October 17, 1997

Problems Related to Bootstrapping Impulse Responses of Autoregressive Processes

Alexander Benkwitz[1], Helmut Lütkepohl[2] and Michael H. Neumann[1]

Abstract

Bootstrap confidence intervals for impulse responses computed from autoregressive processes are considered. A detailed analysis of the methods in current use shows that they are not very reliable in some cases. In particular, there are theoretical reasons for them to have actual coverage probabilities which deviate considerably from the nominal level in some situations of practical importance. For a simple case alternative bootstrap methods are proposed which provide correct results asymptotically.

[1]Humboldt-Universität, Sonderforschungsbereich 373, Spandauer Straße 1, D – 10178 Berlin, Germany
[2]Humboldt-Universität, Institut für Statistik und Ökonometrie, Spandauer Straße 1, D – 10178 Berlin, Germany

We thank Jörg Breitung for helpful comments on this paper. The research was carried out within the Sonderforschungsbereich 373 at Humboldt University Berlin and was printed using funds made available by the Deutsche Forschungsgemeinschaft.

1 Introduction

In vector autoregressive (VAR) analyses impulse responses are often used for interpreting the relation between the variables involved. In that case bootstrap confidence intervals (CIs) or regions are often reported because they are regarded as being more reliable than confidence intervals based on asymptotic theory. Support for this view seemingly comes from the skewness of the bootstrap intervals which contrasts with the symmetry of standard asymptotic intervals. Some Monte Carlo studies have also confirmed this belief (see, e.g., Fachin & Bravetti (1996), Kilian (1995)). On the other hand, it was also found that in some cases bootstrap CIs are not very reliable. In fact, they can lead to extremely poor CIs with actual confidence content substantially different from the nominal level (e.g., Griffiths & Lütkepohl (1989), Kilian (1995, 1996), Fachin & Bravetti (1996)). Of course, this may partly be a small sample problem and hence small sample modifications and corrections have been proposed (e.g., Kilian (1995, 1996), Fachin & Bravetti (1996)). Although these modifications are quite successful in some cases, it will turn out that they do not help with the particular problems we encounter with impulse responses in certain regions of the parameter space.

In this study we will point out that in addition to these small sample problems there are also conceptual problems that prevent the usual bootstrap CIs for impulse responses to have the correct probability content even asymptotically. Sims & Zha (1994) also launched a critique of the usual approaches to construct CIs for impulse responses. Their critique is based on a Bayesian point of view, however. In contrast, we will remain within the classical asymptotic arena and argue that even in this framework problems may arise.

The main problems result from the fact that for the standard bootstrap to work the convergence rate of the estimators to their asymptotic distribution must remain constant over the whole parameter space. It was noted, e.g., by Lütkepohl (1991, Sec. 3.7) that this condition is not even satisfied for stationary VARs let alone nonstationary ones. In particular, it is not satisfied for some cases of interest from an applied point of view. We will discuss the problem in detail for the simplest case of a stationary univariate AR process of order one (AR(1)). For that case we will also consider possible solutions and we will discuss their potential for being generalized to higher order and higher dimensional processes. Unfortunately, it turns out, however, that proposals which work well for the simplest case are not easily generalizable.

The structure of the paper is as follows. The general framework of the analysis is presented in the next section and inference on impulse responses is considered in Section 3. In that section we also draw attention to some basic problems of asymptotic inference in the present context. In Section 4 a detailed analysis of the AR(1) case is provided and possible solutions are offered for this special case. Illustrative simulations are discussed in Section 5. Conclusions follow in Section 6.

The following notation is used throughout. The symbol I_K denotes the $(K \times K)$ identity matrix and the operator vec stacks the columns of a matrix in a column vector. Moreover, $\mathbf{R}$ denotes the real numbers, $O(\cdot), o(\cdot), O_P(\cdot)$ and $o_P(\cdot)$ are the usual symbols for the order of convergence and convergence in probability, respectively. Furthermore, $\stackrel{d}{\rightarrow}$ signifies convergence in distribution. $N(\mu, \sigma^2)$ indicates a normal distribution with mean μ and variance σ^2. More generally, $\mathcal{L}(X)$ denotes the distribution function of the random variable X. We use $P(\cdot)$ to denote the probability of some event and $P_\alpha(\cdot)$ if the probability corresponding to a specific parameter α of the underlying distribution is of interest. log is the natural logarithm, LS stands for least squares and DGP means data generation process.

2 VAR Processes and Impulse Response Functions

Many macroeconomic analyses are based on VAR models of the type

$$A_0 y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \qquad (2.1)$$

where $y_t = [y_{1t}, \ldots, y_{Kt}]'$ is a K-dimensional vector of observable variables, the $A_i, i = 0, 1, \ldots, p$, are $(K \times K)$ coefficient matrices and $u_t = [u_{1t}, \ldots, u_{Kt}]'$ is a white noise process, that is, the u_t are temporally uncorrelated or independent with zero mean and nonsingular (positive definite) covariance matrix Σ_u. The maximum lag length p is usually referred to as the order of the VAR process and the process is briefly called a VAR(p) process. Usually there are also deterministic terms such as intercepts or seasonal dummy variables in the DGP of economic time series. They are deleted in (2.1) because they are not important for our purposes.

The model (2.1) is not identified without any restrictions on the coefficients. The matrix A_0 represents the instantaneous relationships between the variables. Identifying restrictions may be placed on this matrix and also on the other coefficient matrices. If $A_0 = I_K$, the

$(K \times K)$ identity matrix, the process is said to be in *reduced form*. If $A_0 \neq I_K$, (2.1) is a *structural form*. The reduced form may then be obtained by left-multiplying (2.1) by A_0^{-1}, that is, A_0 has to be invertible. If (2.1) is a structural form the u_t are structural residuals. In that case, Σ_u is often assumed to be diagonal. The model is called recursive if A_0 is (upper or lower) triangular with unit diagonal and Σ_u is diagonal. In contrast, in the reduced form the u_t are usually contemporaneously correlated, that is, Σ_u is a general positive definite covariance matrix. In that case the u_t are also the errors of optimal linear 1-step ahead forecasts and, hence, Σ_u is the 1-step ahead forecast error covariance matrix.

The process (2.1) may be stationary or it may contain integrated and cointegrated variables. The problems of interest in the following are present in both cases. For simplicity we will therefore predominantly consider the stationary case.

Regardless of the stationarity properties, the model (2.1) summarizes the instantaneous and intertemporal relations between the variables. The exact form of these relations is usually difficult to see directly from the A_i coefficients. Therefore impulse response functions are often computed which represent the marginal responses of the variables of the system to an impulse in one of the variables. These may be regarded as conditional forecasts of the variables given that they have been zero up to time 0 when an impulse in one of the variables occurs. Depending on the kind of impulse hitting the system there are various different impulse responses that have been used for interpreting VAR models. For detailed discussions see Sims (1980, 1981), Lütkepohl (1990, 1991), Watson (1994), Lütkepohl & Breitung (1997). The important property of these quantities from the point of view of our analysis is that they are particular nonlinear functions of the parameters of the model (2.1), say,

$$\phi_{ij,h} = \phi_{ij,h}(A_0, A_1, \ldots, A_p), \tag{2.2}$$

where $\phi_{ij,h}$ represents the response of variable i to an impulse in variable j, h periods ago. For instance, in the simple case of a reduced form with $A_0 = I_K$ and if the so-called forecast error impulse responses are considered (see Lütkepohl (1991, Sec. 2.3.2)), $\phi_{ij,h}$ is the ijth element of the matrix Φ_h obtained recursively as

$$\Phi_h = \sum_{j=1}^{h} \Phi_{h-j} A_j, \quad h = 1, 2, \ldots, \tag{2.3}$$

where $\Phi_0 = I_K$. If $p = 1$ this is easily seen to imply $\Phi_h = A_1^h$. Note that generally the $\phi_{ij,h}$

are sums of products of the elements of the A_i. In the next section we will discuss potential pitfalls in inference procedures for impulse responses that result from this special structure of these quantities. There are also impulse responses which depend on the elements of the white noise covariance matrix Σ_u in addition to the A_i coefficients. For simplicity we will not consider this possibility here because the potential problems arise even for the simpler case indicated in (2.2).

3 Inference on Impulse Responses

Usually the coefficients of the model (2.1) are estimated by some standard procedure such as LS and estimators of the impulse responses are then obtained as

$$\hat{\phi}_{ij,h} = \phi_{ij,h}(\hat{A}_0, \hat{A}_1, \dots, \hat{A}_p) \tag{3.1}$$

where the $\hat{A}_0, \dots, \hat{A}_p$ are, of course, the estimated VAR coefficient matrices. Assuming that the $\hat{A}_i$ have an asymptotic normal distribution,

$$\sqrt{T}\mathrm{vec}([\hat{A}_0, \dots, \hat{A}_p] - [A_0, \dots, A_p]) \xrightarrow{d} N(0, \Sigma_{\hat{A}}), \tag{3.2}$$

we have that the $\phi_{ij,h}$ have an asymptotic normal distribution as well,

$$\sqrt{T}(\hat{\phi}_{ij,h} - \phi_{ij,h}) \xrightarrow{d} N(0, \sigma^2_{ij,h}), \tag{3.3}$$

where

$$\sigma^2_{ij,h} = \frac{\partial \phi_{ij,h}}{\partial \alpha'} \Sigma_{\hat{A}} \frac{\partial \phi_{ij,h}}{\partial \alpha} \tag{3.4}$$

with $\alpha = \mathrm{vec}[A_0, \dots, A_p]$, and $\partial \phi_{ij,h}/\partial \alpha$ denotes a vector of partial derivatives. The result (3.3) holds if $\sigma^2_{ij,h}$ is nonzero which is a crucial condition for asymptotic inference to work. Note that $\Sigma_{\hat{A}}$ may be singular if there are constraints on the coefficients or if the variables are integrated and/or cointegrated (see Lütkepohl (1991, Chapter 11)). However, even if $\Sigma_{\hat{A}}$ is nonsingular, $\sigma^2_{ij,h}$ may be zero because the partial derivatives in (3.4) may be zero. In fact, they will usually be zero in parts of the parameter space because the $\phi_{ij,h}$ generally consist of sums of products of the VAR coefficients and, hence, the partial derivatives will also be sums of products of such coefficients which may be zero.

