Finite-sample properties of estimators for first and second order autoregressive processes

Sigrunn H. Sørbye[1] · Pedro G. Nicolau[1] · Håvard Rue[2]

Abstract

The class of autoregressive (AR) processes is extensively used to model temporal dependence in observed time series. Such models are easily available and routinely fitted using freely available statistical software like R. A potential problem is that commonly applied estimators for the coefficients of AR processes are severely biased when the time series are short. This paper studies the finite-sample properties of well-known estimators for the coefficients of stationary AR(1) and AR(2) processes and provides bias-corrected versions of these estimators which are quick and easy to apply. The new estimators are constructed by modeling the relationship between the true and originally estimated AR coefficients using weighted orthogonal polynomial regression, taking the sampling distribution of the original estimators into account. The finite-sample distributions of the new bias-corrected estimators are approximated using transformations of skew-normal densities, combined with a Gaussian copula approximation in the AR(2) case. The properties of the new estimators are demonstrated by simulations and in the analysis of a real ecological data set. The estimators are easily available in our accompanying R-package for AR(1) and AR(2) processes of length 10–50, both giving bias-corrected coefficient estimates and corresponding confidence intervals.

Keywords Bias correction · Gaussian copula · Orthogonal polynomial regression · Sampling distribution · Skew-normal approximation

1 Introduction

The class of autoregressive (AR) processes is one of the most central and widely applied time series models. These processes are simple and intuitive, expressing the current value of a time

---

✉ Sigrunn H. Sørbye
  sigrunn.sorbye@uit.no

  Pedro G. Nicolau
  pedro.nicolau@uit.no

  Håvard Rue
  haavard.rue@kaust.edu.sa

[1] UiT The Arctic University of Norway, Tromsø, Norway

[2] King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

series as a linear combination of previous values and additional random noise. Specifically, a pth order autoregressive process (AR(p)) can be defined by

$$x_t - \mu = \sum_{j=1}^{p} \phi_j(x_{t-j} - \mu) + w_t \tag{1}$$

where the set $\{\phi_j\}_{j=1}^{p}$ are fixed coefficients and $\{w_t\}$ is a Gaussian white noise process, $w_t \sim N(0, \sigma^2)$. Throughout this paper, the mean $\mu = \mathrm{E}(x_t)$ is assumed to be unknown. Originated by the famous Yule–Walker equations (Yule 1927; Walker 1931), the statistical properties of AR models have been thoroughly studied and established, see e.g. Brockwell and Davis (2002), Box et al. (2008) and Shumway and Stoffer (2017) for comprehensive introductions. However, one problem still remains. Commonly used estimators for the AR coefficients are severely biased for small sample sizes, e.g. having less than 50 observations (Shaman and Stine 1988; Huitema and McKean 1991; DeCarlo and Tryon 1993; Cheang and Reinsel 2000).

The estimation bias for finite-sample AR processes is problematic in several fields of applications in which realistic time series lengths are inherently short, e.g behavioral data (Arnau and Bono 2001), economic time series (Patterson 2000; Kruse et al. 2018), and in environmental and ecological research studies (Bence 1995; Ives et al. 2010). One example is within population ecology, where the coefficient estimates of fitted AR processes give intrinsic relevant ecological information characterizing population dynamics like density-dependence, cyclical frequency and measures of synchrony (Moran 1953; Bjørnstad et al. 1995; Stenseth et al. 2003; Hugueny 2006; Cohen and Saitoh 2016). In general, the bias in the parameter estimates for finite-samples affects forecast accuracy (Stine 1987; Kim 2003) and even a small bias in the parameter estimates might have a severe influence in estimating non-linear functions of the coefficients, e.g. the impulse response function (Patterson 2000). The estimation bias could also make commonly used hypothesis tests to classify stationary AR processes unreliable for finite-samples (Liu and Maharaj 2013) as such tests typically rely on asymptotic distributions.

This paper studies the finite-sample properties of commonly applied estimators for the coefficients of stationary AR(1) and AR(2) processes and provides bias-corrected versions of such estimators. The AR(1) and AR(2) processes do have considerable practical importance (Box et al. 2008, p. 53), especially for short time series and various estimators for the AR coefficients are easily available. The popular `stats:::ar`-function in R (R Core Team 2020) implements the closed form solution given by the Yule–Walker equations as its default. The `ar`-function also provides estimates using Burg's method (Burg 1967) and the ordinary least squares approach. The latter method is not considered further here as the resulting coefficient estimates often fall outside the stationary area of the AR processes. In addition, the `stats:::ar`-function implements the conditional maximum likelihood estimator (MLE), maximizing the likelihood given initial values of $x_0, \ldots, x_{p-1}$ in (1). The R-package `FitAR:::FitAR` (McLeod and Zhang 2008) provides the exact maximum likelihood estimator, where initial values are set using Burg's method. All of the mentioned estimators give very similar results for large sample sizes while the exact MLE has been claimed to usually perform better than alternatives for short time series (McLeod and Zhang 2006; Box and Luceno 1997).

