalar C denotes a generic constant independent of n. For an $m \times 1$ vector a, let $\|a\|$ denote the Euclidean norm i.e. $\|a\|^2 = a'a$. Given an $m \times n$ matrix A, let $\|A\|$ denote the Schur's matrix norm defined by $\|A\|^2 = tr\left(A'A\right)$, and let $\|A\|_1 = \sup_{x \neq 0}\{\|Ax\| / \|x\|\}$ denote the subordinated matrix norm associated with the Euclidean norm (the so-called spectral norm). Then, $\|A\|_1 = \max\{\sqrt{\lambda} : \lambda \text{ is an eigenvalue of } A'A\}$, i.e. $\|A\|_1^2$ is the largest eigenvalue of $A'A$ (if A is symmetric, then $\|A\|_1^2$ is the square of the largest, in absolute value, eigenvalue of A). The following inequalities relating $\|\cdot\|$ and $\|\cdot\|_1$ are used below:

$$\|A\|_1^2 \leq \|A\|^2 , \text{ and} \tag{A.1}$$

$$\|AB\|^2 \leq \|A\|_1^2 \|B\|^2 \quad \text{and} \quad \|AB\|^2 \leq \|B\|_1^2 \|A\|^2 , \tag{A.2}$$

for any two compatible matrices A and B (cf. Horn and Johnson (1985), p. 314 for (A.2) and p. 313 for (A.1)).

For any bootstrap statistic T_n^* we write $T_n^* \xrightarrow{P^*} 0$ in probability when $\lim_{n\to\infty} P\left[P^*\left(|T_n^*| > \delta\right) > \delta\right] = 0$ for any $\delta > 0$, i.e. $P^*\left(|T_n^*| > \delta\right) = o_P(1)$. We write $T_n^* = O_{P^*}\left(n^{\lambda}\right)$ in probability when for all $\delta > 0$

there exists a $M_\delta < \infty$ such that $\lim_{n\to\infty} P\left[P^*\left(\left|n^{-\lambda}T_n^*\right| > M_\delta\right) > \delta\right] = 0$. 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.

Lemma A.1 below proves the absolute summability of the fourth order joint cumulants of $(y_t, y_{t+l_1}, y_{t+l_2}, y_{t+l_3})$, given the absolute summability of the fourth order joint cumulants of $(\varepsilon_t, \varepsilon_{t+l_1}, \varepsilon_{t+l_2}, \varepsilon_{t+l_3})$. We will let $Cum(\cdot, \ldots, \cdot)$ denote the joint cumulant of the set of random variables involved. Lemma A.1 follows from Theorem 2.8.1 of Brillinger (1981, p. 19) under our Assumption 1. Lemma A.2 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. Its proof uses Lemma A.1.

Lemma A.1. *Let $\{y_t\}$ be generated from (2.1) and assume that Assumption 1 holds. Then*

$$\sum_{l_1=-\infty}^{\infty} \sum_{l_2=-\infty}^{\infty} \sum_{l_2=-\infty}^{\infty} \left|\kappa_y(0, l_1, l_2, l_3)\right| < \infty,$$

where $\kappa_y(0, l_1, l_2, l_3)$ is the fourth order joint cumulant of $(y_t, y_{t+l_1}, y_{t+l_2}, y_{t+l_3})$.

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

$$\left\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 = o_P(1). \tag{A.3}$$

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

$$k^{1/2}\left\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 = o_P(1). \tag{A.4}$$

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

Lemma A.3. *Let $\{y_t\}$ satisfy (2.1) and assume that Assumptions 1 and Assumption 2 (i′) and (ii) hold. Let $\ell(k)$ be an arbitrary sequence of $k \times 1$ vectors satisfying $0 < M_1 \leq \|\ell(k)\|^2 \leq M_2 < \infty$. Then*

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

The following results are the bootstrap (fixed-design WB and pairwise bootstrap) analogues of Lemma A.3.

Lemma A.4. *Under the assumptions of Lemma A.3,*

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

where $r^*_{fwb}=o_{P^*}(1)$ *in probability, i.e.* $P^*\left(\left|r^*_{fwb}\right|>\delta\right)=o_P(1)$ *for any* $\delta>0$.

Lemma A.5. *Under the assumptions of Lemma A.3,*

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

where $r^*_{pb}=o_{P^*}(1)$ *in probability, i.e.* $P^*\left(\left|r^*_{pb}\right|>\delta\right)=o_P(1)$ *for any* $\delta>0$.

Proof of Theorem 2.1. 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 write

$$\hat{\phi}(k)-\phi(k)=\hat{\Gamma}_k^{-1}\left(\hat{\Gamma}_{k,1}-\hat{\Gamma}_k\phi(k)\right)$$
$$=\hat{\Gamma}_k^{-1}(n-k)^{-1}\sum_{t=k+1}^{n}Y_{t-1,k}\left(y_t-Y'_{t-1,k}\phi(k)\right)=\hat{\Gamma}_k^{-1}(n-k)^{-1}\sum_{t=k+1}^{n}Y_{t-1,k}\varepsilon_{t,k}$$
$$=-\hat{\Gamma}_k^{-1}(n-k)^{-1}\sum_{t=k+1}^{n}Y_{t-1,k}\left(\varepsilon_t-\varepsilon_{t,k}\right)+\hat{\Gamma}_k^{-1}(n-k)^{-1}\sum_{t=k+1}^{n}Y_{t-1,k}\varepsilon_t.$$

Using (A.1), we have

$$\left\|\hat{\phi}(k)-\phi(k)\right\|\leq\left\|\hat{\Gamma}_k^{-1}\right\|_1\left\|U_{1n}\right\|+\left\|\hat{\Gamma}_k^{-1}\right\|_1\left\|U_{2n}\right\|,\tag{A.5}$$

where

$$U_{1n}=(n-k)^{-1}\sum_{t=k+1}^{n}Y_{t-1,k}\left(\varepsilon_t-\varepsilon_{t,k}\right),\tag{A.6}$$

and

$$U_{2n}=(n-k)^{-1}\sum_{t=k+1}^{n}Y_{t-1,k}\varepsilon_t.\tag{A.7}$$

Next, we prove: (a) $\left\|\hat{\Gamma}_k^{-1}\right\|_1=O_P(1)$, (b) $\left\|U_{1n}\right\|=O_P\left(\frac{k^{1/2}}{n^{1/2}}\right)$, and (c) $\left\|U_{2n}\right\|=O_P\left(\frac{k^{1/2}}{n^{1/2}}\right)$. For (a), note that $\left\|\hat{\Gamma}_k^{-1}\right\|_1\leq\left\|\Gamma_k^{-1}\right\|_1+\left\|\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right\|_1$. As in Berk (1974, cf. equations (2.14) and (2.18)), under our conditions the spectral density of y_t is bounded above and bounded away from zero, implying that $\|\Gamma_k\|_1$ and $\|\Gamma_k^{-1}\|_1$ are bounded uniformly in k. From Lemma A.2 we have that $\left\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 = o_P(1)$ provided $k^2/n \to 0$, which proves (a). Next, we prove (b). By the triangle inequality for vector norms first, and the Cauchy-Schwartz inequality second, we get

$$E(\|U_{1n}\|) \le (n-k)^{-1} \sum_{t=k+1}^{n} E\|Y_{t-1,k}(\varepsilon_t - \varepsilon_{t,k})\| = (n-k)^{-1} \sum_{t=k+1}^{n} \left(E\|Y_{t-1,k}\|^2\right)^{1/2} \left(E|\varepsilon_t - \varepsilon_{t,k}|^2\right)^{1/2}.$$