To see the problem more clearly, consider the simple case of a one-dimensional AR(1) process $y_t = \alpha y_{t-1} + u_t$. In this case $\phi_h = \alpha^h$. Suppose $\hat{\alpha}$ is an estimator of α satisfying

$$\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \sigma^2_{\hat{\alpha}}) \tag{3.5}$$

with $\sigma^2_{\hat{\alpha}} \neq 0$. Then

$$\sqrt{T}(\hat{\alpha}^2 - \alpha^2) \xrightarrow{d} N(0, \sigma^2_{\hat{\alpha}^2}) \tag{3.6}$$

with $\sigma^2_{\hat{\alpha}^2} = 4\alpha^2\sigma^2_{\hat{\alpha}}$ which is obviously zero if $\alpha = 0$. Of course, this is a well-known result as in that case $T\hat{\alpha}^2$ is known to have a proper asymptotic distribution and thus $\sqrt{T}\hat{\alpha}^2$ must be degenerate. Hence, estimated impulse responses may have a degenerate asymptotic distribution even if the underlying DGP is a well behaved stationary process.

One might be tempted to use (3.6) as a starting point for the construction of confidence intervals for α^2. Since the estimated $\sigma^2_{\hat{\alpha}^2}$ obtained by replacing α and $\sigma^2_{\hat{\alpha}}$ by their usual LS estimators will be nonzero almost surely one may consider the t-ratio $\sqrt{T}(\hat{\alpha}^2 - \alpha^2)/2\hat{\alpha}\hat{\sigma}_{\hat{\alpha}}$ as a basis for constructing a CI. We will see in the next section that this results in a conservative CI for the case $\alpha = 0$. It is not clear that a conservative CI will be obtained in the more interesting cases where impulse responses from higher dimensional processes are considered. Of particular concern is the fact that the procedure fails for a case of special interest, namely when the impulse responses are all zero. This failure is typical also for higher dimensional processes for which the order may also be greater than 1. Of course, the situation where some variable does not react to an impulse in some other variable, i.e. the impulse response is zero, is of particular interest because it means that there is no causal link in a certain part of the system. Hence, the asymptotic CIs fail in situations of particular importance. Note, however, that for stable, stationary VAR(p) processes, the asymptotic CIs work allright for $\phi_{ij,h}$ with $h \leq p$. This fact was used by Lütkepohl & Poskitt (1996) and Saikkonen & Lütkepohl (1995) to point out a possibility for circumventing the problem by assuming that the true DGP is an infinite order VAR process. Although the asymptotic problems can be fixed in this way, simulations reported in Lütkepohl & Poskitt (1996) indicate that this may not be very helpful in samples of the size typically available in macroeconometrics.

It may be worth pointing out that a similar problem exists for forecast error variance decompositions. For simplicity, consider a bivariate VAR(1), $y_t = A_1 y_{t-1} + u_t$, with

$$A_1 = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}$$

and suppose the white noise covariance is known to be $\Sigma_u = I_2$. Then the forecast error covariance matrix of a 2-step forecast is known to be

$$\Sigma_y(2) = I_2 + A_1 A_1' = \begin{bmatrix} 1+\alpha_{11}^2+\alpha_{12}^2 & \bullet \\ \alpha_{21}\alpha_{11}+\alpha_{22}\alpha_{12} & 1+\alpha_{21}^2+\alpha_{22}^2 \end{bmatrix}$$

(e.g., Lütkepohl (1991, Sec. 2.2)). Hence the contribution of the second variable to the forecast error variance of the first variable is

$$\frac{\alpha_{12}^2}{1+\alpha_{11}^2+\alpha_{12}^2}.$$

The estimated version of this fraction has T-convergence if $\alpha_{12} = 0$ and, thus, the second variable does not contribute to the forecast error variance of the first variable. Consequently, asymptotic CIs for these quantities will be problematic in this case of special interest in empirical work.

In practice, bootstrap methods are often used to construct CIs for impulse responses, forecast error variance components etc.. We emphasize that derivations of the properties of these methods usually rely on asymptotic theory. Therefore it should not come as a surprise that standard bootstrap techniques do not work well in general for some cases of interest here. In the next section we will consider in detail the implications of these phenomena for constructing CIs based on asymptotic theory as well as the bootstrap. We will do so for the simplest case of a stationary univariate AR(1) process. We will begin with an analysis of standard methods and then discuss modifications which circumvent the problems of the former.

4 CIs for Impulse Responses from a Univariate AR(1)

Suppose that we observe $y_0, \ldots, y_T$ generated by the univariate process

$$y_t = \alpha y_{t-1} + u_t, \tag{4.1}$$

where the u_t's are i.i.d. with mean 0 and variance σ_u^2. For simplicity, we assume that all moments of u_t are finite. We are interested in a CI for the impulse response coefficient $\phi_h = \alpha^h$. The starting point for all of our methods will be the least squares estimator

$$\widehat{\alpha}_T = \sum_{t=1}^{T} y_t y_{t-1} / \sum_{t=1}^{T} y_{t-1}^2. \tag{4.2}$$

We will first discuss CIs based on asymptotic theory and then consider bootstrap CIs.

4.1 Naive confidence intervals based on first-order asymptotic theory

Let $|\alpha| < 1$. It is well-known that $\sqrt{T}(\hat{\alpha}_T - \alpha) \xrightarrow{d} N(0, \sigma^2_{\hat{\alpha}_T} = 1 - \alpha^2)$ so that with $T^{-1}\sum_{t=1}^{T} y_{t-1}^2 \to \sigma_u^2/(1-\alpha^2)$ we have

$$\frac{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}{\sigma_u}(\widehat{\alpha}_T - \alpha) \xrightarrow{d} N(0,1),$$

see, for example, Anderson (1959). The standard approach uses $\widehat{\alpha}_T^h$ as a starting point for constructing a confidence interval for α^h.

Let

$$\widetilde{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{\alpha}_T y_{t-1})^2.$$

It is easy to see that $\widetilde{\sigma}_u^2 = \sigma_u^2 + o_P(1)$. The asymptotic distribution of $\widehat{\alpha}_T^h - \alpha^h$ can be found by the so-called delta method. Because of

$$\begin{aligned}
\widehat{\alpha}_T^h - \alpha^h &= [\alpha + (\widehat{\alpha}_T - \alpha)]^h - \alpha^h \\
&= \sum_{k=0}^{h-1}\binom{h}{k}\alpha^k(\widehat{\alpha}_T - \alpha)^{h-k} \\
&= h\alpha^{h-1}(\widehat{\alpha}_T - \alpha) + O_P(T^{-1}),
\end{aligned} \tag{4.3}$$

and, using again $\widehat{\alpha}_T = \alpha + O_P(T^{-1/2})$, we obtain for $\alpha \neq 0$ that

$$\frac{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}{\widetilde{\sigma}_u h \widehat{\alpha}_T^{h-1}}(\widehat{\alpha}_T^h - \alpha^h) \xrightarrow{d} N(0,1). \tag{4.4}$$

This asymptotic result may be used to establish an asymptotic CI for α^h with a nominal coverage probability of $1-\gamma$ as

$$CI_1 = \left[\widehat{\alpha}_T^h - \frac{\widetilde{\sigma}_u h|\widehat{\alpha}_T|^{h-1}}{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}c_{(1-\gamma/2)}, \widehat{\alpha}_T^h + \frac{\widetilde{\sigma}_u h|\widehat{\alpha}_T|^{h-1}}{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}c_{(1-\gamma/2)}\right], \tag{4.5}$$

where c_β denotes the β-quantile of the standard normal distribution. It follows immediately from (4.4) that for $\alpha \neq 0$,

$$P\left(\alpha^h \in CI_1\right) \longrightarrow 1 - \gamma \quad \text{as} \quad T \to \infty, \tag{4.6}$$

that is, CI_1 has asymptotically the correct coverage probability.

However, for $\alpha = 0$, it turns out that

$$\frac{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}{\widetilde{\sigma}_u h \widehat{\alpha}_T^{h-1}}(\widehat{\alpha}_T^h - \alpha^h) = \frac{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}{\widetilde{\sigma}_u h}(\widehat{\alpha}_T - \alpha) \xrightarrow{d} N(0, 1/h^2). \tag{4.7}$$

As a consequence CI_1 is conservative, with an asymptotic coverage probability larger than the prescribed $1-\gamma$. In terms of the length of the interval, CI_1 is about h times too large. To see this note that, for $\alpha = 0$, we have that $\sqrt{T}(\widehat{\alpha}_T^h - \alpha^h)/\widehat{\alpha}_T^{h-1} = \sqrt{T}(\widehat{\alpha}_T - \alpha) \xrightarrow{d} N(0, \sigma_{\widehat{\alpha}}^2)$, which means that the CI $[\widehat{\alpha}^h - c_{1-\gamma/2}\frac{\widehat{\alpha}^{h-1}\widehat{\sigma}_{\widehat{\alpha}}}{\sqrt{T}}, \widehat{\alpha}^h + c_{1-\gamma/2}\frac{\widehat{\alpha}^{h-1}\widehat{\sigma}_{\widehat{\alpha}}}{\sqrt{T}}]$ has the desired coverage probability at least asymptotically. Hence, CI_1 is h times as large as a proper $(1-\gamma)$ CI and thus has a substantially greater coverage probability than the intended $1-\gamma$. We quantify the error in coverage probability in our simulations reported in Section 5.

At this point some general comments are in order. The difficulty in getting asymptotically correct confidence intervals is caused by the fact that $\widehat{\alpha}_T^h - \alpha^h$ has a different limiting behaviour for $\alpha \neq 0$ and $\alpha = 0$, respectively. In the first case we have that $\sqrt{T}(\widehat{\alpha}_T^h - \alpha^h)$ has a nondegenerate limit distribution, whereas $T^{h/2}(\widehat{\alpha}_T^h - \alpha^h)$ has a proper limit distribution in the latter case. This change in the limiting behaviour is not fully captured by the factor $\sqrt{\sum_{t=1}^{T} y_{t-1}^2}/(\widetilde{\sigma}_u h \widehat{\alpha}_T^{h-1})$ that leads to a pivotal statistic only in the case $\alpha \neq 0$.