The upper panels of Fig. 1 illustrate the average coefficient estimates of AR(2) processes of length $n = 15$ (left) and $n = 30$ (right), using the exact MLE, Burg's method and the Yule–Walker estimator. The results are based on generating 10,000 time series for each selected combination of (ϕ_1, ϕ_2) within the triangular stationary area of such processes. The AR(2)
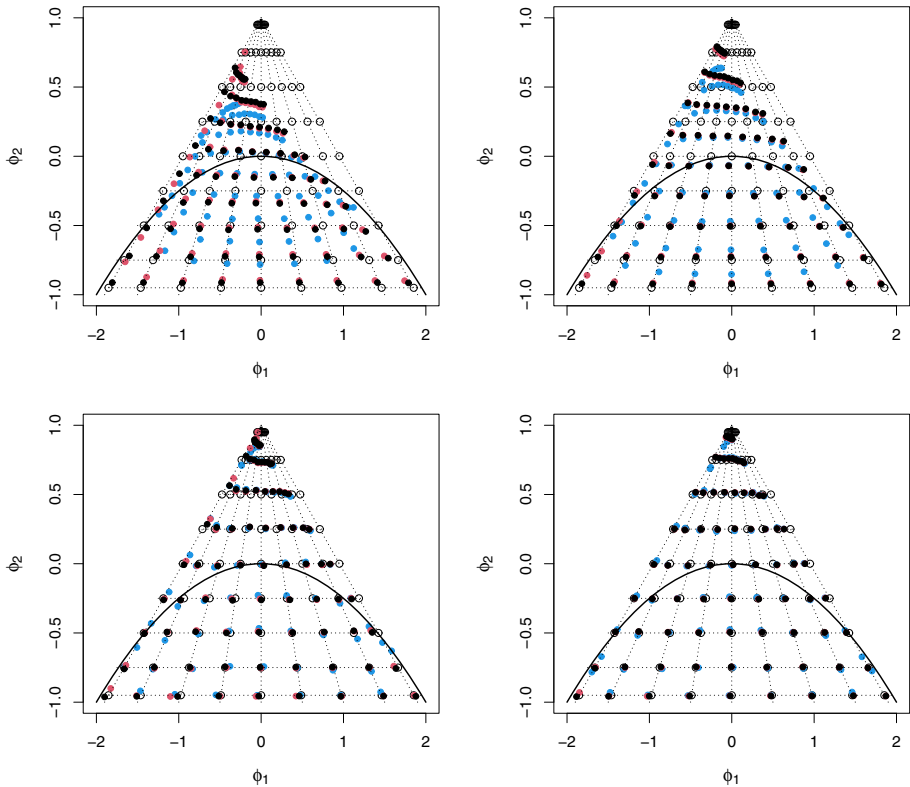
Fig. 1 Mean estimates of selected pairs (ϕ_1, ϕ_2) based on 10,000 simulations. Upper panels: The estimates using the exact MLE (black), Burg's algorithm (red) and the Yule–Walker solution (blue) when $n = 15$ (left) and $n = 30$ (right). Lower panels: The corresponding corrected mean estimates found by our proposed simulation-based approach

process has pseudo-periodic behavior for pairs of coefficients below the given parabolic curve. In the lower part of this region, the bias of the exact MLE and Burg's method is not too severe but it increases with increasing values of ϕ_2. The Yule–Walker approach clearly gives the most biased results for both sample sizes. The average estimates using the conditional MLE are not illustrated as these were not visually distinguishable from the exact MLE.

A number of methods to provide bias-corrected estimators for the AR coefficients have been proposed in literature. These range from asymptotic-based formulas for the bias (Marriott and Pope 1954; Kendall 1954; Shaman and Stine 1988; Tanaka 1984; Cordeiro and Klein 1994) to methods using restricted maximum likelihood (Cheang and Reinsel 2000) and bootstrapping (Thombs and Schucany 1990; Kim 2003). Andrews (1993) introduced a median-unbiased correction for the least squares estimator of the AR(1) coefficient which was generalized to give an approximate median-unbiased estimator of AR(p) processes in Andrews and Chen (1994). This estimator is implemented in the R-package `BootPR` (Kim 2014), also including the estimators of Shaman and Stine (1988) and Roy and Fuller (2001). The resulting estimates are not constrained to fall within the stationary area of the processes.