Now, $E\|Y_{t-1,k}\|^2 = E\left(\sum_{j=1}^{k} y_{t-j}^2\right) = \sum_{j=1}^{k} E\left(y_{t-j}^2\right) \le Ck$, given that $E\left(y_{t-j}^2\right) = \sigma^2 \sum_{i=1}^{\infty} \psi_i^2 \le C < \infty$ uniformly in j under our conditions. Observing that $\varepsilon_t - \varepsilon_{t,k} = -\sum_{j=k+1}^{\infty} \phi_j y_{t-j}$, by Minkowski's inequality,

$$\left(E|\varepsilon_t - \varepsilon_{t,k}|^2\right)^{1/2} = \left(E\left|-\sum_{j=k+1}^{\infty} \phi_j y_{t-j}\right|^2\right)^{1/2} \le \sum_{j=k+1}^{\infty} |\phi_j| \left(E|y_{t-j}^2|\right)^{1/2} \le C \sum_{j=k+1}^{\infty} |\phi_j|, \qquad \text{(A.8)}$$

implying that $E(\|U_{1n}\|) \le Ck^{1/2} \sum_{j=k+1}^{\infty} |\phi_j|$. Thus, by the Markov inequality, for any $\delta > 0$,

$$P\left(\frac{n^{1/2}}{k^{1/2}} \|U_{1n}\| > \delta\right) \le \frac{1}{\delta} \frac{n^{1/2}}{k^{1/2}} E(\|U_{1n}\|) \le \frac{C}{\delta} n^{1/2} \sum_{j=k+1}^{\infty} |\phi_j| \to 0,$$

by Assumption 2 (ii). Thus, $\|U_{1n}\| = o_P\left(\frac{k^{1/2}}{n^{1/2}}\right)$ and therefore $O_P\left(\frac{k^{1/2}}{n^{1/2}}\right)$. Lastly, we show (c). We can write

$$E\left(\|U_{2n}\|^2\right) = (n-k)^{-2} \sum_{j=1}^{k} \sum_{t=1+k}^{n} \sum_{s=1+k}^{n} E(y_{t-j} y_{s-j} \varepsilon_t \varepsilon_s) = (n-k)^{-2} \sum_{j=1}^{k} \sum_{t=1+k}^{n} E\left(y_{t-j}^2 \varepsilon_t^2\right), \qquad \text{(A.9)}$$

since $E(y_{t-j} y_{s-j} \varepsilon_t \varepsilon_s) = 0$ for $t \ne s$. Also, using the MA(∞) representation of y_t,

$$\begin{aligned} E\left(y_{t-j}^2 \varepsilon_t^2\right) &= E\left(\left(\sum_{l=0}^{\infty} \psi_l \varepsilon_{t-j-l}\right)^2 \varepsilon_t^2\right) = \sum_{l_1=0}^{\infty} \sum_{l_2=0}^{\infty} \psi_{l_1} \psi_{l_2} E\left(\varepsilon_{t-j-l_1} \varepsilon_{t-j-l_2} \varepsilon_t^2\right) \\ &= \sum_{l_1=0}^{\infty} \sum_{l_2=0}^{\infty} \psi_{l_1} \psi_{l_2} \alpha_{l_1+j, l_2+j} \le C \left(\sum_{l=0}^{\infty} \psi_l\right)^2 < \infty, \end{aligned}$$

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

Proof of Theorem 2.2. First, note that $v_k^2 = \ell'(k) \Gamma_k^{-1} B_k \Gamma_k^{-1} \ell(k)$ is bounded above and bounded away from zero, given our assumptions. In particular, $v_k^2 > 0$ holds by assumption. The fact that

$v_k^2 < \infty$ follows from our assumptions. It suffices to show that (i) $\|\ell(k)\| \leq M_2 < \infty$; (ii) $\left\|\Gamma_k^{-1}\right\|_1 \leq C_2 < \infty$, and (iii) $\|B_k\|_1 \leq D_2 < \infty$. (i) holds by assumption and (ii) follows by Brockwell and Davis' (1991) Proposition 4.5.3. under our assumptions. To obtain (iii), write $Y_{t-1,k} = \sum_{j=1}^{\infty} b_{j,k}\varepsilon_{t-j}$, where $b_{j,k} = \left(\psi_{j-1}, \ldots, \psi_{j-k}\right)'$ with $\psi_j = 0$ for $j < 0$ and $\psi_0 = 1$. It follows that

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

where $\alpha_{i,j} = \kappa_\varepsilon(0, -i, -j, 0)$ for $i \neq j$ and $\alpha_{i,j} = \sigma^4 + \kappa_\varepsilon(0, -i, -j, 0)$ for $i = j$. Using these expressions for $\alpha_{i,j}$ and the fact that we can write $\Gamma_k = \sigma^2 \sum_{j=1}^{\infty} b_{j,k}b'_{j,k}$, we obtain

$$\|B_k\|_1 \leq \sigma^2 \|\Gamma_k\|_1 + \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \left\|b_{j,k}b'_{i,k}\right\|_1 \left|\kappa_\varepsilon(0, -i, -j, 0)\right|.$$

Given (ii), it suffices to show that the second term is bounded to establish that $\|B_k\|_1$ is bounded. Note that $\left\|b_{j,k}b'_{i,k}\right\|_1 \leq \left\|b_{j,k}b'_{i,k}\right\| = \left(tr\left(b_{i,k}b'_{j,k}b_{j,k}b'_{i,k}\right)\right)^{1/2} \leq \sum_{l=0}^{\infty}\psi_l^2 < \infty$. This together with Assumption 1 (iii) implies the desired result.

Thus, to show $\ell(k)'\sqrt{n-k}v_k^{-1}\left(\hat{\phi}(k) - \phi(k)\right) \Rightarrow N(0,1)$ it suffices that $(n-k)^{-1/2}\sum_{t=1+k}^{n}\ell(k)'\Gamma_k^{-1}Y_{t-1,k}\varepsilon_t/v_k \Rightarrow N(0,1)$, given Lemma A.3 and the Asymptotic Equivalence Lemma (cf. White 2000, Lemma 4.7). 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 $E(z_{nt}|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = \ell(k)'\Gamma_k^{-1}Y_{t-1,k}v_k^{-1}E(\varepsilon_t|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = 0$ under Assumption 1. Hence we need to show that (a) $(n-k)^{-1}\sum_{t=1+k}^{n} z_{nt}^2 - 1 \xrightarrow{P} 0$, and (b) $\max_{1+k \leq t \leq n}(n-k)^{-1/2}|z_{nt}| \xrightarrow{P} 0$. We start with (a). Note that $Var(z_{nt}) = \ell(k)'\Gamma_k^{-1}B_k\Gamma_k^{-1}\ell(k)/v_k^2 = 1$, with $B_k = E\left(Y_{t-1,k}Y'_{t-1,k}\varepsilon_t^2\right)$. We can write

$$\begin{aligned}\frac{1}{n-k}\sum_{t=1+k}^{n} z_{nt}^2 - 1 &= v_k^{-2}\left\{\ell(k)'\Gamma_k^{-1}\left(\frac{1}{n-k}\sum_{t=1+k}^{n} Y_{t-1,k}Y'_{t-1,k}\varepsilon_t^2 - E\left(Y_{t-1,k}Y'_{t-1,k}\varepsilon_t^2\right)\right)\Gamma_k^{-1}\ell(k)\right\} \\ &\equiv v_k^{-2}\left\{\ell(k)'\Gamma_k^{-1}S_k\Gamma_k^{-1}\ell(k)\right\},\end{aligned}$$

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

$$\left|\ell(k)'\Gamma_k^{-1}S_k\Gamma_k^{-1}\ell(k)\right| \leq \|\ell(k)\|\left\|\Gamma_k^{-1}\right\|_1\|S_k\|_1\left\|\Gamma_k^{-1}\right\|_1\|\ell(k)\| \leq M_2C_2^2\|S_k\|_1,$$

since $\|\ell(k)\|^2 \leq M_2$ and $\left\|\Gamma_k^{-1}\right\|_1 \leq C_2$ 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, by (A.1), that $E\|S_k\|^2 \to 0$. We have

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

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

$$(n-k)\,E\left([S_k]_{i,j}^2\right) = (n-k)\,E\left[\left((n-k)^{-1}\sum_{t=1+k}^{n}\left(y_{t-i}y_{t-j}\varepsilon_t^2 - E\left(y_{t-i}y_{t-j}\varepsilon_t^2\right)\right)\right)^2\right] \tag{A.10}$$

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 $\eta > 0$ and for some $r > 1$,

$$P\left(\max_{1+k\leq t\leq n}|z_{nt}| > \sqrt{n-k}\,\eta\right) \leq \sum_{t=1+k}^{n} P\left(|z_{nt}| > \sqrt{n-k}\,\eta\right) \leq \sum_{t=1+k}^{n}\frac{E|z_{nt}|^r}{(n-k)^{r/2}\,\eta^r}. \tag{A.11}$$