Such a situation is already known for $\widehat{\alpha}_T - \alpha$ for the critical case $|\alpha| = 1$. The three cases, $|\alpha| < 1$, $|\alpha| = 1$ and $|\alpha| > 1$, lead to very different limit distributions. Assume for a moment that $\sigma_u^2 = 1$. According to Theorem 4.3 of Anderson (1959),

$$\sqrt{T}(\widehat{\alpha}_T - \alpha) \xrightarrow{d} N(0, 1-\alpha^2)$$

holds for $|\alpha| < 1$. For one of the critical cases, $\alpha = 1$, White (1958) states that

$$T(\widehat{\alpha}_T - \alpha) \xrightarrow{d} \frac{1}{2}(W_1^2 - 1)/\int_0^1 W_s^2\, ds,$$

where W_s is a standard Wiener process. Heimann & Kreiss (1996) showed a similar result,

$$T(\widehat{\alpha}_T - \alpha) \xrightarrow{d} \frac{1}{2}(1 - W_1^2)/\int_0^1 W_s^2\, ds,$$

for $\alpha = -1$; see also Chan & Wei (1988, Section 3.2). Finally, under the additional conditions $y_0 \equiv 0$ and $u_t \sim N(0, \sigma_u^2)$, White (1958) showed for the case $|\alpha| > 1$ that

$$\frac{|\alpha|^T}{\alpha^2 - 1}(\widehat{\alpha}_T - \alpha) \xrightarrow{d} \eta,$$

where η has a Cauchy distribution. Hence, if we extend the parameter space and also allow for nonstationary processes the problem of incorrect CIs arises also in other situations than the simple one considered in detail in the foregoing. Of course, the problem also becomes more severe when higher order and higher dimensional processes are considered. In the following we will now focus exclusively on stationary univariate AR(1) processes. The reader should be aware, however, that similar problems also arise in other situations.

4.2 Confidence intervals based on the standard bootstrap

We consider the following bootstrap method:

1) Estimate $\widehat{\alpha}_T$ by least squares.

2) Generate bootstrap residuals $u_1^*, \ldots, u_T^*$ by randomly drawing with replacement from the set of estimated and recentered residuals, $\{\widehat{u}_1 - \bar{u}_., \ldots, \widehat{u}_T - \bar{u}_.\}$, where $\widehat{u}_t = y_t - \widehat{\alpha}_T y_{t-1}$, and $\bar{u}_. = T^{-1} \sum \widehat{u}_t$.

3) Set $y_0^* = y_0$ and construct bootstrap time series recursively by

$$y_t^* \; = \; \widehat{\alpha}_T y_{t-1}^* \, + \, u_t^*, \quad t = 1, \ldots, T. \tag{4.8}$$

4) Calculate a bootstrap version of the statistic of interest, in our case

$$\widehat{\alpha}_T^* \; = \; \sum_{t=1}^{T} y_t^* y_{t-1}^* / \sum_{t=1}^{T} (y_{t-1}^*)^2.$$

A slightly different method was proposed by Efron & Tibshirani (1986) who centered the original data $\{y_t\}$ first, rather than centering the estimated residuals. Such a scheme was also proposed by de Wet & van Wyk (1986) in the context of a linear regression model, where the errors were assumed to be generated by a linear AR(1) process.

Let $t_{\gamma/2}^*$ and $t_{(1-\gamma/2)}^*$ be the $\gamma/2$- and $(1-\gamma/2)$-quantiles of $\mathcal{L}((\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h \mid y_0, \ldots, y_T)$, respectively. Define

$$CI_2 \; = \; \left[\widehat{\alpha}_T^h \, - \, t_{(1-\gamma/2)}^*, \widehat{\alpha}_T^h \, - \, t_{\gamma/2}^* \right].$$

Since the limit distribution of $\sqrt{T}(\widehat{\alpha}_T^h - \alpha^h)$ depends in a continuous manner on the index α, as long as $\alpha \neq 0$ and $|\alpha| < 1$, it follows immediately that $\mathcal{L}(\sqrt{T}((\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h \mid y_0, \ldots, y_T))$ has the same limit distribution as $\mathcal{L}(\sqrt{T}(\widehat{\alpha}_T^h - \alpha^h))$. Hence,

$$P_\alpha \left(\alpha^h \in CI_2 \right) \longrightarrow 1 \, - \, \gamma, \tag{4.9}$$

for all $\alpha \neq 0$, $|\alpha| < 1$.

Now we suppose that $\alpha = 0$. From Theorem 4.3 of Anderson (1959) it follows that

$$P\left(\sqrt{T}(\widehat{\alpha}_T^* - \widehat{\alpha}_T) \leq x \mid y_0, \ldots, y_T\right) - F_{N(0,1)}\left(x/\sqrt{1-\widehat{\alpha}_T^2}\right) = o_P(1).$$

Integrating out the $y_0, \ldots, y_T$ shows that the unconditional distribution of $\sqrt{T}(\widehat{\alpha}_T^* - \widehat{\alpha}_T)$ converges to a standard normal distribution. Moreover, $\widehat{\alpha}_T^* - \widehat{\alpha}_T$ is asymptotically independent of $\widehat{\alpha}_T - \alpha$, which implies that

$$\sqrt{T}\begin{pmatrix} \widehat{\alpha}_T^* - \widehat{\alpha}_T \\ \widehat{\alpha}_T - \alpha \end{pmatrix} \stackrel{d}{\longrightarrow} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1-\alpha^2 & 0 \\ 0 & 1-\alpha^2 \end{pmatrix}\right).$$

Using

$$(\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h = \sum_{k=0}^{h-1} \binom{h}{k} \widehat{\alpha}_T^k (\widehat{\alpha}_T^* - \widehat{\alpha}_T)^{h-k}$$

we obtain that

$$T^{h/2}((\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h) \stackrel{d}{\longrightarrow} \sum_{k=0}^{h-1} \binom{h}{k} Z_1^k Z_2^{h-k}, \tag{4.10}$$

where $Z_1, Z_2 \sim N(0, 1-\alpha^2)$ are independent. In contrast, for the present case where $\alpha = 0$, we have that

$$T^{h/2}(\widehat{\alpha}_T^h - \alpha^h) \stackrel{d}{\longrightarrow} Z^h \tag{4.11}$$

with $Z \sim N(0, 1-\alpha^2)$. However, to show that the bootstrap CI has the correct asymptotic confidence level we need the result

$$P\left(T^{h/2}(\widehat{\alpha}_T^h - \alpha^h) \leq x\right) - P\left(T^{h/2}((\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h) \leq x \mid y_0, \ldots, y_T\right) = o_P(1).$$

This would at least require that the distribution of $\widehat{\alpha}_T^h - \alpha^h$ is approximated by the *unconditional* distribution of $(\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h$, that is

$$P\left(T^{h/2}(\widehat{\alpha}_T^h - \alpha^h) \leq x\right) - P\left(T^{h/2}((\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h) \leq x\right) = o(1).$$

In view of (4.10) and (4.11) this obviously does not hold. Moreover, it is clear that usual small sample corrections which aim at reducing the bias do not help in this context.

It is well-known that a general statistic μ_T can be better approximated by the bootstrap if it depends to a lesser extent on the unknown distribution that governs the data generating process. For example, concerning the sample mean of i.i.d. random variables it is well known that studentizing leads to a better rate of approximation by the bootstrap; see Hall (1988).

Therefore, we use the statistic $(\widehat{\alpha}_T^h - \alpha^h)/\sqrt{\widehat{\mathrm{var}}(\widehat{\alpha}_T^h)}$ as a basis for the construction of a confidence interval, and determine a bootstrap quantile from the statistic $((\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h)/\sqrt{\widehat{\mathrm{var}}((\widehat{\alpha}_T^*)^h)}$. The variances were estimated by the bootstrap method, that is

$$\widehat{\mathrm{var}}(\widehat{\alpha}_T^h) = \frac{1}{B^*}\sum_{i=1}^{B^*}(\widehat{\alpha}_T^{*,i})^{2h} - \left[\frac{1}{B^*}\sum_{i=1}^{B^*}(\widehat{\alpha}_T^{*,i})^h\right]^2$$

and

$$\widehat{\mathrm{var}}((\widehat{\alpha}_T^*)^h) = \frac{1}{B^{**}}\sum_{i=1}^{B^{**}}(\widehat{\alpha}_T^{**,i})^{2h} - \left[\frac{1}{B^{**}}\sum_{i=1}^{B^{**}}(\widehat{\alpha}_T^{**,i})^h\right]^2,$$

where B^* and B^{**} are the respective numbers of bootstrap replications. Note in particular that $\widehat{\alpha}_T^{**,i}$ is obtained by a double bootstrap, that is pseudo-data are generated according to a process with the parameter $\widehat{\alpha}_T^*$.

Let $t_{\gamma/2}^{**}$ and $t_{(1-\gamma/2)}^{**}$ be the $\gamma/2$- and $(1-\gamma/2)$-quantiles, respectively, of $\mathcal{L}\left([(\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h]/\sqrt{\widehat{\mathrm{var}}((\widehat{\alpha}_T^*)^h)} \mid y_0, \ldots, y_T\right)$. Based on our studentized statistics, we obtain the interval

$$CI_3 = \left[\widehat{\alpha}_T^h - t_{(1-\gamma/2)}^{**}\sqrt{\widehat{\mathrm{var}}(\widehat{\alpha}_T^h)},\ \widehat{\alpha}_T^h - t_{\gamma/2}^{**}\sqrt{\widehat{\mathrm{var}}(\widehat{\alpha}_T^h)}\right].$$

However, although studentizing improves the accuracy of the bootstrap in many "regular" cases, we do not believe that it helps in our context. Since, for $\alpha = 0$ and $h > 1$, the distributions of $(\widehat{\alpha}_T^h - \alpha^h)$ and $((\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h)$ are totally different, one cannot expect that those of $(\widehat{\alpha}_T^h - \alpha^h)/\sqrt{\widehat{\mathrm{var}}(\widehat{\alpha}_T^h)}$ and $((\widehat{\alpha}_T^*)^h - \widehat{\alpha}_T^h)/\sqrt{\widehat{\mathrm{var}}((\widehat{\alpha}_T^*)^h)}$ coincide asymptotically. In the simulations reported in Section 5 we will take a closer look at the performance of the latter bootstrap CI.