Our bias-correcting approach differs from previous suggestions in the sense that we model the true AR coefficients as a function of original estimates, accounting for the sampling

distribution of the original estimator. This is achieved by a brute-force simulation approach where we generate AR(1) and AR(2) processes for a fine grid of underlying true values of the coefficients. The relationship between the true and estimated coefficients is then described using a weighted orthogonal polynomial regression model. In the AR(2) case, the regression model is fitted based on a total of more than 59 million time series for a given sample size n. The resulting bias-corrected average estimates are illustrated in the lower panels of Fig. 1, again using the exact MLE, Burg's method and the Yule–Walker solution as original estimators. We do see a clear improvement in the bias properties for all of the methods. Admittedly, the bias-corrected estimators will not be completely unbiased as the relationship between the true and originally estimated coefficients cannot be modeled perfectly. This is especially seen for coefficient combinations along the borders of the triangular area.

The given brute-force simulation approach makes it possible to also derive sampling distributions for both the original and bias-corrected estimators. Specifically, we have fitted skew-normal distributions to transformations of the originally estimated coefficients of AR(1) and AR(2) processes. The parameters of the skew-normal distribution are then modeled in terms of the true underlying AR coefficients, again using orthogonal polynomial regression. In the AR(1) case, it is straightforward to obtain confidence intervals by Monte Carlo sampling, both when the original and bias-corrected estimators are used. In the AR(2) case, confidence intervals are found by combining Monte Carlo sampling and a Gaussian copula representation to preserve correlation between the estimated AR coefficients.

This paper is structured as follows. Section 2 outlines our modeling approach giving bias-corrected estimators of the first-lag autocorrelation coefficient of AR(1) processes and derive their sampling distributions. In Sect. 3, the suggested approach is extended to give bias correction and approximate sampling distributions in estimating the coefficients of AR(2) processes. Section 4 illustrates the bias correction for AR(2) processes for a real ecological data set, where the autoregressive coefficients represent direct and delayed density dependence for a vole species. Concluding remarks are given in Sect. 5. The Appendix describes our accompanying R-package which can be used to obtain the bias-corrected estimates and 95% confidence intervals for time series of length $n = 10, 11, \ldots, 50$.

2 Finite-sample properties for AR(1) processes

The AR(1) model has a simple one-step dependence as defined by (1) where $p = 1$. The bias in estimating the first-order autocorrelation coefficient $\phi_1 = \phi$ has been studied by several authors, see e.g Krone et al. (2017) for a recent comparative simulation study of AR(1) estimators in short time series. The bias of the exact MLE is illustrated for different sample sizes in Fig. 2, giving empirical averages for 10,000 simulations for a fine grid of ϕ-values in the stationary area $(-1, 1)$.

A common way to explicitly construct an unbiased estimate for a parameter ϕ is simply to subtract an estimate of the bias from the original estimator, providing a bias-corrected estimator of the form

$$\hat{\phi}_c = \hat{\phi} - \hat{E}(\hat{\phi} - \phi).$$

For example, the exact MLE can be corrected using the asymptotic bias $-(1+3\phi)/n$ (Tanaka 1984; Cordeiro and Klein 1994). Naturally, such a linear correction would not be accurate enough for small sample sizes. In Arnau and Bono (2001), the ordinary autocorrelation estimator for ϕ is bias-corrected by adding the absolute value of a polynomial fit to the

Fig. 2 The estimated empirical bias $\hat{\phi} - \phi$ of the exact MLE for AR(1) processes with length $n = 10$ (black), $n = 15$ (red), $n = 20$ (green), $n = 30$ (blue), $n = 40$ (light blue) and $n = 50$ (pink)
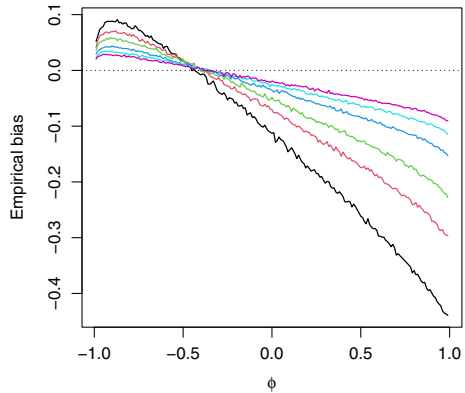
empirical bias. This has a slight resemblance to our approach, but an important difference is that we do not try to estimate or model the bias. We model the true parameter value ϕ directly as a function of original estimates using a large number of simulations. Similarly, we provide approximate sampling distributions by modeling parameters of skew-normal approximations as functions of the true values of ϕ.