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

$$E|z_{nt}|^r = E|v_{t,k}\varepsilon_t/v_k|^r \leq \left|v_k^{-1}\right|^r\left(E|v_{t,k}|^{2r}\right)^{1/2}\left(E|\varepsilon_t|^{2r}\right)^{1/2} \leq C\left(E|v_{t,k}|^{2r}\right)^{1/2}, \tag{A.12}$$

given that $v_k^{-1} = O(1)$ and $E|\varepsilon_t|^{2r} = O(1)$, for $r \leq 4$ under our assumptions. We now prove that $E|v_{t,k}|^{2r} = O(k^r)$. Given the definition of $v_{t,k}$ and the norm properties of vectors and matrices, $|v_{t,k}| = \left|\ell(k)'\,\Gamma_k^{-1}Y_{t-1,k}\right| \leq \|\ell(k)\|\left\|\Gamma_k^{-1}\right\|_1\|Y_{t-1,k}\|$, so that

$$|v_{t,k}|^{2r} = \left|\ell(k)'\,\Gamma_k^{-1}Y_{t-1,k}\right|^{2r} \leq \|\ell(k)\|^{2r}\left\|\Gamma_k^{-1}\right\|_1^{2r}\|Y_{t-1,k}\|^{2r} \leq M_2^rC_2^r\|Y_{t-1,k}\|^{2r} = M_2^rC_2^r\left|\sum_{j=1}^{k}y_{t-j}^2\right|^r.$$

Thus,

$$\begin{aligned} E|v_{t,k}|^{2r} &\leq M_2^rC_2^rE\left|\sum_{j=1}^{k}y_{t-j}^2\right|^r = M_2^rC_2^r\left[\left(E\left|\sum_{j=1}^{k}y_{t-j}^2\right|^r\right)^{1/r}\right]^r \\ &\leq M_2^rC_2^r\left(\sum_{j=1}^{k}\left(E\left|y_{t-j}^2\right|^r\right)^{1/r}\right)^r \leq M_2^rC_2^r\left(\sum_{j=1}^{k}\Delta^{1/r}\right)^r \leq M_2^rC_2^r\Delta k^r, \end{aligned}$$

where the second inequality holds by the Minkowski inequality and the third inequality holds by the fact that $E|y_{t-j}|^{2r} \leq \Delta < \infty$ for all $j = 1,\ldots,k$ and for some $r \leq 4$. It follows from (A.12) that

$E\left|z_{nt}\right|^r \leq C\left(M_2^r C_2^r \Delta k^r\right)^{1/2} \leq Ck^{r/2}$, for some finite constant C, which implies from (A.11) that

$$P\left(\max_{1+k\leq t\leq n}\left|z_{nt}\right| > \sqrt{n-k}\eta\right) \leq \sum_{t=1+k}^{n}\frac{Ck^{r/2}}{(n-k)^{r/2}\eta^r} = O\left(\frac{k^{r/2}}{(n-k)^{r/2-1}}\right). \tag{A.13}$$

Letting $r=3$ implies that (A.13) is $O\left(\frac{k^{3/2}}{(n-k)^{1/2}}\right)$, which is $o(1)$ provided $k^3/n \to 0$, as we assume.

To conclude the proof, we show that (A.10) 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.10) is equal to

$$\begin{aligned}
&(n-k)^{-1}\sum_{t=1+k}^{n}\sum_{s=1+k}^{n} Cov\left(y_{t-i}y_{t-j}\varepsilon_t^2, y_{s-i}y_{s-j}\varepsilon_s^2\right)\\
=&\sum_{l_1,\ldots,l_4=-\infty}^{\infty}\psi_{l_1}\psi_{l_2}\psi_{l_3}\psi_{l_4}(n-k)^{-1}\sum_{t=1+k}^{n}\sum_{s=1+k}^{n} Cov\left(\varepsilon_{t-i-l_1}\varepsilon_{t-j-l_2}\varepsilon_t^2, \varepsilon_{s-i-l_3}\varepsilon_{s-j-l_4}\varepsilon_s^2\right).
\end{aligned}$$

Next, we show that

$$(n-k)^{-1}\sum_{t=1+k}^{n}\sum_{s=1+k}^{n} Cov\left(\varepsilon_{t-i-l_1}\varepsilon_{t-j-l_2}\varepsilon_t^2, \varepsilon_{s-i-l_3}\varepsilon_{s-j-l_4}\varepsilon_s^2\right) \leq C, \tag{A.14}$$

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 $Cov\left(\varepsilon_{t-i-l_1}\varepsilon_{t-j-l_2}\varepsilon_t^2, \varepsilon_{s-i-l_3}\varepsilon_{s-j-l_4}\varepsilon_s^2\right)$ as the sum of products of cumulants of ε_t of order eight and lower. 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

$$Cum\left(Y_1, Y_2\right) = \sum_v Cum\left(X_{ij} : i, j \in v_1\right)\ldots Cum\left(X_{ij} : i, j \in v_p\right) \tag{A.15}$$

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

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

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.15).

Consider $p = 1$, i.e. consider $v = \left\{\varepsilon_{t-i-l_1}, \varepsilon_{t-j-l_2}, \varepsilon_t, \varepsilon_t, \varepsilon_{s-i-l_3}, \varepsilon_{s-j-l_4}, \varepsilon_s, \varepsilon_s\right\}$. This term contributes towards the sum with the 8th order joint cumulant $Cum\left(\varepsilon_{t-i-l_1}, \varepsilon_{t-j-l_2}, \varepsilon_t, \varepsilon_t, \varepsilon_{s-i-l_3}, \varepsilon_{s-j-l_4}, \varepsilon_s, \varepsilon_s\right)$, which by stationarity can be written as $\kappa_\varepsilon\left(t-s-i-l_1, t-s-j-l_2, t-s, t-s, -i-l_3, -j-l_4, 0, 0\right) \equiv \kappa_\varepsilon\left(\tau-i-l_1, \tau-j-l_2, \tau, \tau, -i-l_3, -j-l_4, 0, 0\right)$, if we set $\tau = t-s$. Thus, by a change of variables the contribution of this term to (A.14) is

$$\begin{aligned}
&\sum_{\tau=-n+(1+k)}^{n-(1+k)} \left(1-\frac{|\tau|}{n-k}\right)\kappa_{\varepsilon}\left(\tau-i-l_1,\tau-j-l_2,\tau,\tau,-i-l_3,-j-l_4,0,0\right) \\
\leq\ &\sum_{\tau=-\infty}^{\infty} \left|\kappa_{\varepsilon}\left(\tau-i-l_1,\tau-j-l_2,\tau,\tau,-i-l_3,-j-l_4,0,0\right)\right| \\
\leq\ &\sum_{\tau_1=-\infty}^{\infty}\sum_{\tau_2=-\infty}^{\infty}\cdots\sum_{\tau_7=-\infty}^{\infty} \left|\kappa_{\varepsilon}\left(\tau_1,\tau_2,\ldots,\tau_7,0\right)\right| < \infty,
\end{aligned}$$

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.15) with a non-zero 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.14). 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 2.3. Adding and subtracting appropriately, we can write

$$\hat{v}_k^2-v_k^2=\ell(k)'\hat{\Gamma}_k^{-1}\left(\hat{B}_k-B_k\right)\Gamma_k^{-1}\ell(k)+\ell(k)'\left(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right)B_k\Gamma_k^{-1}\ell(k)+\ell(k)'\hat{\Gamma}_k^{-1}\hat{B}_k\left(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right)\ell(k),$$

which is bounded by $\left|\hat{v}_k^2-v_k^2\right|\leq C\left(\left\|\hat{B}_k-B_k\right\|+\left\|\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right\|_1\right)$, for some constant C, given that $\|\ell(k)\|$, $\left\|\Gamma_k^{-1}\right\|_1$, and $\|B_k\|_1$ are bounded, and that $\left\|\hat{\Gamma}_k^{-1}\right\|_1$ and $\left\|\hat{B}_k\right\|_1$ are bounded in probability. Under our assumptions, $\left\|\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right\|=o_P(1)$, see Lemma A.2. Next, we show that $\left\|\hat{B}_k-B_k\right\|=o_P(1)$, which proves the result. We can write $\left\|\hat{B}_k-B_k\right\|\leq A_1+A_2+A_3$, where

$$A_1=\left\|(n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}Y_{t-1,k}'\left(\hat{\varepsilon}_{t,k}^2-\varepsilon_{t,k}^2\right)\right\|,\ A_2=\left\|(n-k)^{-1}\sum_{t=1+k}^{n}Y_{t-1,k}Y_{t-1,k}'\left(\varepsilon_{t,k}^2-\varepsilon_t^2\right)\right\|,$$