In the following we will present some modifications of the bootstrap which may be used to get asymptotically correct CIs for the presently considered simple AR(1) case.

4.3 Improved confidence intervals based on a superefficient estimator

The main reason why the standard bootstrap fails at the point $\alpha = 0$ is that $P_{\alpha_T}\left(T^{h/2}(\widehat{\alpha}_T^h - \alpha_T^h) \le x\right)$ remains different from $P_0\left(T^{h/2}(\widehat{\alpha}_T^h - \alpha^h) \le x\right)$, even if α_T tends to 0 with the rate $T^{-1/2}$.

Since $\widehat{\alpha}_T$ converges to the true value just with this rate, the bootstrap is not able to recognize the presence of the case $\alpha = 0$.

A well-known remedy to such problems with singularities in the limit distribution is the use of a so-called superefficient estimator that converges at a faster rate just at these critical points in the parameter space. Datta (1995) used this idea to devise a bootstrap for AR(1) processes that estimates $\mathcal{L}((\mathrm{var}(\widehat{\alpha}_T))^{-1/2}(\widehat{\alpha}_T - \alpha))$ consistently for all $\alpha \in \mathbf{R}$.

Whereas Datta (1995) used an estimator that is superefficient at $\alpha = \pm 1$, we need this property for $\alpha = 0$. Let $\{c_T\}$ be any sequence satisfying $c_T \to 0$ and $T^{1/2}c_T \to \infty$ as $T \to \infty$. Then the threshold estimator

$$\widetilde{\alpha}_T = \begin{cases} \widehat{\alpha}_T, & \text{if } |\widehat{\alpha}_T| > c_T \\ 0 & \text{otherwise} \end{cases} \tag{4.12}$$

is superefficient at $\alpha = 0$, that is $\widetilde{\alpha}_T$ converges with a faster rate than $T^{-1/2}$ to the true value. This estimator allows to switch between the two cases, $\alpha = 0$ and $\alpha \neq 0$. We define the following quantity:

$$\begin{aligned} S_T &= \frac{\sqrt{\sum_{t=1}^T y_{t-1}^2}}{\widetilde{\sigma}_u[\widehat{\alpha}_T^{h-1} + (h-1)\widetilde{\alpha}_T^{h-1}]}(\widehat{\alpha}_T^h - \alpha^h) \\ &= \begin{cases} \frac{\sqrt{\sum_{t=1}^T y_{t-1}^2}}{\widetilde{\sigma}_u[h\widehat{\alpha}_T^{h-1} + o_P(1)]}(\widehat{\alpha}_T^h - \alpha^h), & \text{if } \alpha \neq 0 \\ \frac{\sqrt{\sum_{t=1}^T y_{t-1}^2}}{\widetilde{\sigma}_u[\widehat{\alpha}_T^{h-1} + o_P(T^{-1/2})]}(\widehat{\alpha}_T^h - \alpha^h), & \text{if } \alpha = 0 \end{cases} . \end{aligned} \tag{4.13}$$

Hence, it can be seen easily that

$$S_T \xrightarrow{d} N(0,1) \tag{4.14}$$

for all $|\alpha| < 1$. Therefore,

$$CI_4 = \left[\widehat{\alpha}_T^h - \frac{\widetilde{\sigma}_u|\widehat{\alpha}_T^{h-1} + (h-1)\widetilde{\alpha}_T^{h-1}|}{\sqrt{\sum_{t=1}^T y_{t-1}^2}}c_{(1-\gamma/2)},\ \widehat{\alpha}_T^h + \frac{\widetilde{\sigma}_u|\widehat{\alpha}_T^{h-1} + (h-1)\widetilde{\alpha}_T^{h-1}|}{\sqrt{\sum_{t=1}^T y_{t-1}^2}}c_{(1-\gamma/2)}\right] \tag{4.15}$$

is a confidence interval for α^h to an asymptotic level $1-\gamma$, that is,

$$P_\alpha\left(\alpha^h \in CI_4\right) \longrightarrow 1 - \gamma \quad \text{for all } |\alpha| < 1. \tag{4.16}$$

Usually, results are formulated in a pointwise (in α) manner as in (4.16). However, a closer look at the proposed procedure indicates that there is no uniformity in α. We conjecture that (4.16) fails, if we consider instead of any *fixed* α a sequence $\{\alpha_T\}$ tending to zero at the rate $T^{-1/2}$.

Rather than relying on the asymptotic distribution, we could also use a bootstrap approximation of $\mathcal{L}(\widehat{\alpha}_T^h - \alpha^h)$ in connection with the above superefficient estimator $\widetilde{\alpha}_T$. This was done by Datta (1995) for estimating the distribution of $\widehat{\alpha}_T - \alpha$ around $|\alpha| = 1$. We believe, however, that the same pointwise result as above is obtained in that case with the same problem concerning uniformity in α. Of course, for practical purposes one may be satisfied with pointwise convergence. Even then it will be difficult to generalize this approach to higher order and higher dimensional processes because it requires that care has to be taken for every possible singularity point of the asymptotic distribution. In general this may be a difficult or impossible task. The use of superefficient estimators can solve problems with different limit distributions at *known isolated* points in the parameter space. Since any estimator can only be superefficient on sets with measure 0, it is impossible to apply such a strategy in the case of rapidly changing limit distributions, where these changes occur at *unknown* points in the parameter space. Therefore other procedures have been considered which do not require the user to identify all singularity points prior to using the bootstrap. One such procedure will be described in the following subsection.

4.4 Subsampling

It can be seen from the calculations in Subsection 4.2 that the standard bootstrap would have been consistent in the case $\alpha = 0$, if the resampling scheme (4.8) were based on an estimator $\widehat{\widehat{\alpha}}_T$ with $E_0(\widehat{\widehat{\alpha}}_T - \alpha)^2 = o(T^{-1})$, where E_0 denotes the expectation evaluated under $\alpha = 0$. This would imply that the resampling scheme adjusts with a sufficiently fast rate to the change of the distribution of $(\text{var}(\widehat{\alpha}_T^h))^{-1/2}(\widehat{\alpha}_T^h - \alpha^h)$ from $\alpha = 0$ to $\alpha \neq 0$.

Subsampling, that is, resampling fewer than T observations, is a relatively new technique that aims at improving the relation between the rate of convergence of the bootstrap version of the estimator and the rate at which the parameter that controls the data generating process in the bootstrap world converges. Surveys on this technique are given by Bertail, Politis & Romano (1995), in the discussion to Li & Maddala (1996), and by Bickel, Götze & van Zwet (1997).

Subsampling is relatively straightforward, if the rate of convergence is constant over the whole parameter space and if only the shape of the corresponding limit distributions is different. Although recent work of Bertail et al. (1995) also allows for the case of different rates of convergence to be estimated separately, we try to avoid these complications by multiplying the statistic of interest, $\widehat{\alpha}_T^h - \alpha^h$, with an appropriate normalizing factor. This was also done by Heimann & Kreiss (1996) in the case of estimating the distribution of $(\mathrm{var}(\widehat{\alpha}_T))^{-1/2}(\widehat{\alpha}_T - \alpha)$ around $|\alpha| = 1$.

We consider the statistic

$$S_T \;=\; \frac{\sqrt{\sum_{t=1}^T y_{t-1}^2}}{\widehat{\alpha}_T^{h-1}}(\widehat{\alpha}_T^h - \alpha^h). \tag{4.17}$$

It may be seen from (4.4) and (4.7) that S_T has a nondegenerate limit distribution for all $|\alpha| < 1$. Let $N = N(T)$ be the size of the bootstrap sample, where $N(T) \to \infty$ and $N(T)/T \to 0$ as $T \to \infty$. Our bootstrap mimics $S_{N(T)}$ rather than S_T. Nevertheless, this is asymptotically correct, since both quantities have the same limit distribution. There are two obvious possibilities for bootstrapping:

1) A model-based method

   a) draw bootstrap residuals $u_1^*, \ldots, u_N^*$ randomly with replacement from $\{\widehat{u}_1 - \bar{u}_., \ldots, \widehat{u}_T - \bar{u}_.\}$, where $\widehat{u}_t = y_t - \widehat{\alpha}_T y_{t-1}$,

   b) set $y_0^* \equiv y_0$ and define recursively

$$y_t^* \;=\; \widehat{\alpha}_T y_{t-1}^* \;+\; u_t^*, \quad t = 1, \ldots, N,$$

   c) calculate

$$\widehat{\alpha}_N^* \;=\; \sum_{t=1}^N y_t^* y_{t-1}^* / \sum_{t=1}^N (y_{t-1}^*)^2.$$

2) A model-free method

   a) take all $T - N + 1$ blocks of $N + 1$ consecutive observations from $y_0, y_1, \ldots, y_T$,

   b) calculate, for each block $y_s^*, \ldots, y_{s+N}^*$, $s = 0, \ldots, T - N$, the bootstrap estimate

$$\widehat{\alpha}_{N,t}^* \;=\; \sum_{t=1}^N y_{s+t}^* y_{s+t-1}^* / \sum_{t=1}^N (y_{s+t-1}^*)^2.$$