2.1 Deriving bias-corrected estimators by simulation

Let $\hat{\phi}$ denote an original estimator for the first-lag autocorrelation coefficient of AR(1) processes. Our goal is to construct a bias-corrected estimator for ϕ such that $\mathrm{E}(\hat{\phi}_c) = \phi$ for all values of ϕ. To achieve this we model the relationship between the true and estimated parameter values using a weighted orthogonal polynomial regression model, using the true parameters as response variables. The coefficients of the regression model are calculated by minimizing the weighted squared error between the corrected estimates and the true parameter values for a grid of true values used as a training set. This provides a predictive model which can then be used to provide bias-corrected estimates for any original estimate $\hat{\phi}$. The specific steps in constructing this bias-corrected estimator can be summarized as follows:

1. To avoid constraints on the support of ϕ, we first introduce a monotonic transformation

$$g(\phi) = \mathrm{logit}\left(\frac{\phi + 1}{2}\right), \tag{2}$$

which has infinite support. This facilitates optimization and implies that our inverse transformed bias-corrected estimate will always be within the stationary area of the AR(1) process.

2. Let $\hat{\phi}$ denote an original estimator of ϕ. We model the true AR coefficient using an orthogonal polynomial model

$$\phi = f(\hat{\phi}, \boldsymbol{\beta}) = g^{-1}\left(\sum_{k=0}^{K} \beta_k h_k(g(\hat{\phi}))\right), \quad \hat{\phi} \in (-1, 1) \tag{3}$$

where $\boldsymbol{\beta} = \{\beta_k\}_{k=0}^{K}$ denotes a fixed set of regression coefficients while $\{h_k(.)\}_{k=0}^{K}$ represents a set of orthogonal polynomials of order k. Here, we choose to use the probabilists' Hermite polynomials which are orthogonal with respect to the standard normal density.

These polynomials are defined by

$$h_0(x) = 1, \quad h_1(x) = x, \quad h_{k+1}(x) = xh_k(x) - kh_{k-1}(x), \quad k \geq 1. \tag{4}$$

3. To estimate $\boldsymbol{\beta}$ for a given sample size n, we generate a total of $m = 10{,}000$ time series for a fine grid of ϕ-values, which can then be considered as our training set. Note that the suggested estimator is a non-linear function of $\hat{\phi}$ implying that

$$E(f(\hat{\phi}, \boldsymbol{\beta})) \neq f(E(\hat{\phi}), \boldsymbol{\beta}).$$

This means that the optimization takes the estimated value for each time series into account not just the average estimate of the m simulations for each ϕ. The regression coefficients are thus found by solving the optimization problem

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{r=1}^{l} \frac{1}{s_r^2} \left(\frac{1}{m} \sum_{j=1}^{m} g^{-1} \left(\sum_{k=0}^{K} \beta_k h_k(g(\hat{\phi}_{rj})) \right) - \phi_r \right)^2$$

$$= \arg\min_{\boldsymbol{\beta}} \sum_{r=1}^{l} \frac{1}{s_r^2} \left(\frac{1}{m} \sum_{j=1}^{m} f(\hat{\phi}_{rj}, \boldsymbol{\beta}) - \phi_r \right)^2. \tag{5}$$

The quantity $\hat{\phi}_{rj}$ represents the estimate of ϕ_r in simulation j. Specifically, we choose the grid of true parameter values defined by $\phi_r \in (-0.95, -0.94, \ldots, 0.95)$, implying that $l = 191$. The sample variances s_r^2 of the m parameter estimates of ϕ_r are used as weights. This gives a unique set of regression coefficients $\hat{\boldsymbol{\beta}}$ for each sample size n and for each original estimator $\hat{\phi}$, implying that $\hat{\boldsymbol{\beta}}$ accounts for the sampling distribution of the estimator $\hat{\phi}$. The bias-corrected estimator is then given by $\hat{\phi}_c = f(\hat{\phi}, \hat{\boldsymbol{\beta}})$ which is used to predict the true value of ϕ for a new estimate $\hat{\phi}$.

In minimizing (5), we have chosen to exclude values of ϕ close to the edges of the stationary interval in the training set. This is to avoid a severe inflation of the variance caused by forcing the estimator to give unbiased estimates at the edges of the interval. Also, we have chosen to use weighted regression where the weights s_r^2 represents the ordinary unbiased sample variances of the $m = 10{,}000$ estimates $\hat{\phi}_r$. This choice was made to robustify the given modeling approach to increasing sample variances close to the boundaries of the given interval but this did not have any major impact on the optimized values for $\boldsymbol{\beta}$.

In practice, the given approach can be used to find corrected estimates for any estimator $\hat{\phi}$ giving values within the stationary range, not only the four estimators considered here. The corrected estimates might be equal to ± 1, but will never fall outside of the interval $[-1, 1]$. Our implementation includes the exact and conditional MLEs, Burg's algorithm and the Yule–Walker solution. We have stored all the sets of regression coefficients for these four estimators for AR(1) series of length $n = 10, 11, \ldots, 50$. To facilitate use of the given bias-corrected estimates we have provided an accompanying R-package, see the Appendix.