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

We can show that $A_3=O_P\left(\frac{k}{(n-k)^{1/2}}\right)$ under our conditions (see proof of Theorem 2.2. A_3 here corresponds to S_k there). Next, we will show that $A_1 = O_P\left(\left(\frac{k^4}{n}\right)^{1/2}\right)$, which is $o_P(1)$ if $\frac{k^4}{n} \to 0$, and $A_2 = O_P\left(k \sum_{j=1+k}^{\infty} |\phi_j|\right)$, 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

$$A_1 = \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y'_{t-1,k} \left(\hat{\varepsilon}_{t,k} - \varepsilon_{t,k}\right)\left(\hat{\varepsilon}_{t,k} + \varepsilon_{t,k}\right) \right\| \leq \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y'_{t-1,k} \hat{\varepsilon}_{t,k} \left(\hat{\varepsilon}_{t,k} - \varepsilon_{t,k}\right) \right\|$$
$$+ \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y'_{t-1,k} \varepsilon_{t,k} \left(\hat{\varepsilon}_{t,k} - \varepsilon_{t,k}\right) \right\| \equiv A_{11} + A_{12}.$$

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

$$\begin{aligned} A_{11} &= \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y'_{t-1,k} Y'_{t-1,k} \hat{\varepsilon}_{t,k} \left(\hat{\phi}(k) - \phi(k)\right) \right\| \\ &\leq \left\| \hat{\phi}(k) - \phi(k) \right\| (n-k)^{-1} \sum_{t=1+k}^{n} \left\| Y_{t-1,k} \right\|^3 \left| \hat{\varepsilon}_{t,k} \right| \\ &\leq \left\| \hat{\phi}(k) - \phi(k) \right\| \left((n-k)^{-1} \sum_{t=1+k}^{n} \left\| Y_{t-1,k} \right\|^6 \right)^{1/2} \left((n-k)^{-1} \sum_{t=1+k}^{n} \left| \hat{\varepsilon}_{t,k} \right|^2 \right)^{1/2} \\ &= O_P\left(\frac{k^{1/2}}{n^{1/2}}\right) O_P\left(k^{3/2}\right) O_P(1) = O_P\left(\frac{k^{4/2}}{n^{1/2}}\right), \end{aligned}$$

where the second equality holds by Theorem 2.1, the fact that $E\left\|Y_{t-1,k}\right\|^6 = O\left(k^3\right)$ (since $E\left|y_t\right|^8 \leq \Delta < \infty$) and $(n-k)^{-1} \sum_{t=1+k}^{n} \left|\hat{\varepsilon}_{t,k}\right|^2 = O_P(1)$. 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 $\varepsilon_{t,k}^2 - \varepsilon_t^2 = \left(\varepsilon_{t,k} - \varepsilon_t\right)\left(\varepsilon_{t,k} + \varepsilon_t\right)$, 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

$$\begin{aligned} A_{22} &\leq (n-k)^{-1} \sum_{t=1+k}^{n} \left\| Y_{t-1,k} \right\|^2 \left|\varepsilon_t\right| \sum_{j=1+k}^{\infty} \left|\phi_j\right| \left|y_{t-j}\right| \leq \sum_{j=1+k}^{\infty} \left|\phi_j\right| (n-k)^{-1} \sum_{t=1+k}^{n} \left\| Y_{t-1,k} \right\|^2 \left|\varepsilon_t y_{t-j}\right| \\ &\leq \sum_{j=1+k}^{\infty} \left|\phi_j\right| \left((n-k)^{-1} \sum_{t=1+k}^{n} \left\| Y_{t-1,k} \right\|^4 \right)^{1/2} \left((n-k)^{-1} \sum_{t=1+k}^{n} \left|\varepsilon_t y_{t-j}\right|^2 \right)^{1/2} = O_P\left(k \sum_{j=1+k}^{\infty} \left|\phi_j\right| \right), \end{aligned}$$

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 Lemma A.4, it suffices to show that $(n-k)^{-1/2}\sum_{t=1+k}^{n} w_{nt}^* \Rightarrow^{d_{P^*}} N(0,1)$, in probability, where $w_{nt}^* = \ell(k)' \Gamma_k^{-1} v_k^{-1} Y_{t-1,k}\hat{\varepsilon}_{t,k}^*$. 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). Note $E^*\left((n-k)^{-1/2}\sum_{t=1+k}^{n} w_{nt}^*\right) = 0$ 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 $v_k^{*2} \equiv \ell(k)'\Gamma_k^{-1}\hat{B}_k\Gamma_k^{-1}\ell(k)$, $\hat{B}_k = (n-k)^{-1}\sum_{t=1+k}^{n} Y_{t-1,k}Y_{t-1,k}'\hat{\varepsilon}_{t,k}^2$. The proof will consist of the following 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. By the triangle inequality, $\left|v_k^{*2} - v_k^2\right| \leq \|\ell(k)\|^2 \left\|\Gamma_k^{-1}\right\|_1^2 \left\|\hat{B}_k - B_k\right\| \leq C\left\|\hat{B}_k - B_k\right\|$, for some constant C, given that $\|\ell(k)\|$ and $\left\|\Gamma_k^{-1}\right\|_1$ are bounded. By Theorem 2.3, $\left\|\hat{B}_k - B_k\right\| = o_P(1)$, which proves step 1.

Proof of Step 2. We will show Lyapunov's condition 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

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

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

By an argument similar to that used in Lemma A.4, we can show that $(n-k)^{-1}\sum_{t=1+k}^{n}|\hat{\varepsilon}_{t,k}|^{4r} \leq (n-k)^{-1}\sum_{t=1+k}^{n}|\varepsilon_t|^{4r} + O_P\left(\left(\sum_{j=1+k}^{\infty}|\phi_j|\right)^{4r}\right) + O_P\left(\frac{k^{4r}}{n^{4r/2}}\right)$, provided $E|y_t|^{4r} \leq \Delta < \infty$ for all t. Thus, with $r = \frac{3}{2}$, under our conditions 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 $r = \frac{3}{2}$. Thus,

$$(n-k)^{-r}\sum_{t=1+k}^{n} E^*|w_{nt}^*|^{2r} = O_P\left(\frac{k^r}{(n-k)^{r-1}}\right) = O_P\left(\left(\frac{k^3}{n-k}\right)^{1/2}\right) = o_P(1),$$

if $k^3/n \to 0$.

Proof of Theorem 3.2. Given Lemma A.5, it suffices to show that $(n-k)^{-1/2}\sum_{t=1+k}^{n} w_{nt}^* \Rightarrow^{d_{P^*}} N(0,1)$ in probability, where $w_{nt}^* = \ell(k)' v_k^{-1}\Gamma_k^{-1}Y_{t-1,k}^*\hat{\varepsilon}_{t,k}^*$. Note that, conditional on the data, w_{nt}^* is independent with $E^*(w_{nt}^*) = 0$ and $Var^*\left((n-k)^{-1/2}\sum_{t=1+k}^{n} w_{nt}^*\right) = \frac{v_k^{*2}}{v_k^2}$, where $v_k^{*2} = \ell(k)' \Gamma_k^{-1} \hat{B}_k \Gamma_k^{-1} \ell(k)$, as in the proof of Theorem 3.1. Thus, $\frac{v_k^{*2}}{v_k^2} \to 1$ in probability and we only need to check Lyapunov's condition (cf. Step 2 in the proof of Theorem 3.1). Using the properties of the pairwise bootstrap yields, for some $r > 1$,

$$E^*\left(|w_{nt}^*|^{2r}\right) \leq \|\ell(k)\|^{2r} \left\|\Gamma_k^{-1}\right\|_1^{2r} \left|v_k^{-1}\right|^{2r} E^*\left(\left\|Y_{t-1,k}^* \hat{\varepsilon}_{t,k}^*\right\|^{2r}\right) \leq C(n-k)^{-1} \sum_{t=1+k}^{n} \|Y_{t-1,k}\hat{\varepsilon}_{t,k}\|^{2r}$$

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

provided $\sup_t E|\varepsilon_t|^{4r} < C < \infty$. We choose $r = \frac{3}{2}$ and get

$$(n-k)^{-r}\sum_{t=1+k}^{n} E^*\left(|w_{nt}^*|^{2r}\right) \leq (n-k)^{1-r} O_P(k^r) = O_P\left(\left(\frac{k^3}{n-k}\right)^{1/2}\right) = o_P(1),$$

if $k^3/n \to 0$.