Since we assume to know the structure of the data generating process, we opt for the first method. It was shown in Heimann & Kreiss (1996) that $\sqrt{\sum_{t=1}^{N}(y_{t-1}^*)^2}(\widehat{\alpha}_N^* - \widehat{\alpha}_T)$ converges weakly to the same limit as $\sqrt{\sum_{t=1}^{T} y_{t-1}^2}(\widehat{\alpha}_T - \alpha)$, for all $\alpha \in \mathbf{R}$. For $\alpha \neq 0$, we have that

$$\begin{aligned}
\frac{\sqrt{\sum_{t=1}^{N}(y_{t-1}^*)^2}}{(\widehat{\alpha}_N^*)^{h-1}}\left((\widehat{\alpha}_N^*)^h - \widehat{\alpha}_T^h\right) &= \sum_{k=0}^{h-1}\binom{h}{k}\widehat{\alpha}_T^k \frac{\sqrt{\sum_{t=1}^{N}(y_{t-1}^*)^2}}{(\widehat{\alpha}_N^*)^{h-1}}\left(\widehat{\alpha}_N^* - \widehat{\alpha}_T\right)^{h-k} \\
&= h\frac{\widehat{\alpha}_T^{h-1}}{(\widehat{\alpha}_N^*)^{h-1}}\sqrt{\sum_{t=1}^{N}(y_{t-1}^*)^2}\left(\widehat{\alpha}_N^* - \widehat{\alpha}_T\right) + o_P(1) \\
&= h\sqrt{\sum_{t=1}^{N}(y_{t-1}^*)^2}\left(\widehat{\alpha}_N^* - \widehat{\alpha}_T\right) + o_P(1),
\end{aligned}$$

and, for $\alpha = 0$,

$$\begin{aligned}
\frac{\sqrt{\sum_{t=1}^{N}(y_{t-1}^*)^2}}{(\widehat{\alpha}_N^*)^{h-1}}\left((\widehat{\alpha}_N^*)^h - \widehat{\alpha}_T^h\right) &= \sum_{k=0}^{h-1}\binom{h}{k}\widehat{\alpha}_T^k \frac{\sqrt{\sum_{t=1}^{N}(y_{t-1}^*)^2}}{(\widehat{\alpha}_N^*)^{h-1}}\left(\widehat{\alpha}_N^* - \widehat{\alpha}_T\right)^{h-k} \\
&= \sqrt{\sum_{t=1}^{N}(y_{t-1}^*)^2}\frac{\left(\widehat{\alpha}_N^* - \widehat{\alpha}_T\right)^h}{(\widehat{\alpha}_N^*)^{h-1}} + o_P(1) \\
&= \sqrt{\sum_{t=1}^{N}(y_{t-1}^*)^2}\left(\widehat{\alpha}_N^* - \widehat{\alpha}_T\right) + o_P(1).
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\frac{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}{\widehat{\alpha}_T^{h-1}}(\widehat{\alpha}_T^h - \alpha^h) &= \frac{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}{\widehat{\alpha}_T^{h-1}}\sum_{k=0}^{h-1}\binom{h}{k}\alpha^k(\widehat{\alpha}_T - \alpha)^{h-k} \\
&= \begin{cases} h\sqrt{\sum_{t=1}^{T} y_{t-1}^2}(\widehat{\alpha}_T - \alpha) + o_P(1), & \text{if } \alpha \neq 0 \\ \sqrt{\sum_{t=1}^{T} y_{t-1}^2}(\widehat{\alpha}_T - \alpha), & \text{if } \alpha = 0 \end{cases}.
\end{aligned}$$

Comparing the right-hand sides of the latter three displayed formulas we see that the (conditional) distribution of

$$\frac{\sqrt{\sum_{t=1}^{N}(y_{t-1}^*)^2}}{(\widehat{\alpha}_N^*)^{h-1}}\left((\widehat{\alpha}_N^*)^h - \widehat{\alpha}_T^h\right)$$

approximates that of

$$\frac{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}{\widehat{\alpha}_T^{h-1}}(\widehat{\alpha}_T^h - \alpha^h)$$

for all $|\alpha| < 1$.

Let t^*_γ be the $(1-\gamma)$-quantile of the distribution of $\left|\frac{\sqrt{\sum_{t=1}^{N}(y^*_{t-1})^2}}{(\widehat{\alpha}^*_N)^{h-1}}\left((\widehat{\alpha}^*_N)^h - \widehat{\alpha}^h_T\right)\right|$. Then

$$P\left(\left|\frac{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}{\widehat{\alpha}_T^{h-1}}(\widehat{\alpha}^h_T - \alpha^h)\right| < t^*_{\gamma/2}\right) \longrightarrow 1-\gamma,$$

which implies that

$$CI_5 = \left[\widehat{\alpha}^h_T - t^*_{\gamma/2}\frac{|\widehat{\alpha}_T|^{h-1}}{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}},\ \ \widehat{\alpha}^h_T + t^*_{\gamma/2}\frac{|\widehat{\alpha}_T|^{h-1}}{\sqrt{\sum_{t=1}^{T} y_{t-1}^2}}\right]$$

is an asymptotic $(1-\gamma)$-confidence interval for $|\alpha| < 1$.

In more general situations where higher order and higher dimensional processes are considered it will not be easy to find a suitable normalization of estimated impulse responses analogous to the factor $\sqrt{\sum_{t=1}^{T} y_{t-1}^2}/\widehat{\alpha}_T^{h-1}$ in our simple case which guarantees a constant rate of convergence. In the next subsection we will present a subsampling approach which is theoretically suitable in such a case. It may be computationally quite demanding, however.

4.5 Subsampling with estimated rate of convergence

The problem with the subsampling procedure of the previous subsection is that it may be difficult to find a quantity with constant rate of convergence in all of the feasible parameter space. For this situation, Bertail et al. (1997) proposed to estimate the rate of convergence, τ_T, say.

The convergence rate is being estimated using two subsampling distributions based on the subsampling sizes $N_1 = N_1(T)$ and $N_2 = N_2(T)$. If $N_1, N_2 \to \infty$ as $T \to \infty$ we get

$$\begin{aligned} P(\tau_{N_1}((\widehat{\alpha}^*_{N_1})^h - \widehat{\alpha}^h_T) \le t|\underline{Y}) &= P(\tau_{N_2}((\widehat{\alpha}^*_{N_2})^h - \widehat{\alpha}^h_T) \le t|\underline{Y}) + o_P(1) \\ &= \mathcal{L}_\infty(t) + o_P(1), \end{aligned} \tag{4.18}$$

where $\underline{Y}$ denotes the sample $(y_0, \dots, y_T)$ and $\mathcal{L}_\infty(t)$ the limit distribution of $\tau_N((\hat{\alpha}_N)^h - \widehat{\alpha}^h_T)$.

The inverse of these cumulative distribution functions is needed in the form

$$F^{-1}_{\tau_{N_i}((\widehat{\alpha}^*_{N_i})^h - \widehat{\alpha}^h_T)}(t|\underline{Y}) = \tau_{N_i} F^{-1}_{((\widehat{\alpha}^*_{N_i})^h - \widehat{\alpha}^h_T)}(t|\underline{Y}), \quad i = 1, 2, \tag{4.19}$$

where t is a continuity point of $\mathcal{L}_\infty(t)$. From equations (4.18) and (4.19) we get

$$\tau_{N_1} F^{-1}_{((\widehat{\alpha}^*_{N_1})^h - \widehat{\alpha}^h_T)}(t|\underline{Y}) = \tau_{N_2} F^{-1}_{((\widehat{\alpha}^*_{N_2})^h - \widehat{\alpha}^h_T)}(t|\underline{Y}) + o_P(1)$$

or, equivalently, assuming $\tau_T = T^{\delta}$ and $N_i = T^{\beta_i}$, $1 > \beta_1 > \beta_2 > 0$,

$$\delta = \frac{\log(F^{-1}_{((\widehat{\alpha}^*_{N_2})^h - \widehat{\alpha}^h_T)}(t|\underline{Y})) - \log(F^{-1}_{((\widehat{\alpha}^*_{N_1})^h - \widehat{\alpha}^h_T)}(t|\underline{Y}))}{(\log(N_1) - \log(N_2))} + o_P(\log(T)^{-1}).$$

Therefore, we estimate δ by

$$\widehat{\delta} = \frac{\log(F^{-1}_{((\widehat{\alpha}^*_{N_2})^h - \widehat{\alpha}^h_T)}(t|\underline{Y})) - \log(F^{-1}_{((\widehat{\alpha}^*_{N_1})^h - \widehat{\alpha}^h_T)}(t|\underline{Y}))}{(\log(N_1) - \log(N_2))}.$$

It follows from the consistency established in (4.18) that $\widehat{\delta}$ is a consistent estimate of δ. In this paper we have taken the mean over several points t_j, $j = 1, \dots, J$. Then we use the estimator $\widehat{\tau}_T = T^{\widehat{\delta}}$.

Now it is possible to proceed with constructing confidence intervals using subsampling as introduced in Section 4.4. The difference in both methods is found in the norming factor for the statistic $((\widehat{\alpha}^*_N)^h - \widehat{\alpha}^h_T)$. Here, the norming factor is estimated for each α^h separately, whereas the method in Section 4.4 uses algebraic manipulation to handle the problem of different convergence rates. This implies that each new investigated statistic demands new analytical work for this manipulation.

The unknown distribution of interest, $\mathcal{L}(\widehat{\tau}_T(\widehat{\alpha}^h_T - \alpha^h))$, is approximated by

$$\mathcal{L}\left(\widehat{\tau}_{N(T)}((\widehat{\alpha}^*_{N(T)})^h - \widehat{\alpha}^h_T)|\underline{Y}\right), \tag{4.20}$$

with subsample size $N(T)$. Let $t^*_{\gamma/2}$ and $t^*_{1-\gamma/2}$ be the $\gamma/2$ and $1-\gamma/2$ quantiles of (4.20), respectively. Then

$$P\left(t^*_{\gamma/2} < \widehat{\tau}_T(\widehat{\alpha}^h_T - \alpha^h) < t^*_{1-\gamma/2}\right) \rightarrow (1-\gamma).$$

Hence,

$$CI_6 = \left[\widehat{\alpha}^h_T - \frac{t^*_{1-\gamma/2}}{\widehat{\tau}_T},\ \ \widehat{\alpha}^h_T - \frac{t^*_{\gamma/2}}{\widehat{\tau}_T}\right]$$

is a confidence interval which has asymptotically the correct coverage probability of $(1-\gamma)$.

4.6 Nearly exact confidence intervals

Our fifth method of constructing confidence intervals uses ideas of Sims & Zha (1994) and may be motivated as follows. Assume for a moment that we know the distribution of the innovations u_t exactly. Then we can calculate, for each fixed α, the (hypothetical) distribution of $\widehat{\alpha}_T^h$. Let $t_{\alpha,\gamma/2}$ and $t_{\alpha,(1-\gamma/2)}$ be the $\gamma/2$- and $(1-\gamma/2)$-quantile of the corresponding distribution. Define

$$\widetilde{CI}_7 \;=\; \left\{\alpha^h \;|\;\; \widehat{\alpha}_T^h \in [t_{\alpha,\gamma/2}, t_{\alpha,(1-\gamma/2)}]\right\}.$$

By construction,

$$P_\alpha\left(\alpha^h \in \widetilde{CI}_7\right) \;=\; P_\alpha\left(\widehat{\alpha}_T^h \in [t_{\alpha,\gamma/2}, t_{\alpha,(1-\gamma/2)}]\right) \;=\; 1 \;-\; \gamma,$$

that is, $\widetilde{CI}_7$ is an *exact* confidence set for α^h, for all values of $\alpha \in \mathbf{R}$. This approach was proposed by Andrews (1993, Section 4) in the case of a known distribution of the innovations.