2.2 Bias-correcting curves

Figure 3 illustrates how the correction works for AR(1) processes of length $n = 15$ and $n = 30$ for a fine grid of estimated ϕ values in the interval $(-1,1)$. The given curves correspond to using the exact MLE, Burg's algorithm and the Yule–Walker solution where the corrected estimators are calculated using up to cubic Hermite polynomials ($K = 3$). For an original estimate $\hat{\phi}$ given at the horizontal axis, we calculate the corresponding corrected estimate $\hat{\phi}_c$
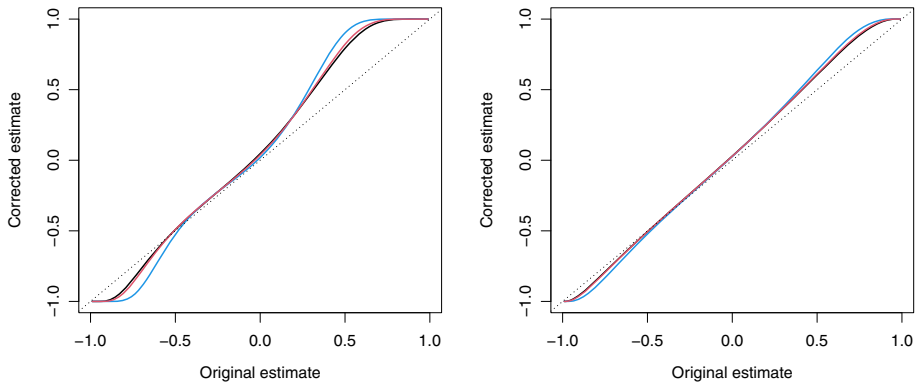
Fig. 3 The computed correction curves when $n = 15$ (left) and $n = 30$ (right) for the exact MLE (black), Burg's method (red) and the Yule–Walker solution (blue)

at the vertical axis. Naturally, this gives a quite large bias-correction when $n = 15$, where original estimates above 0.5 are corrected to a value close to 1. In the case of $n = 30$, we notice that the correction curves are close to linear for an internal subset of the interval.

The original and the corresponding corrected average estimates for $m = 10,000$ simulations are displayed in Fig. 4. Using the correction, we do get close to unbiased results both when $n = 15$ and $n = 30$. The overall average bias and sample variance for the estimators are given in Table 1. We also compute the overall root mean squared error (RMSE), which in the case of the corrected estimator is defined by

$$\text{RMSE}(\hat{\phi}_c) = \sqrt{\frac{1}{ml} \sum_{r=1}^{l} \sum_{j=1}^{m} \left(\hat{\phi}_{c,rj} - \phi_r \right)^2}$$

where $\hat{\phi}_{c,rj}$ is the corrected estimate of ϕ_r in simulation j. The RMSE for the original estimator and the bias and variances are computed correspondingly.

The results illustrate the well-known bias-variance trade-off implying that a decrease in the bias of an estimator will inherently cause an increase in the variance. Using the bias-corrected estimators we get approximately unbiased results only causing a negligible increase in RMSE, especially when $n = 30$. By increasing the order K of the Hermite polynomials, the bias can be further decreased but this also increases the variance giving higher RMSE and $K = 3$ was opted to be the best choice. In the minimization we could also have excluded more of the ϕ-values at the ends of the unit interval, e.g. using $\phi \in (-0.9, 0.9)$. This would reduce the variance but naturally also increase the bias close to the limits of the interval $(-1, 1)$.

2.3 Sampling distributions for the original and corrected estimators

Commonly applied estimators for the coefficients of AR(p) processes are asymptotically normal (Hannan 1970) but the finite-sample distributions of these estimators have not been thoroughly studied. Figure 5 illustrates the sampling distribution for the logit transformation $g(\hat{\phi})$ where $\hat{\phi}$ is the exact MLE for AR(1) processes of length $n = 30$ and where the underlying true values are $\phi \in (-0.9, 0.6, 0.9)$. The fitted curves are skew-normal densities which are seen to give very good approximations to the given sampling distributions. These
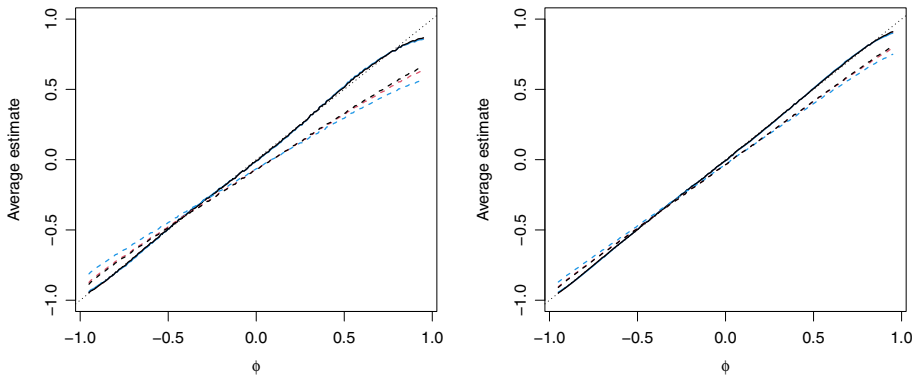
Fig. 4 The average original (dashed) and resulting corrected average estimates (solid) for 10,000 simulations, generating time series of length $n = 15$ (left) and $n = 30$ (right). The estimators used include the exact MLE (black), Burg's method (red) and the Yule–Walker solution (blue)