Proof of Corollary 4.1. First, we show that $\sqrt{n-k}\left(g\left(\hat{\phi}(k)\right) - g(\phi(k))\right)/v_{g,k} \Rightarrow N(0,1)$, where $v_{g,k} \equiv \nabla' g(\phi(k))\left(\Gamma_k^{-1}B_k\Gamma_k^{-1}\right)\nabla g(\phi(k))$. A mean value expansion yields

$$\sqrt{n-k}\left(g\left(\hat{\phi}(k)\right) - g(\phi(k))\right) = \nabla' g(\phi(k))\sqrt{n-k}\left(\hat{\phi}(k) - \phi(k)\right) + r_k, \tag{A.16}$$

where $r_k = \left(\nabla' g\left(\bar{\phi}(k)\right) - \nabla' g(\phi(k))\right)\sqrt{n-k}\left(\hat{\phi}(k) - \phi(k)\right)$, $\bar{\phi}(k)$ lies in the segment connecting $\hat{\phi}(k)$ and $\phi(k)$, and we can show that $r_k = O_P\left(\frac{k}{n^{1/2}}\right)$ by Theorem 2.1 and Assumption 3 (ii). We then apply Theorem 2.2 to the first term in the RHS of (A.16) with $\ell(k) = \nabla' g(\phi(k))$. Second, we show that $\sqrt{n-k}\left(g\left(\hat{\phi}^*(k)\right) - g\left(\hat{\phi}(k)\right)\right)/v_{g,k} \Rightarrow^{d_{P^*}} N(0,1)$ in probability. A mean value expansion yields

$$\sqrt{n-k}\left(g\left(\hat{\phi}^*(k)\right) - g\left(\hat{\phi}(k)\right)\right) = \nabla' g(\phi(k))\sqrt{n-k}\left(\hat{\phi}^*(k) - \hat{\phi}(k)\right) + r_k^*, \tag{A.17}$$

where $r_k^* = \left(\nabla' g\left(\bar{\phi}^*(k)\right) - \nabla' g(\phi(k))\right)\sqrt{n-k}\left(\hat{\phi}^*(k) - \hat{\phi}(k)\right)$ and $\bar{\phi}^*(k)$ lies in the segment connecting $\hat{\phi}^*(k)$ and $\hat{\phi}(k)$. Thus,

$$\begin{aligned} |r_k^*| &\leq \left\|\nabla' g\left(\bar{\phi}^*(k)\right) - \nabla' g(\phi(k))\right\|\sqrt{n-k}\left\|\hat{\phi}^*(k) - \hat{\phi}(k)\right\| \leq M\left\|\bar{\phi}^*(k) - \phi(k)\right\|\sqrt{n-k}\left\|\hat{\phi}^*(k) - \hat{\phi}(k)\right\| \\ &\leq M\left(\left\|\hat{\phi}^*(k) - \hat{\phi}(k)\right\| + \left\|\hat{\phi}(k) - \phi(k)\right\|\right)\sqrt{n-k}\left\|\hat{\phi}^*(k) - \hat{\phi}(k)\right\|, \end{aligned}$$

since $\left\|\bar{\phi}^*(k) - \phi(k)\right\| \leq \left\|\bar{\phi}^*(k) - \hat{\phi}(k)\right\| + \left\|\hat{\phi}(k) - \phi(k)\right\| \leq \left\|\hat{\phi}^*(k) - \hat{\phi}(k)\right\| + \left\|\hat{\phi}(k) - \phi(k)\right\|$. For each of our two bootstrap methods we can show that $\left\|\hat{\phi}^*(k)-\hat{\phi}(k)\right\| = O_{P^*}\left(\frac{k^{1/2}}{n^{1/2}}\right)$ with probability approaching one. For the FWB, we have that $\hat{\phi}^*(k)-\hat{\phi}(k) = \hat{\Gamma}_k^{-1}(n-k)^{-1}\sum_{t=1+k}^{n} Y_{t-1,k}\hat{\varepsilon}_{t,k}^*$, which implies $\left\|\hat{\phi}^*(k)-\hat{\phi}(k)\right\| \leq \left\|\hat{\Gamma}_k^{-1}\right\|_1 \left\|(n-k)^{-1}\sum_{t=1+k}^{n} Y_{t-1,k}\hat{\varepsilon}_{t,k}^*\right\|$. Since $\left\|\hat{\Gamma}_k^{-1}\right\|_1 = O_P(1)$, as we argued before, and $\left\|(n-k)^{-1}\sum_{t=1+k}^{n} Y_{t-1,k}\hat{\varepsilon}_{t,k}^*\right\| = O_{P^*}\left(\frac{k^{1/2}}{n^{1/2}}\right)$ by (A.21) (cf. proof of Lemma A.4) and the Markov inequality, it follows that $\left\|\hat{\phi}^*(k)-\hat{\phi}(k)\right\| = O_{P^*}\left(\frac{k^{1/2}}{n^{1/2}}\right)$ in probability. The proof for the pairwise bootstrap follows similarly and is therefore omitted. Thus, for our two bootstrap methods, $r_k^* = O_{P^*}\left(\frac{k}{n^{1/2}}\right)$ in probability. Finally, an application of Theorems 3.1 and 3.2 with $\ell(k) = \nabla' g(\phi(k))$ delivers the result.

Proof of Lemma A.1. We apply Brillinger's (1981) Theorem 2.8.1. Define $\psi_j = 0$ for $j < 0$ and let $Cum(\cdot,\cdot,\cdot,\cdot)$ denote the fourth order joint cumulant a set of random variables. Using the MA(∞) representation of y_t, we have that

$$\begin{aligned}
\kappa_y(0,l_1,l_2,l_3) &= \sum_{j_1,\ldots,j_4=-\infty}^{\infty} \psi_{j_1}\psi_{j_2}\psi_{j_3}\psi_{j_4} Cum\left(\varepsilon_{t-j_1},\varepsilon_{t+l_1-j_2},\varepsilon_{t+l_2-j_3},\varepsilon_{t+l_3-j_4}\right) \\
&= \sum_{j_1,\ldots,j_4=-\infty}^{\infty} \psi_{j_1}\psi_{j_2}\psi_{j_3}\psi_{j_4} Cum\left(\varepsilon_{t},\varepsilon_{t+l_1+j_1-j_2},\varepsilon_{t+l_2+j_1-j_3},\varepsilon_{t+l_3+j_1-j_4}\right) \\
&\equiv \sum_{j_1,\ldots,j_4=-\infty}^{\infty} \psi_{j_1}\psi_{j_2}\psi_{j_3}\psi_{j_4} \kappa_\varepsilon\left(0,l_1+j_1-j_2,l_2+j_1-j_3,l_3+j_1-j_4\right),
\end{aligned}$$

where the first equality follows from the properties of cumulants (see e.g. Brillinger, 1981, p. 19), and the second equality follows from the stationarity of ε_t. It follows that

$$\begin{aligned}
&\sum_{l_1=-\infty}^{\infty}\sum_{l_2=-\infty}^{\infty}\sum_{l_3=-\infty}^{\infty} \left|\kappa_y(0,l_1,l_2,l_3)\right| \\
\leq &\sum_{j_1,\ldots,j_4=-\infty}^{\infty} \left|\psi_{j_1}\psi_{j_2}\psi_{j_3}\psi_{j_4}\right| \sum_{l_1=-\infty}^{\infty}\sum_{l_2=-\infty}^{\infty}\sum_{l_3=-\infty}^{\infty} \left|\kappa_\varepsilon\left(0,l_1+j_1-j_2,l_2+j_1-j_3,l_3+j_1-j_4\right)\right| < \infty,
\end{aligned}$$

given Assumption 1 (iii) and the absolute summability of ψ_j. ■

Proof of Lemma A.2. As in the proof of Berk's (1974) Lemma 3, we have that

$$\begin{aligned}
&\left\|\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right\|_1 = \left\|\hat{\Gamma}_k^{-1}\left(\Gamma_k-\hat{\Gamma}_k\right)\Gamma_k^{-1}\right\|_1 \leq \left\|\hat{\Gamma}_k^{-1}\right\|_1 \left\|\hat{\Gamma}_k-\Gamma_k\right\|_1 \left\|\Gamma_k^{-1}\right\|_1 \\
\leq &\left(\left\|\Gamma_k^{-1}\right\|_1 + \left\|\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right\|_1\right)\left\|\hat{\Gamma}_k-\Gamma_k\right\|_1 \left\|\Gamma_k^{-1}\right\|_1 \leq \left(C+\left\|\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right\|_1\right)\left\|\hat{\Gamma}_k-\Gamma_k\right\|_1 C,
\end{aligned}$$

using the fact that $\left\|\Gamma_k^{-1}\right\|_1$ is bounded. Thus,

$$\left\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 \leq \frac{C^2 \left\|\hat{\Gamma}_k - \Gamma_k\right\|_1}{1 - C\left\|\hat{\Gamma}_k - \Gamma_k\right\|_1}. \tag{A.18}$$