Since the distribution of the innovations u_t is usually unknown, we propose to estimate it by the bootstrap. Let $u_1^*, \ldots, u_T^*$ be drawn with replacement from $\{\widehat{u}_1 - \bar{u}_\cdot, \ldots, \widehat{u}_T - \bar{u}_\cdot\}$, where $\widehat{u}_t = y_t - \widehat{\alpha}_T y_{t-1}$, as before. For each value of α we generate (hypothetical) bootstrap processes by setting $y_0^{\alpha,*} \equiv y_0$ and

$$y_t^{\alpha,*} \;=\; \alpha y_{t-1}^{\alpha,*} \;+\; u_t^*, \quad t = 1, \ldots, T,$$

and corresponding estimators $\widehat{\alpha}_T^{\alpha,*} = \sum_{t=1}^T y_t^{\alpha,*} y_{t-1}^{\alpha,*} / \sum_{t=1}^T (y_{t-1}^{\alpha,*})^2$. Let $t^*_{\alpha,\gamma/2}$ and $t^*_{\alpha,(1-\gamma/2)}$ be the $\gamma/2$- and $(1-\gamma/2)$-quantiles of $\mathcal{L}((\widehat{\alpha}_T^{\alpha,*})^h)$, respectively. According to the theoretical set $\widetilde{CI}_7$ above, we construct an asymptotic confidence set as

$$CI_7 \;=\; \left\{\alpha^h \;|\;\; \widehat{\alpha}_T^h \in [t^*_{\alpha,\gamma/2}, t^*_{\alpha,(1-\gamma)/2}]\right\}. \tag{4.21}$$

Although we do not have a rigorous proof, we conjecture that the set CI_7 is an interval under nonrestrictive regularity conditions. To underline the uniformity of the consistency of CI_7, we formulate a result for the supremum of the error in coverage probability over a set of α's containing the critical region around 0.

Theorem 1

Let $c < 1$ and suppose that $E|u_t|^M < \infty$ for all $M < \infty$. Then

$$\sup_{\alpha \in [-c,c]} \left\{\left|P_\alpha(\alpha^h \in CI_7) \;-\; (1-\gamma)\right|\right\} \;=\; o(1).$$

□

This result is proven in the Appendix. Moreover, it is possible to prove the consistency of this method in a pointwise manner for all $\alpha \in \mathbf{R}$. Again this method becomes rather complicated and computer intensive in more general situations with higher order and higher dimensional processes. However, in principle it can be extended. In the next section we report some simulations which illustrate the small sample aspects of the asymptotic results of the present section.

5 Simulations

The simulation study compares the performance of CI_1 through CI_7 of impulse responses 1, 2, 10, and 20 periods after an impulse hits the system. This is done by estimating the real coverage probability, $\hat{p}$, of 95% confidence intervals ($p = .95$). Additionally the length of the confidence intervals, l, is evaluated.

Data was artificially generated by the univariate AR(1) process

$$y_t = \alpha y_{t-1} + u_t, \qquad t = 1, \ldots, T, \quad y_0 = 0, \quad u_t \sim N(0,1), \quad \text{and} \quad \alpha = 0, .2, .5 \qquad (5.1)$$

We considered two sample sizes $T = 100$ and $T = 1,000$. The first is a common sample size when analysing macro data. The second is thought to be a "step" away from small sample size. The idea is to see how the described methods react when the sample size increases. For each α, $M = 1,000$ Monte Carlo (MC) replications were performed. Assume $\mathrm{E}(\widehat{p}) = p$. Then, the standard error of $\widehat{p}$ after 1,000 MC replications is approximated by $\sqrt{p(1-p)/M} \approx 0.007$. In order to control variability the error vector $U = (u_1, \ldots, u_T)'$ in (5.1) for the m-th MC replication is identical for each α. Furthermore, in the m-th MC replication the bootstrap based methods resampled the same indices of $\widehat{U}$ in order to get U^*. Thus, the methods are fully accountable for the different performance.

The quantiles in the bootstrap based intervals $(CI_2, CI_3, CI_5, CI_6, CI_7)$ were estimated using two–sided symmetric quantiles. The quantile t_γ is thus the $(\gamma \cdot B)$th element of the bootstrap realizations in ascending order of the unknown distribution. A distribution was constructed using $B = 1,000$ bootstrap replications. We also tried another approach. Choosing two quantiles $t_1 < t_2$ of the estimated distribution $\mathcal{L}(\cdot)$ such that it minimizes $t_2 - t_1$ subject to the constraint $\mathcal{L}(t_2) - \mathcal{L}(t_1) = 1 - \gamma$ (highest density). The simulations showed that the latter seems to have a slight edge over the former in terms of interval length. The gains were highest in the case of $\alpha = 0$. This might be explained with the skewed distribution of $(\widehat{\alpha}_N^*)^h - \widehat{\alpha}_T^h$. Finally, it does not make much difference in terms of coverage frequency whether symmetric or highest density quantiles are used in setting up the CIs. The results reported for CI_2, CI_3, CI_5, CI_6, and CI_7 are therefore based on two–sided symmetric quantiles.

The threshold c_T for the superefficient estimator in Section 4.3 is chosen to be

$$c_T = \frac{\sqrt{2\log T}}{\sqrt{\sum_{t=1}^T y_t^2}}.$$

While the denominator is just a scaling factor the particular choice of the numerator, $\sqrt{2\log T}$, is motivated by the fact that a standard normal random variable in absolute value exceeds this bound with a probability of T^{-1}, which we consider as a sufficient value. The actual thresholds are given in Table 7.

The subsample length in method 5 is $N(T) = 90$ for $T = 100$, and $N(T) = 900$ for $T = 1,000$. Finding a suitable subsample size is actually a difficult task. Politis & Romano (1994) found the order $N(T) \asymp T^{2/3}$ to be optimal on the basis of second-order asymptotic theory. In our simulations, we tried different values for $N(T)$ including the proposed ones. The results showed that the estimated coverage probability does not change much when changing $N(T)$. Moreover, for $T = 100$ we found that "small" $N(T)$ (e.g. 25, 30) resulted in larger confidence intervals than rather "large" values for $N(T)$ (e.g. 80,90). Whereas this phenomenon was not really observable for $\alpha = 0$, the enlargement of the confidence intervals became quite substantial for $\alpha = 0.5$. Therefore, we decided to choose $N(T) = 90$ in the current setting. For the estimation of the rate of convergence, method 6 uses many different subsample sizes $N(T)$ ($N(100) = 90, 80, 50, 30$, $N(1,000) = 900, 800, 500, 300$) and quantiles $t = .95, .93, .91, .89, .87$. The results are reported in Table 8 where it is seen that although the theoretical convergence rate for $\alpha = .2$ and $\alpha = .5$ is the same, the estimated rate for the former value of α is an intermediate value between the rate for $\alpha = 0$ and $\alpha = .5$. The confidence intervals are constructed with subsample sizes $N(T)$=50 and 500 for $T = 100$ and 1,000, respectively. Again we tried several subsample sizes and found that the chosen size performs best in the current setting.

This paper does not investigate the quite popular percentile method. It basically takes the $t^*_{\gamma/2}$ and $t^*_{1-\gamma/2}$ quantiles of the bootstrap distribution $\mathcal{L}((\widehat{\alpha}^*_T)^h | y_1, \ldots, y_T)$ as the lower and upper confidence bound, respectively, for a $(1-\gamma)$ confidence interval of α. The reason is that we cannot find a sound derivation of its asymptotic coverage performance. However, it seems to perform well when $\mathcal{L}((\widehat{\alpha}^*_T)^h | y_1, \ldots, y_T)$ is symmetric. In our example we would face a complete failure for the case $\alpha = 0$ and h even. In a simulation we got a real

Table 1: Estimated coverage probability ($\widehat{p}$) and average length (l) of estimated 95% CIs for AR(1) process with $T = 100$ and $\alpha = 0$

	α		α^2		α^{10}		α^{20}	
	$\hat{p}$	l	$\hat{p}$	l	$\hat{p}$	l $\times 10^{-5}$	$\hat{p}$	l $\times 10^{-8}$
Asymptotic (CI_1)	.951	.392	1.000	.063	1.000	.146	1.000	.006
Standard bootstrap (CI_2)	.939	.388	.980	.077	.976	1.237	.976	.327
Studentized bootstrap (CI_3)	.957	.401	.978	.051	.976	1.200	.974	3.312
Superefficient est. (CI_4)	.951	.392	.956	.032	.956	.113	.956	.006
Subsampling (CI_5)	.949	.385	.997	.132	1.000	$*$	1.000	$**$
Subsampling with $\widehat{\tau}_T$ (CI_6)	.925	.380	.999	.062	1.000	.761	1.000	.246
Indirect (CI_7)	.956	.399	.951	.078	.951	2.630	.951	2.833

* The average length was estimated > 2

** The average length was estimated > 1500

coverage probability of 0.0%. This can be explained with the highly asymmetric shape of $\mathcal{L}((\widehat{\alpha}^*_T)^h|y_1, \ldots, y_T)$ in this case.