Table 1 The average bias, variance and root mean square error for the original and bias-corrected estimators of ϕ when $n = 15$ and $n = 30$

	Method	Original estimator			Bias-corrected estimator		
		Bias	Variance	RMSE	Bias	Variance	RMSE
$n = 15$	Exact MLE	-0.080	0.055	0.263	-0.0015	0.084	0.286
	Conditional MLE	-0.080	0.055	0.263	-0.0015	0.084	0.286
	Burg's method	-0.081	0.052	0.263	-0.0015	0.085	0.288
	Yule–Walker	-0.079	0.046	0.265	-0.0013	0.089	0.296
$n = 30$	Exact MLE	-0.037	0.026	0.171	-0.0004	0.031	0.174
	Conditional MLE	-0.037	0.026	0.171	-0.0004	0.031	0.174
	Burg's method	-0.038	0.025	0.172	-0.0005	0.031	0.175
	Yule–Walker	-0.038	0.024	0.175	-0.0005	0.032	0.178

The averages are computed based on $m = 10{,}000$ simulations for each value of $\phi \in (-0.95, -0.94, \ldots, 0.95)$

have been fitted using the function `fGarch:::snormFit` in R (Wuertz et al. 2020), implementing the skew-normal density as defined by Fernandez and Steel (1998), i.e.

$$\pi_{\text{sn}}(x) = \frac{2}{\xi + \frac{1}{\xi}} \left(\pi_G(x/\xi) H(x) + \pi_G(x\xi) H(-x) \right). \tag{6}$$

The function $\pi_G(.)$ denotes the standard normal density while $H(x)$ is the ordinary Heaviside or unit step function. In addition to the skewness parameter $\xi > 0$, the skew-normal density is parameterized in terms of a mean μ and standard deviation σ, using an input argument $(x - \mu)/\sigma$.

Based on the skew-normal density approximation for $g(\hat{\phi})$, we can derive finite-sample distributions for the estimators $\hat{\phi}$ and the corresponding bias-corrected estimators $\hat{\phi}_c$. Let $\tilde{\pi}_{\text{sn}}(.)$ denote the skew-normal approximation for the logit transformation $g(\hat{\phi})$. The approximate sampling distribution for the original estimator is easily expressed analytically by the ordinary change of variable transformation

$$\tilde{\pi}(\hat{\phi}) = \tilde{\pi}_{\text{sn}}(g(\hat{\phi})) \left| (1 + \hat{\phi})^{-1} + (1 - \hat{\phi})^{-1} \right|. \tag{7}$$
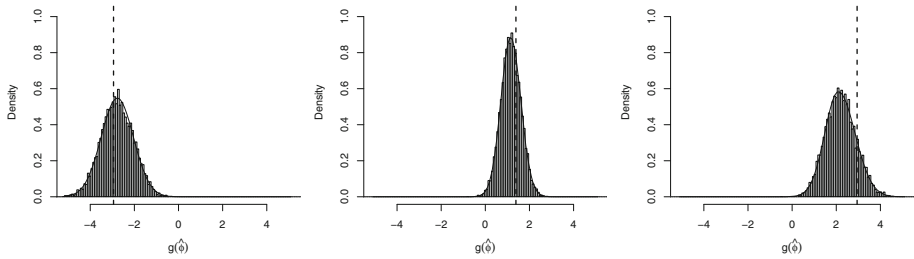
Fig. 5 Sampling distributions for $g(\hat{\phi})$ where $\hat{\phi}$ is the exact MLE. The dashed lines represent the true values of $g(\phi)$ where $\phi = -0.9$ (left), $\phi = 0.6$ (middle) and $\phi = 0.9$ (right)
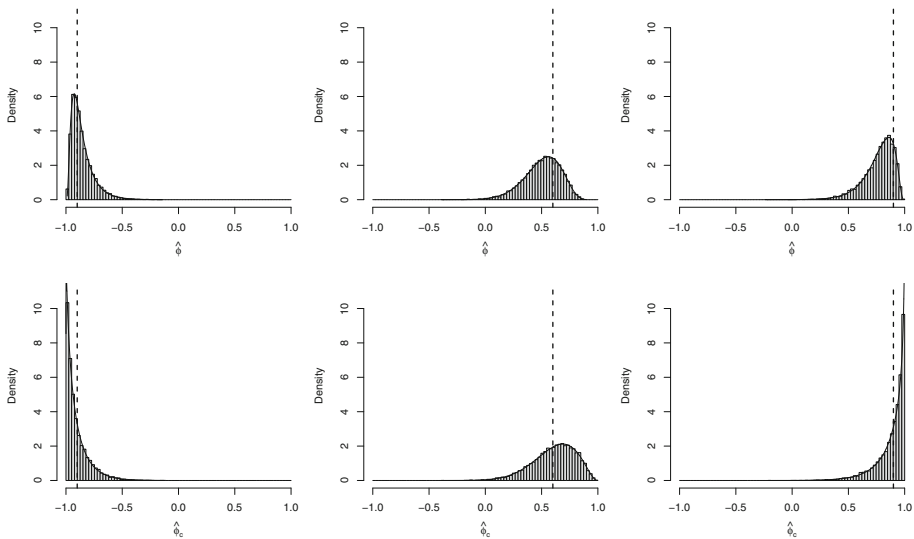