Now, using (A.1), we have

$$E\left\|\hat{\Gamma}_k - \Gamma_k\right\|_1^2 \leq E\left\|\hat{\Gamma}_k - \Gamma_k\right\|^2$$
$$= \sum_{i=1}^{k}\sum_{j=1}^{k} E\left((n-k)^{-1}\sum_{t=1+k}^{n}\left(y_{t-i}y_{t-j} - E\left(y_{t-i}y_{t-j}\right)\right)\right)^2. \tag{A.19}$$

Note that

$$(n-k)\,E\left((n-k)^{-1}\sum_{t=1+k}^{n}\left(y_{t-i}y_{t-j} - E\left(y_{t-i}y_{t-j}\right)\right)\right)^2 \tag{A.20}$$
$$= \quad (n-k)^{-1}\sum_{t=1+k}^{n}\sum_{s=1+k}^{n}[E\left(y_{t-i}y_{t-j}y_{s-i}y_{s-j}\right) - E\left(y_{t-i}y_{t-j}\right)E\left(y_{s-i}y_{s-j}\right)].$$

Letting $R(m) = E(y_t y_{t+m})$ for all $m \in \mathbb{Z}$, it follows that $E(y_{t-i}y_{t-j})\,E(y_{s-i}y_{s-j}) = [R(i-j)]^2$. By the stationarity of y_t, $E(y_{t-i}y_{t-j}y_{s-i}y_{s-j}) = E(y_t y_{t+i-j} y_s y_{s+i-j})$, and because y_t is a zero-mean process, it also follows that $E(y_t y_{t+i-j} y_s y_{s+i-j})$ can be written in terms of the fourth-order cumulant of $(y_t, y_{t+i-j}, y_s, y_{s+i-j})$ as (see Hannan, 1970, p. 209)

$$E\left(y_t y_{t+i-j} y_{t+(s-t)} y_{t+(s-t)+i-j}\right) = \kappa_y\left(t, t+i-j, t+(s-t), t+(s-t)+i-j\right)$$
$$+ \left(R(i-j)\right)^2 + \left(R(s-t)\right)^2 + R(s-t+i-j)\,R(s-t-(i-j)).$$

Stationarity of y_t implies $\kappa_y\left(t, t+i-j, t+(s-t), t+(s-t)+i-j\right) = \kappa_y\left(0, i-j, s-t, s-t+i-j\right)$, so that letting $l = s-t$ it follows from (A.20) that

$$(n-k)\,E\left((n-k)^{-1}\sum_{t=1+k}^{n}\left(y_{t-i}y_{t-j} - E\left(y_{t-i}y_{t-j}\right)\right)\right)^2$$
$$= \sum_{l=-(n-k-1)}^{n-k-1}\left(1 - \frac{|l|}{n-k}\right)\Big(\kappa_y\left(0, i-j, l, l+i-j\right) + \left(R(i-j)\right)^2 + \left(R(l)\right)^2$$
$$+R(l+i-j)\,R(l-(i-j)) - \left(R(i-j)\right)^2\Big)$$
$$= \sum_{l=-(n-k-1)}^{n-k-1}\left(1 - \frac{|l|}{n-k}\right)\Big(\kappa_y\left(0, i-j, l, l+i-j\right) + \left(R(l)\right)^2 + R(l+i-j)\,R(l-(i-j))\Big).$$

Thus, (A.20) is bounded by

$$\sum_{l=-\infty}^{\infty}\left(|\kappa_y(0,i-j,l,l+i-j)|+(R(l))^2+|R(l+i-j)R(l-(i-j))|\right)$$

$$\leq \sum_{l=-\infty}^{\infty}|\kappa_y(0,i-j,l,l+i-j)|+\sum_{l=-\infty}^{\infty}(R(l))^2+\left(\sum_{l=-\infty}^{\infty}(R(l))^2\right)^{1/2}\left(\sum_{l=-\infty}^{\infty}(R(l))^2\right)^{1/2}$$

$$\leq \sum_{l=-\infty}^{\infty}|\kappa_y(0,i-j,l,l+i-j)|+2\sum_{l=-\infty}^{\infty}(R(l))^2,$$

where the second term is bounded by $\sigma^4\left(\sum_{j=0}^{\infty}\psi_j\right)^2$, as in Berk (1974, p. 491), and the first term is bounded by $\sum_{l_1=-\infty}^{\infty}\sum_{l_2=-\infty}^{\infty}\sum_{l_3=-\infty}^{\infty}|\kappa_y(0,l_1,l_2,l_3)|$, which is bounded by Lemma A.1. Hence (A.19) is $O\left(k^2/(n-k)\right)$ and

$$E\left(\left\|\hat{\Gamma}_k-\Gamma_k\right\|_1^2\right)\leq E\left(\left\|\hat{\Gamma}_k-\Gamma_k\right\|^2\right)\leq Ck^2/(n-k)\to 0,$$

if $k^2/n\to 0$. By the Markov inequality, this implies that $\left\|\hat{\Gamma}_k-\Gamma_k\right\|_1=o_P(1)$ and thus from (A.18),

$$\left\|\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right\|_1\leq\frac{C^2\left\|\hat{\Gamma}_k-\Gamma_k\right\|_1}{1-C\left\|\hat{\Gamma}_k-\Gamma_k\right\|_1}=o_P(1),$$

proving (A.3). If instead $k^3/n\to 0$, then

$$E\left(\left(k^{1/2}\left\|\hat{\Gamma}_k-\Gamma_k\right\|_1\right)^2\right)\leq kE\left(\left\|\hat{\Gamma}_k-\Gamma_k\right\|^2\right)\leq Ck^3/(n-k)\to 0$$

showing that $k^{1/2}\left\|\hat{\Gamma}_k-\Gamma_k\right\|_1=o_P(1)$ and consequently $k^{1/2}\left\|\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right\|_1=o_P(1)$, which proves (A.4). ■

Proof of Lemma A.3. Following Lewis and Reinsel's (1985) proof of their Theorem 2, we can write

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

$$= w_{1n}+w_{2n}+w_{3n},$$

where $w_{1n}=\ell(k)'\left(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right)\sqrt{n-k}U_{1n}$, $w_{2n}=\ell(k)'\left(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right)\sqrt{n-k}U_{2n}$ and $w_{3n}=\ell(k)'\,\Gamma_k^{-1}\sqrt{n-k}U_{1n}$, with U_{1n} and U_{2n} defined as in the proof of Theorem 2.1. Using (A.2) we have $|w_{1n}|\leq\|\ell(k)\|\,k^{1/2}\left\|\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right\|_1\left\|k^{-1/2}\sqrt{n-k}U_{1n}\right\|=o_P(1)$, given that $\|\ell(k)\|\leq M_2^{1/2}$, that $k^{1/2}\left\|\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}\right\|_1=o_P(1)$ by Lemma A.2 if $\frac{k^3}{n}\to 0$, and that $\left\|k^{-1/2}\sqrt{n-k}U_{1n}\right\|=O_P(1)$, as we showed in the proof of Theorem 2.1. Similarly, $|w_{2n}| \leq \|\ell(k)\| k^{1/2} \left\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 \left\|k^{-1/2}\sqrt{n-k}U_{2n}\right\| = o_P(1)$, given Lemma A.2, eq. (A.4), and given that $\|U_{2n}\| = O_P\left(\frac{k^{1/2}}{(n-k)^{1/2}}\right)$ (cf. proof of Theorem 2.1). Finally, following Lewis and Reinsel (1985, p. 399) we let $v_{t,k} = \ell(k)' \Gamma_k^{-1} Y_{t-1,k}$. Then,