Table 2: Estimated coverage probability ($\widehat{p}$) and average length (l) of estimated 95% CIs for AR(1) process with $T = 100$ and $\alpha = .2$

	α		α^2		α^{10}		α^{20}	
	$\hat{p}$	l	$\hat{p}$	l	$\hat{p}$	l $\times 10^{-4}$	$\hat{p}$	l $\times 10^{-6}$
Asymptotic (CI_1)	.955	.384	.885	.155	.672	.913	.618	.091
Standard bootstrap (CI_2)	.937	.381	.710	.146	.541	2.935	.519	.552
Studentized bootstrap (CI_3)	.955	.394	.791	.172	.779	4.738	.787	2.479
Superefficient est. (CI_4)	.955	.384	.760	.102	.366	.855	.292	.091
Subsampling (CI_5)	.952	.378	.908	.287	.963	*	.967	**
Subsampling with $\widehat{\tau}_T$ (CI_6)	.932	.374	.696	.123	.550	1.773	.526	.372
Indirect (CI_7)	.956	.391	.954	.163	.954	5.962	.954	3.482

* The average length was estimated > 4

** The average length was estimated > 5500

Table 3: Estimated coverage probability ($\widehat{p}$) and average length (l) of estimated 95% CIs for AR(1) process with $T = 100$ and $\alpha = 0.5$

	α		α^2		α^{10}		α^{20}	
	$\hat{p}$	l	$\hat{p}$	l	$\hat{p}$	l $\times 10^{-2}$	$\hat{p}$	l $\times 10^{-3}$
Asymptotic (CI_1)	.951	.340	.944	.334	.786	1.345	.708	.231
Standard bootstrap (CI_2)	.928	.342	.873	.316	.630	1.790	.561	.675
Studentized bootstrap (CI_3)	.950	.350	.975	.377	.972	3.340	.979	2.044
Superefficient est. (CI_4)	.951	.340	.944	.332	.786	1.345	.708	.231
Subsampling (CI_5)	.945	.334	.971	.374	1.000	17.712	1.000	*
Subsampling with $\widehat{\tau}_T$ (CI_6)	.916	.338	.883	.303	.623	1.223	.560	.365
Indirect (CI_7)	.953	.347	.953	.344	.953	3.248	.953	2.335

* The average length was estimated > 10

We consider a precise coverage probability as the most important feature of a confidence interval. However, from the point of view of the usefulness of the intervals, their average length is also an important factor. These two points are discussed in the following. Possible generalizations of the methods to multivariate and higher order autoregressive processes are discussed afterwards. Recall from the previous section that methods 4 to 7 are asymptotically correct for all α, whereas methods 1 to 3 are only correct if $\alpha \neq 0$.

The following observations emerge from Tables 1–6. Tables 1–3 represent the results for the sample size of $T = 100$ and Tables 4–6 for $T = 1,000$. First, all methods lead to nearly identical results in the case of $h = 1$, even for $\alpha = 0$. The indirect method is overall the best in terms of coverage. As expected it produces the nominal coverage level almost exactly for all cases considered. Although the average length of the CIs is in some cases a bit larger than that of other methods it is often comparable to the asymptotic CIs. Its main disadvantage then is the difficulty to extend it to higher dimensional and higher order cases.

For the critical case where $\alpha = 0$ the CIs based on standard asymptotics are clearly conservative for $h > 1$ and have a considerably larger coverage probability than the nominal 95%. The length of the intervals is surprisingly small given that we found in the asymptotic analysis that even further reductions of the length may be possible. In this respect the CI_1 intervals are just outperformed by the CIs based on the superefficient estimator which is overall clearly the best method for the case $\alpha = 0$. However, for nonzero α both the CIs based on standard asymptotics and on the superefficient estimator are problematic in the small sample context because their actual coverage level deteriorates substantially for growing α. This is true in particular for the latter CIs. For instance, the coverage frequency of CI_4 is only 29.2% for $\alpha = 0.2$ and $T = 100$ when confidence intervals for α^{20} are considered. Increasing the sample size to $T = 1,000$ leads in this case to a far better coverage of 75.2%. Further simulations showed that the real coverage becomes 94.0% when the sample size grows to $T = 100,000$.

The standard bootstrap leads to similar results as method 1. In the majority of cases, the performance of the bootstrap intervals is even slightly worse with respect to both the coverage probability and the length of the intervals. The coverage probability deteriorates again for $\alpha \neq 0$ and $h > 1$. This seems to be a small sample problem, however, which might be tackled by specific corrections; see, e.g., Kilian (1995, 1996) and Fachin & Bravetti

Table 4: Estimated coverage probability ($\widehat{p}$) and average length (l) of estimated 95% CIs for AR(1) process with $T = 1,000$ and $\alpha = 0$

	α		α^2		α^{10}		α^{20}	
	$\hat{p}$	l	$\hat{p}$	l $\times 10^{-3}$	$\hat{p}$	l $\times 10^{-10}$	$\hat{p}$	l $\times 10^{-18}$
Asymptotic (CI_1)	.949	.124	1.000	6.315	1.000	.145	1.000	.003
Standard bootstrap* (CI_2)	.940	.123	.980	8.178	.978	1.826	.978	.355
Studentized bootstrap* (CI_3)	.950	.126	.980	5.558	.978	1.151	.975	.585
Superefficient est. (CI_4)	.949	.124	.949	3.158	.949	.014	.949	$< .001$
Subsampling (CI_5)	.951	.124	.997	12.848	1.000	$**$	1.000	$***$
Subsampling with $\widehat{\tau}_T$ (CI_6)	.944	.122	1.000	6.471	1.000	1.201	1.000	.490
Indirect (CI_7)	.946	.124	.946	7.689	.946	2.646	.946	1.703

* Numbers reported based on 400 Monte–Carlo–Replications, $\sigma_{\widehat{p}} \approx 0.011$ for $p = .95$

** The average length was estimated $> 2 \times 10^{-5}$

*** The average length was estimated $> 5 \times 10^{-7}$

Table 5: Estimated coverage probability ($\widehat{p}$) and average length (l) of estimated 95% CIs for AR(1) process with $T = 1,000$ and $\alpha = 0.2$

	α		α^2		α^{10}		α^{20}	
	$\hat{p}$	l	$\hat{p}$	l $\times 10^{-2}$	$\hat{p}$	l $\times 10^{-6}$	$\hat{p}$	l $\times 10^{-11}$
Asymptotic (CI_1)	.953	.121	.939	4.865	.831	1.314	.752	.291
Standard bootstrap* (CI_2)	.940	.121	.893	4.792	.653	2.311	.548	1.299
Studentized bootstrap* (CI_3)	.940	.124	.965	5.197	.973	3.046	.983	2.513
Superefficient est. (CI_4)	.953	.121	.939	4.858	.831	1.314	.752	.291
Subsampling (CI_5)	.952	.121	.971	5.178	.998	5.063	1.000	7.340
Subsampling with $\widehat{\tau}_T$ (CI_6)	.950	.120	.924	4.657	.648	1.694	.582	.926
Indirect (CI_7)	.954	.122	.954	4.887	.954	2.614	.954	1.995

* Numbers reported based on 400 Monte–Carlo–Replications, $\sigma_{\widehat{p}} \approx 0.011$ for $p = .95$

Table 6: Estimated coverage probability ($\widehat{p}$) and average length (l) of estimated 95% CIs for AR(1) process with $T = 1,000$ and $\alpha = 0.5$

	α		α^2		α^{10}		α^{20}	
	$\hat{p}$	l	$\hat{p}$	l	$\hat{p}$	l $\times 10^{-3}$	$\hat{p}$	l $\times 10^{-5}$
Asymptotic (CI_1)	.955	.107	.956	.107	.908	2.273	.854	.624
Standard bootstrap* (CI_2)	.943	.107	.938	.106	.810	2.428	.710	.916
Studentized bootstrap* (CI_3)	.953	.109	.950	.110	.973	2.810	.978	1.217
Superefficient est. (CI_4)	.955	.107	.956	.107	.908	2.273	.854	.624
Subsampling (CI_5)	.958	.107	.959	.108	.967	2.883	.977	1.385
Subsampling with $\widehat{\tau}_T$ (CI_6)	.948	.106	.949	.105	.841	2.153	.713	.690
Indirect (CI_7)	.955	.108	.955	.107	.955	2.616	.955	1.070

* Numbers reported based on 400 Monte–Carlo–Replications, $\sigma_{\widehat{p}} \approx 0.011$ for $p = .95$

Table 7: Average threshold c_T of the superefficient estimator for sample sizes $T = 100$, and 1,000. Numbers in paranthesis are standard deviation.

T	$\alpha = 0$	$\alpha = .2$	$\alpha = .5$
100	.306 (.022)	.300 (.023)	.266 (.025)
$1,000$	.118 $(2.65 \cdot 10^{-3})$	.115 $(2.74 \cdot 10^{-3})$	.102 $(3.00 \cdot 10^{-3})$

Table 8: Average rate of convergence $\widehat{\delta}$ for the distribution of different impulse responses. Standard deviation in paranthesis (), true values in brackets [ ].

	T	$\alpha = 0$	$\alpha = .2$	$\alpha = .5$
$\widehat{\alpha}_T^* - \widehat{\alpha}_T$	100	.543 [0.5] (.005)	.529 [0.5] (.005)	.516 [0.5] (.006)
	$1,000$	.513 [0.5] (.004)	.507 [0.5] (.004)	.497 [0.5] (.004)
$(\widehat{\alpha}_T^*)^2 - \widehat{\alpha}_T^2$	100	.841 [1.0] (.022)	.702 [0.5] (.021)	.559 [0.5] (.008)
	$1,000$	.859 [1.0] (.020)	.554 [0.5] (.004)	.514 [0.5] (.004)
$(\widehat{\alpha}_T^*)^{10} - \widehat{\alpha}_T^{10}$	100	3.660 [5.0] (1.072)	2.256 [0.5] (1.007)	.981 [0.5] (.048)
	$1,000$	3.782 [5.0] (1.034)	1.031 [0.5] (.022)	.661 [0.5] (.007)
$(\widehat{\alpha}_T^*)^{20} - \widehat{\alpha}_T^{20}$	100	7.311 [10.0] (4.345)	4.425 [0.5] (4.256)	1.641 [0.5] (.213)
	$1,000$	7.557 [10.0] (4.184)	1.784 [0.5] (.090)	.878 [0.5] (.013)

(1996). The studentized version of the standard bootstrap (CI_3) shows in principle the same performance for $\alpha = 0$. In the case of $\alpha \neq 0$ the coverage is better but still deteriorates for $\alpha = .2$, $h > 1$, and $T = 100$.

The CIs based on the subsampling bootstrap (CI_5) produce roughly the correct coverage level. In some cases their length is considerably greater than that of other methods, especially for large h. For α^{10} and α^{20} the interval length is unacceptable in the case of $\alpha = 0$. In addition, taking into account that for more complicated situations the subsampling bootstrap involves a substantially larger computational burden, the virtue of this method is difficult to see at least in the present context.