Fig. 6 Sampling distributions for $\hat{\phi}$ (upper panels) and $\hat{\phi}_c$ (lower panels) where $\hat{\phi}$ is the exact MLE. The dashed lines give the true values of ϕ where $\phi = -0.9$ (left), $\phi = 0.6$ (middle) and $\phi = 0.9$ (right)

Likewise, an approximation to the sampling distribution for the corrected estimator can be derived numerically as

$$\tilde{\pi}(\hat{\phi}_c) = \tilde{\pi}_{\text{sn}}(s(\hat{\phi}_c)) \left| \frac{ds(\hat{\phi}_c)}{d\hat{\phi}_c} \right|, \tag{8}$$

where $s(\hat{\phi}_c)$ represents a spline function approximating the monotonic relationship between $\hat{\phi}_c$ and $g(\hat{\phi})$. The resulting sampling distributions for the original and corrected estimators are shown in Fig. 6, where $n = 30$ and $\phi \in (-0.9, 0, 6, 0.9)$.

To calculate the sampling distributions in (7)–(8) for an estimator $\hat{\phi}$, we first need to assess the relationships between the parameters of the skew-normal approximation and ϕ. Using generic notation, let $\hat{\boldsymbol{\theta}}_r = (\hat{\theta}_{1,r}, \hat{\theta}_{2,r}, \hat{\theta}_{3,r}) = (\hat{\mu}_r, \hat{\sigma}_r, \ln(\hat{\xi}_r))$ denote the parameters of the skew-normal approximation to $m = 10{,}000$ samples of $g(\hat{\phi}_r)$, for each $\phi_r \in (-0.99, -0.98, \ldots, 0.99)$. Each of the estimated skew-normal parameters are used as response variables in separate orthogonal polynomial regression models
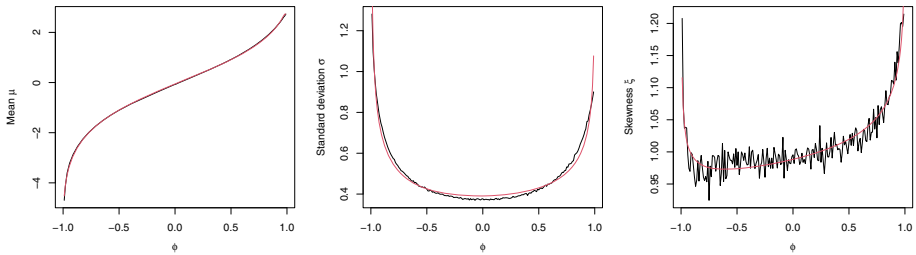
Fig. 7 The mean, standard deviation and skewness parameter of the skew-normal approximation for $g(\hat{\phi})$ as smoothed functions (red) of ϕ, where $\hat{\phi}$ represents the exact MLE when $n = 30$

$$\hat{\theta}_{s,r} = \sum_{k=0}^{K} b_{k,s} h_k(g(\hat{\phi}_r)), \quad s = 1, 2, 3, \quad r = 1, \ldots, 199 \tag{9}$$

The resulting estimates of the coefficients, $\{\hat{b}_{k,s}\}$, are found straightforwardly by ordinary least squares, using a polynomial order of $K = 3$. Figure 7 illustrates the relationships between each of the parameters of the skew-normal approximation to $g(\hat{\phi}_r)$ and ϕ_r, where $\hat{\phi}_r$ denotes the exact MLE for series of length $n = 30$. The red curves illustrate the smoothed version of these curves which can then be used to predict the parameters of the skew-normal approximation for a new value of $g(\hat{\phi})$. Especially, we notice that the skewness parameter is quite close to 1 for the interior of the given interval, implying that the sampling distributions for $g(\hat{\phi})$ are not very far from being Gaussian. The skewness increases as ϕ increases towards the upper limit of the stationary area.