$$\begin{aligned} E|w_{3n}| &= (n-k)^{-1/2} E\left|\sum_{t=k+1}^{n} \ell(k)' \Gamma_k^{-1} Y_{t-1,k}(\varepsilon_t - \varepsilon_{t,k})\right| = (n-k)^{-1/2} E\left|\sum_{t=k+1}^{n} v_{t,k}(\varepsilon_t - \varepsilon_{t,k})\right| \\ &\leq (n-k)^{-1/2} \sum_{t=k+1}^{n} E\left|v_{t,k}(\varepsilon_t - \varepsilon_{t,k})\right| \leq (n-k)^{-1/2} \sum_{t=k+1}^{n} \left(E\left(v_{t,k}^2\right)\right)^{1/2} \left(E(\varepsilon_t - \varepsilon_{t,k})^2\right)^{1/2} \\ &\leq C\left\|\Gamma_k^{-1}\right\|_1^{1/2} M_2^{1/2} \sqrt{n-k} \sum_{j=1+k}^{\infty} |\phi_j| \to 0, \end{aligned}$$

given Assumption 2 (ii), where the first inequality holds by the triangle inequality, the second inequality holds by Cauchy-Schwartz and the third inequality holds because $\left(E(\varepsilon_t - \varepsilon_{t,k})^2\right)^{1/2} \leq C\sum_{j=k+1}^{\infty} |\phi_j|$ (cf. proof of Theorem 2.1, (A.18)) and because

$$E\left(v_{t,k}^2\right) = \ell(k)' \Gamma_k^{-1} \ell(k) = \left\|\Gamma_k^{-1/2} \ell(k)\right\|^2 \leq \left\|\Gamma_k^{-1/2}\right\|_1^2 \|\ell(k)\|^2 \leq \left\|\Gamma_k^{-1}\right\|_1 M_2,$$

which is bounded under our assumptions. ■

Proof of Lemma A.4. Adding and subtracting appropriately yields

$$r_{fwb}^* = \ell(k)' \left(\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right) v_k^{-1} (n-k)^{-1/2} \sum_{t=1+k}^{n} Y_{t-1,k} \hat{\varepsilon}_{t,k}^*,$$

implying that

$$\begin{aligned} \left|r_{fwb}^*\right| &\leq \|\ell(k)\| \left\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 \left|v_k^{-1}\right| \left\|(n-k)^{-1/2} \sum_{t=1+k}^{n} Y_{t-1,k} \hat{\varepsilon}_{t,k}^*\right\| \\ &\leq Ck^{1/2} \left\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 (n-k)^{1/2} k^{-1/2} \left\|(n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} \hat{\varepsilon}_{t,k}^*\right\|, \end{aligned}$$

for some finite constant C, given that $\|\ell(k)\|$ and $\left|v_k^{-1}\right|$ are bounded. Since

$$E^*\left(\left\|(n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} \hat{\varepsilon}_{t,k}^*\right\|^2\right) = O_P\left(\frac{k}{(n-k)}\right), \tag{A.21}$$

we have $E^*\left|r_{fwb}^*\right| \leq Ck^{1/2} \left\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\right\|_1 O_P(1) = o_P(1)$, where the equality holds by an application of Lemma A.2, see (A.4). By the Markov inequality, for any $\delta > 0$, we have $P^*\left(\left|r_{fwb}^*\right| > \delta\right) \leq \frac{1}{\delta} E^*\left(\left|r_{fwb}^*\right|\right) = o_P(1)$ and the desired result follows.

To complete the proof, we prove (A.21). We can write

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

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

where the last inequality follows because $E^*\left(\hat{\varepsilon}^*_{t,k}\hat{\varepsilon}^*_{s,k}\right)=0$ if $t\neq s$ and $E^*\left(\hat{\varepsilon}^*_{t,k}\hat{\varepsilon}^*_{s,k}\right)=\hat{\varepsilon}^2_{t,k}$ otherwise. Next, we show that $\chi_1=O_P(k)$, which in turn implies (A.21). Applying the triangle inequality first and then the Cauchy-Schwartz inequality, we have that

$$\begin{aligned}|\chi_{1n}| &\leq (n-k)^{-1}\sum_{t=1+k}^{n}\left|Y'_{t-1,k}Y_{t-1,k}\right|\left|\hat{\varepsilon}^2_{t,k}\right|\leq\left((n-k)^{-1}\sum_{t=1+k}^{n}\left|Y'_{t-1,k}Y_{t-1,k}\right|^2\right)^{1/2}\left((n-k)^{-1}\sum_{t=1+k}^{n}\hat{\varepsilon}^4_{t,k}\right)^{1/2}\\ &= \left((n-k)^{-1}\sum_{t=1+k}^{n}\left\|Y_{t-1,k}\right\|^4\right)^{1/2}\left((n-k)^{-1}\sum_{t=1+k}^{n}\hat{\varepsilon}^4_{t,k}\right)^{1/2}\equiv A_1.A_2.\end{aligned}$$

Because $\sup_t E|y_t|^4\leq C<\infty$ under our conditions, $E\left\|Y_{t-1,k}\right\|^4=O\left(k^2\right)$, which implies that $A_1=O_P(k)$. Since $A_2=O_P(1)$, as we show next, this proves the result.

To show that $A_2=O_P(1)$, note that $\hat{\varepsilon}_{t,k}=\varepsilon_t-\sum_{j=1+k}^{\infty}\phi_j y_{t-j}-\left(\hat{\phi}(k)-\phi(k)\right)'Y_{t-1,k}$. By the c$_r$-inequality (Davidson, 1994, p. 140), we have that

$$(n-k)^{-1}\sum_{t=1+k}^{n}\hat{\varepsilon}^4_{t,k}\leq C(n-k)^{-1}\sum_{t=1+k}^{n}\left(\varepsilon_t^4+\left|\sum_{j=1+k}^{\infty}\phi_j y_{t-j}\right|^4+\left|\left(\hat{\phi}(k)-\phi(k)\right)'Y_{t-1,k}\right|^4\right)\equiv B_1+B_2+B_3.$$

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

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

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

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

given Theorem 2.1, the fact that $\|Y_{t-1,k}\|^4 = O_P\left(k^2\right)$ and $k^2/n \to 0$ under our assumptions.

Proof of Lemma A.5. Simple algebra shows that $r^*_{pb} = A_1 + A_2$, where

$$\begin{aligned} A_1 &= \ell(k)' \sqrt{n-k}\left(\hat{\Gamma}^{*-1}_{pb,k} - \hat{\Gamma}^{-1}_k\right) v_k^{-1} (n-k)^{-1} \sum_{t=1+k}^{n} Y^*_{t-1,k}\hat{\varepsilon}^*_{t,k} \\ A_2 &= \ell(k)' \sqrt{n-k}\left(\hat{\Gamma}^{-1}_k - \Gamma^{-1}_k\right) v_k^{-1} (n-k)^{-1} \sum_{t=1+k}^{n} Y^*_{t-1,k}\hat{\varepsilon}^*_{t,k}. \end{aligned}$$

Consider A_2 first. We have that

$$|A_2| \leq C \|\ell(k)\| k^{1/2} \left\|\hat{\Gamma}^{-1}_k - \Gamma^{-1}_k\right\| k^{-1/2} (n-k)^{1/2} \left\|(n-k)^{-1} \sum_{t=1+k}^{n} Y^*_{t-1,k}\hat{\varepsilon}^*_{t,k}\right\|. \tag{A.22}$$

Next we will show that

$$E^*\left(\left\|(n-k)^{-1} \sum_{t=1+k}^{n} Y^*_{t-1,k}\hat{\varepsilon}^*_{t,k}\right\|^2\right) = O_P\left(\frac{k}{n-k}\right), \tag{A.23}$$

which, combined with Lemma A.2 and (A.22), shows that $A_2 = o_{P^*}(1)$ in probability. To prove (A.23), note that

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

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

$$E^*\left(Y^{*\prime}_{t-1,k}\hat{\varepsilon}^*_{t,k} Y^*_{s-1,k}\hat{\varepsilon}^*_{s,k}\right) = E^*\left(Y^{*\prime}_{t-1,k}\hat{\varepsilon}^*_{t,k}\right) E^*\left(Y^*_{s-1,k}\hat{\varepsilon}^*_{s,k}\right) = \left\|(n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}\hat{\varepsilon}_{t,k}\right\|^2 = 0,$$

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

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

Thus, the LHS of (A.23) equals $(n-k)^{-2} \sum_{t=1+k}^{n} Y'_{t-1,k} Y_{t-1,k}\hat{\varepsilon}^2_{t,k}$, which is $O_P\left(\frac{k}{n-k}\right)$, as we showed in the proof of Lemma A.4.