In the case of $\alpha = 0$ the subsampling procedure which additionally estimates the rate of convergence (CI_6) is superior to the other subsampling method for $h > 1$. Even though it estimates more conservative CIs it seems to have a slight advantage to the standard bootstrap in terms of interval length. This observation changes for $\alpha \neq 0$. The interval length is still shorter compared to the other subsampling method but it seems to be achieved at the cost of reduced coverage probability. A reason for that might be a poor estimator of δ which implies poor estimation results for τ_T.

Methods 1, 2, 3 and 6 allow for a straightforward generalization to the case of multivariate autoregressive processes of higher order. Remember, however, that methods 1, 2, and 3 are asymptotically incorrect in particular cases of interest, which are mimicked by $\alpha = 0$ and $h > 0$ in our simplified context. The subsampling method 5 can also be easily generalized if one finds a norming factor leading to a statistic with a nondegenerate limit distribution. Even the indirect method, which was clearly the winner in our competition, can in principle be generalized. However, this will certainly lead to quite involved computational problems, and, hence this method will suffer from the "curse of dimensionality".

In conclusions, the simulations show that all the CIs have severe drawbacks. Since some of them work very poorly for the presently considered simplest case there is clearly not much hope that they behave well in more complicated situations where higher order or higher dimensional processes are of interest.

6 Conclusions

In this study we have considered CIs of impulse responses computed from estimated AR processes. A detailed analysis of the simple univariate first order stationary AR case reveals that the presently used methods are all beset with severe drawbacks. CIs based on standard asymptotic theory and standard bootstrap may be drastically too large in the case of a degenerate limit distribution and, in small samples, may have real coverage probabilities quite different from the corresponding nominal levels. The former problem can not be cured by simple bias adjustment or other procedures which have been proposed in the impulse response literature. Modified procedures exist, however, that can take care of the problems asymptotically. Unfortunately, the small sample performance of these procedures is not impressive and it is also not obvious how they can be extended to general higher dimensional processes of order greater than one. Therefore, at present we are unable to recommend any of the methods to assess the sampling variability of impulse responses in practice. In fact, given this state of affairs, it may be a reasonable strategy to use different methods to determine whether the main results are robust with respect to the method used for setting up CIs. Clearly, reporting no measures of sampling variability of impulse responses at all is also unaccaptable because that may result in spurious conclusions.

Appendix

Proof of Theorem 1

First we rewrite $\widetilde{CI}_7$ in an equivalent form:

$$\widetilde{CI}_7 = \begin{cases} \{\alpha^h \mid \quad \widehat{\alpha}_T \in [r_{\alpha,\gamma/2}, r_{\alpha,1-\gamma/2}]\}, & \text{if } h \text{ is odd} \\ \{\alpha^h \mid \quad |\widehat{\alpha}_T| \in [s_{\alpha,\gamma/2}, s_{\alpha,1-\gamma/2}]\}, & \text{if } h \text{ is even} \end{cases}$$

where $r_{\alpha,\gamma}$ and $s_{\alpha,\gamma}$ are corresponding quantiles of the distribution of $\widehat{\alpha}_T$ and $|\widehat{\alpha}_T|$, respectively. Further, CI_7 can be rewritten in an analogous manner, based on empirical quantiles $r^*_{\alpha,\gamma}$ and $s^*_{\alpha,\gamma}$. This means that we can base the proof on a comparison of the distributions of $\sqrt{T}(\widehat{\alpha}_T - \alpha)$ and $\sqrt{T}(\widehat{\alpha}_T^{\alpha,*} - \alpha)$.

It is well-known that $\sqrt{T}(\widehat{\alpha}_T - \alpha)$ is asymptotically normally distributed. However, to prove uniformity in α, we need some explicit bounds for the error of approximation. It can be shown, for example by a Skorokhod embedding of $\sum y_{t-1}u_t$ in a Wiener process (see, e.g., Hall & Heyde (1980, Appendix 1)), that

$$\sup_{\alpha\in[-c,c]} \sup_x \left\{ \left| P_\alpha \left((E\sum y_{t-1}^2)^{-1/2} \sum y_{t-1}u_t \le x\sigma_u \right) - F_{N(0,1)}(x) \right| \right\} = O(T^{-\delta}),$$

for some $\delta > 0$. Here and in the following we use δ to denote an appropriate positive constant which may attain different values at different places.

Since, for $\alpha \in [-c, c]$, the y_t are absolutely regular (β-mixing) with exponentially decaying mixing coefficients, one can show

$$\sup_{\alpha\in[-c,c]} \left\{ \left| P_\alpha \left(\left| \frac{1}{T}\sum y_{t-1}^2 - E\frac{1}{T}\sum y_{t-1}^2 \right| + \left| \frac{1}{T}\sum y_{t-1}^2 - \sigma_u^2/(1-\alpha^2) \right| > T^{-\delta} \right) \right| \right\} = O(T^{-\delta}),$$

which implies that

$$\sup_{\alpha\in[-c,c]} \sup_x \left\{ \left| P_\alpha \left(\sqrt{T}(\widehat{\alpha}_T - \alpha) \le x\sqrt{1-\alpha^2} \right) - F_{N(0,1)}(x) \right| \right\} = O(T^{-\delta}). \qquad (A.1)$$

Analogously we get

$$\sup_{\alpha\in[-c,c]} \sup_x \left\{ \left| P_\alpha \left(\sqrt{T}(\widehat{\alpha}_T^{\alpha,*} - \alpha) \le x\sqrt{1-\alpha^2} \middle| y_0, \ldots, y_T \right) - F_{N(0,1)}(x) \right| \right\} = O(T^{-\delta}) \qquad (A.2)$$

for $(y_0, \ldots, y_T) \in \Omega_T$, where Ω_T is an appropriate set with $P(\Omega_T) \ge 1 - O(T^{-\delta})$. From (A.1) and (A.2) we obtain that

$$\sup_{\alpha\in[-c,c]} \sup_y \left\{ \left| P_\alpha(\sqrt{T}(\widehat{\alpha}_T - \alpha) \le y) - P_\alpha(\sqrt{T}(\widehat{\alpha}_T^{\alpha,*} - \alpha) \le y \mid y_0, \ldots, y_T) \right| \right\} = O(T^{-\delta}) \qquad (A.3)$$

holds with a probability exceeding $1 - O(T^{-\delta})$. Hence,

$$\sup_{\alpha\in[-c,c]} \left\{ \left| P_\alpha \left(\widehat{\alpha}_T \in [r_1, r_2] \right) \big|_{r_1 = r^*_{\alpha,\gamma/2}, r_2 = r^*_{\alpha,1-\gamma/2}} - (1-\gamma) \right| \right\} = O(T^{-\delta}) \qquad (A.4)$$

is valid for all $r^*_{\alpha,\gamma/2}$ and $r^*_{\alpha,1-\gamma/2}$ that correspond to $(y_0, \ldots, y_T) \in \Omega_T$. This yields immediately

$$\sup_{\alpha\in[-c,c]} \left\{ \left| P_\alpha \left(\widehat{\alpha}_T \in [r^*_{\alpha,\gamma/2}, r^*_{\alpha,1-\gamma/2}] \right) - (1-\gamma) \right| \right\} = O(T^{-\delta}), \qquad (A.5)$$

which completes the proof in the case of odd h. The proof for even h is analogous.

References

Anderson, T. W. (1959). On asymptotic distribution of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30**, 676–687.

Andrews, D. W. K. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica* **61**, 139–165.

Bertail, P., D. N. Politis & J. P. Romano (1997). On subsampling estimators with unknown rate of convergence, Technical Report no. 95–20, Dept of Statistics, Purdue University (submitted to JASA).

Bickel, P. J., F. Götze & W. R. van Zwet (1997). Resampling fewer than n observations: gains, losses and remedies for losses. *Statist. Sinica* **7**, 1–31.

Chan, N. H. & Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16**, 367–401.

Datta, S. (1995). On a modified bootstrap for certain asymptotically nonnormal statistics. *Statist. Probab. Lett.* **24**, 91–98.

Efron, B. & R. J. Tibshirani (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Science* **1**, 54–77.

Fachin, S. & L. Bravetti (1996). Asymptotic normal and bootstrap inference in structural VAR analysis, *Journal of Forecasting* **15**, 329–341.

Griffiths, W. & H. Lütkepohl (1989). Confidence intervals for impulse responses from VAR models: A comparison of asymptotic theory and simulation approaches, Discussion Paper, University of New England, Armidale, Australia.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. (with discussion.) *Ann. Statist.* **16**, 927–953.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*, New York: Academic Press.

Heimann, G. & J.-P. Kreiss (1996). Bootstrapping general first order autoregression, *Statistics and Probability Letters* **30**, 87–98.

Kilian, L. (1995). Small-sample confidence intervals for impulse response functions, Discussion Paper, University of Pennsylvania, Department of Economics.

Kilian, L. (1996). An endogenous lag order bootstrap algorithm with an application to impulse response analysis, Discussion paper.

Li, H. and G. S. Maddala (1996). Bootstrapping time series models. *Econometric Reviews* **15**, 115–158.

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

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag.

Lütkepohl, H. & J. Breitung (1997). Impulse response analysis of vector autoregressive processes, Discussion Paper No. 86, SFB 373, Humboldt University Berlin.

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

Politis, D. N. & J. P. Romano (1994). Large sample confidence regions based on subsamples under minimal assumptions, *Ann. Statist.* **22**, 2031–2050.

Saikkonen, P. & H. Lütkepohl (1995). Asymptotic inference on nonlinear functions of the coefficients of infinite order cointegrated VAR processes, Discussion Paper No. 66, SFB 373, Humboldt University Berlin.

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**, 1–48.

Sims, C. A. (1981). An autoregressive index model for the U.S. 1948-1975, in: J. Kmenta & J.B. Ramsey (Hrsg.), *Large-Scale Macro-Econometric Models*, Amsterdam: North-Holland, pp.283-327.

Sims, C. A. & T. Zha (1994). Error bands for impulse responses, Discussion paper.

Watson, M. W. (1994). Vector autoregressions and cointegration, in: Engle, R.F. and D.L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, New York: Elsevier.

de Wet, T. and J. W. J. van Wyk (1986). Bootstrap confidence intervals for regression coefficients when the residuals are dependent, *J. Statist. Comput. Simul.* **23**, 317–327.

White, J. S. (1958). The limiting distribution in the serial correlation coefficient in the explosive case, *Ann. Math. Statist.* **29**, 1188–1197.