Next, we show that $A_1 = o_{P^*}(1)$ in probability. We can write

$$\begin{aligned} |A_1| &\leq C \|\ell(k)\| \left\| \hat{\Gamma}_{pb,k}^{*-1} - \hat{\Gamma}_k^{-1} \right\|_1 (n-k)^{1/2} \left\| (n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k}^{*} \hat{\varepsilon}_{t,k}^{*} \right\| \\ &\leq C \left\| \hat{\Gamma}_{pb,k}^{*-1} - \hat{\Gamma}_k^{-1} \right\|_1 (n-k)^{1/2} O_{P^*}\left(\frac{k^{1/2}}{(n-k)^{1/2}} \right), \end{aligned}$$

conditional on the data, given (A.23) and the Markov inequality. So, it suffices to show that $k^{1/2} \left\| \hat{\Gamma}_{pb,k}^{*-1} - \hat{\Gamma}_k^{-1} \right\|_1 = o_{P^*}(1)$ in probability. Following the same argument as in the proof of Lemma A.2, it is enough to show that $k^{1/2} \left\| \hat{\Gamma}_{pb,k}^{*} - \hat{\Gamma}_k \right\|_1 = o_{P^*}(1)$ in probability, or by the Markov inequality and the inequality (A.1), that $kE^*\left(\left\| \hat{\Gamma}_{pb,k}^{*} - \hat{\Gamma}_k \right\|^2 \right) = o_P(1)$. By definition of the Euclidean matrix norm,

$$\begin{aligned} E^*\left(\left\| \hat{\Gamma}_{pb,k}^{*} - \hat{\Gamma}_k \right\|^2 \right) &= tr\left((n-k)^{-2} \sum_{t=1+k}^{n} \sum_{s=1+k}^{n} E^*\left[\left(Y_{t-1,k}^{*} Y_{t-1,k}^{*\prime} - \hat{\Gamma}_k \right) \left(Y_{s-1,k}^{*} Y_{s-1,k}^{*\prime} - \hat{\Gamma}_k \right) \right] \right) \\ &= tr\left((n-k)^{-2} \sum_{t=1+k}^{n} \left(Y_{t-1,k} Y_{t-1,k}' - \hat{\Gamma}_k \right) \left(Y_{t-1,k} Y_{t-1,k}' - \hat{\Gamma}_k \right) \right) \\ &= (n-k)^{-1}\, tr\left((n-k)^{-1} \sum_{t=1+k}^{n} Y_{t-1,k} Y_{t-1,k}' Y_{t-1,k} Y_{t-1,k}' - \hat{\Gamma}_k \right), \end{aligned}$$

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

References

[1] Andrews, D.W.K. (1991): "Heteroskedastic and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 59, 817-858.

[2] Berk, K. (1974): "Consistent Autoregressive Spectral Estimates", *Annals of Statistics*, 2, 489-502.

[3] Bhansali, R. (1978): "Linear Prediction by Autoregressive Model Fitting in the Time Domain", *Annals of Statistics*, 6, 224-231.

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

[5] Brockwell, P.J., and R.A. Davis (1991): *Time Series: Theory and Methods.* Second edition, Springer-Verlag, New York.

[6] Bühlmann, P. (1995): "Moving-Average Representation for Autoregressive Approximations", *Stochastic Processes and Their Applications*, 60, 331-342.

[7] Bühlmann, P. (1997): "Sieve Bootstrap for Time Series", *Bernoulli*, 3, 123-148.

[8] Canova, F. (1995): "Vector Autoregressions: Specification, Estimation, Inference and Forecasting", in: *The Handbook of Applied Econometrics*, Vol. 1, edited by H. Pesaran and M. Wickens, Basil Blackwell, London.

[9] Choi, E., and P. Hall (2000): "Bootstrap Confidence Regions Computed on Autoregressions of Arbitrary Order", *Journal of the Royal Statistical Society, Series B*, 62, 461-477.

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

[11] Diebold, F.X., and L. Kilian (2001): "Measuring Predictability: Theory and Macroeconomic Applications", *Journal of Applied Econometrics*, 16, 657-669.

[12] Durrett, R. (1996): *Probability: Theory and Examples.* Second edition, Duxury Press, Belmont, California.

[13] Eicker, F. (1963): "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions", *Annals of Mathematical Statistics*, 34, 447-456.

[14] Freedman, D.A. (1981): "Bootstrapping Regression Models", *Annals of Statistics*, 9, 1218-1228.

[15] Galbraith, J.W. (2003): "Content Horizons for Univariate Time-Series Forecasts", *International Journal of Forecasting*, 19, 43-55.

[16] Geman, S., and C.-R. Hwang (1982): "Nonparametric Maximum Likelihood Estimation by the Method of Sieves", *Annals of Statistics*, 10, 401-414.

[17] Gonçalves, S., and H. White (2002): "Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models", CIRANO working paper s-41.

[18] Gonçalves, S., and L. Kilian (2003): "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form", mimeo, Université de Montréal. (available at: http://www.mapageweb.umontreal.ca/goncals/research.html)

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

[20] Hannan, E. (1970): *Multiple Time Series*, Wiley, New York.

[21] Hansen, B.E. (2000): "Testing for Structural Change in Conditional Models", *Journal of Econometrics*, 97, 93-115.

[22] Horn, R. and C. Johnson (1985): *Matrix Analysis*, Cambridge University Press, UK.

[23] Inoue, A., and L. Kilian (2002): "Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in VAR(∞) Models", *International Economic Review,* 43, 309-332.

[24] Kreiss, J.P. (1997): "Asymptotic Properties of Residual Bootstrap for Autoregressions", mimeo, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.

[25] Künsch, H.R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations", *Annals of Statistics*, 17, 1217-1241.

[26] Kuersteiner, G.M. (2001): "Optimal Instrumental Variables Estimation for ARMA Models", *Journal of Econometrics*, 104, 359-405.

[27] Kuersteiner, G.M. (2002): "Efficient IV Estimation for Autoregressive Models with Conditional Heteroskedasticity", *Econometric Theory*, 18, 547-583.

[28] Lewis, R. and G. Reinsel (1985): "Prediction of Multivariate Time Series of Autoregressive Model Fitting", *Journal of Multivariate Analysis*, 16, 393-411.

[29] Liu, R.Y. (1988): "Bootstrap Procedure under Some Non-i.i.d. Models", *Annals of Statistics*, 16, 1696-1708.

[30] Lütkepohl, H. (1988a): "Asymptotic Distribution of the Moving-Average Coefficients of an Estimated Vector Autoregressive Process", *Econometric Theory*, 4, 77-85.

[31] Lütkepohl, H. (1988b): "Prediction Tests for Structural Stability", *Journal of Econometrics*, 39, 267-296.

[32] Lütkepohl, H., and D.S. Poskitt (1991): "Estimating Orthogonal Impulse Responses via Vector Autoregressive Models", *Econometric Theory*, 7, 487-496.

[33] Lütkepohl, H., and D.S. Poskitt (1996): "Testing for Causation Using Infinite Order Vector Autoregressive Processes", *Econometric Theory*, 12, 61-87.

[34] Lütkepohl, H., and P. Saikkonen (1997): "Impulse Response Analysis in Infinite-Order Cointegrated Vector Autoregressive Processes", *Journal of Econometrics*, 81, 127-157.

[35] Mammen, E. (1993): "Bootstrap and Wild Bootstrap for High Dimensional Linear Models", *Annals of Statistics*, 21, 255-285.

[36] Ng, S. and P. Perron (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.

[37] Nicholls, D.F., and A.R. Pagan. (1983): "Heteroskedasticity in Models with Lagged Dependent Variables", *Econometrica*, 51, 1233-1242.

[38] Paparoditis, E. (1996): "Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes", *Journal of Multivariate Analysis*, 57, 277-296.

[39] Saikkonen, P. (1992): "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation", *Econometric Theory*, 8, 1-27.

[40] Saikkonen, P. , and H. Lütkepohl (1996): "Infinite-Order Cointegrated Vector Autoregressive Processes", *Econometric Theory*, 12, 814-844.

[41] Stock, J.H., and M.W. Watson (2001): "Vector Autoregressions", *Journal of Economic Perspectives*, 15, Fall, 101-115.

[42] White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48, 817-838.

[43] White, H. (2000): *Asymptotic Theory for Econometricians*, Revised Edition, Academic Press, San Diego.

[44] Wu, C.F.J. (1986): "Jackknife Bootstrap and other Resampling Methods in Regression Analysis", *Annals of Statistics*, 14, 1261-1295.