By storing the coefficients $\{\hat{b}_{k,s}\}$ for all the four original estimators and each sample size n, we can now calculate confidence intervals for ϕ for all cases using Monte Carlo simulations. Given an estimate $g(\hat{\phi})$, we predict the parameters of the corresponding skew-normal approximation and sample from this distribution. These samples are easily transformed to give confidence intervals for ϕ, using the relevant percentiles for the distributions of $\hat{\phi}$ and $\hat{\phi}_c$. To investigate the coverage probability of the resulting confidence intervals, we performed a simulation study generating 10,000 AR(1) process with a uniformly drawn coefficient, i.e. $\phi \sim \text{Uniform}(-1, 1)$. The simulation study was performed for sample sizes $n = 10, 15, 20, 30, 40, 50$, in which the coefficient ϕ was estimated using each of the four original methods. We then calculated the bias-corrected estimates and found 95% equi-tailed confidence intervals for all cases by Monte Carlo simulation. The results demonstrate that the coverage probabilities for the original estimators are below the nominal level of 0.95 for all estimators and all sample sizes, see Table 2. Especially, the nominal level using the Yule–Walker solution is below 0.90 for all sample sizes. Using the bias-corrected estimators, the coverage properties are clearly better being quite close to the nominal level of 0.95 in all cases. For the smaller sample sizes, this can partly be explained by the larger variance of the bias-corrected estimators giving wider confidence intervals. For the large sample sizes, the confidence lengths are approximately the same.

3 Finite-sample properties for AR(2) processes

In this section we extend the given model-based approach to construct bias-corrected estimators for the pair of coefficients (ϕ_1, ϕ_2) of an AR(2) process. This is far more computationally

Table 2 Coverage probabilities of 95% confidence intervals for ϕ using the original estimators $\hat{\phi}$ and the respective bias-corrected versions $\hat{\phi}_c$ for a total of 10,000 simulations

n	Exact MLE		Conditional MLE		Burg's method		Yule–Walker	
	$\hat{\phi}$	$\hat{\phi}_c$	$\hat{\phi}$	$\hat{\phi}_c$	$\hat{\phi}$	$\hat{\phi}_c$	$\hat{\phi}$	$\hat{\phi}_c$
10	0.8092	0.9771	0.8097	0.9768	0.7798	0.9829	0.6692	0.9942
15	0.8326	0.9741	0.8330	0.9736	0.8012	0.9754	0.7198	0.9791
20	0.8463	0.9644	0.8468	0.9635	0.8233	0.9620	0.7415	0.9624
30	0.8617	0.9527	0.8608	0.9532	0.8460	0.9508	0.7814	0.9476
40	0.8728	0.9506	0.8773	0.9502	0.8621	0.9488	0.8088	0.9437
50	0.8813	0.9474	0.8878	0.9457	0.8722	0.9456	0.8234	0.9396

The true value of ϕ is randomly generated from $(-1, 1)$. The confidence intervals are found by Monte Carlo sampling after fitting a skew-normal approximation to $g(\hat{\phi})$

expensive than in the AR(1) case as we need to generate time series for a fine two-dimensional grid of the coefficients in the triangular stationary area. For each pair of the true parameters, we then fit weighted polynomial regression models to the generated time series and store the resulting optimized regression coefficients. As previously, the original estimators used include the exact and conditional MLE, Burg's algorithm and the Yule–Walker solution for AR(2) processes of length $n = 10, 11, \ldots, 50$. We obtain approximate sampling distributions for the bias-corrected versions of these estimators by constructing Gaussian copulas where the marginals are generated as transformations of skew-normal densities.

3.1 Modeling approach in two dimensions

The AR(2) process is defined by (1) where $p = 2$ and it is stationary within the triangular area constrained by $\phi_2 + |\phi_1| < 1$ where $|\phi_2| < 1$. A more appealing parameterization of this process is given by the partial autocorrelations,

$$\psi_1 = \frac{\phi_1}{1 - \phi_2}, \quad \psi_2 = \phi_2, \tag{10}$$

as the stationary area of the AR(2) process is then defined by the square $\psi_i \in (-1, 1)$, $i = 1, 2$. The area in which the process has pseudo-periodic behavior is characterized by $\psi_1^2(1 - \psi_2)^2 + 4\psi_2 < 0$.

We now extend the algorithm in Sect. 2 to construct bias-corrected estimators $(\hat{\phi}_{c,1}, \hat{\phi}_{c,2})$, again taking the sampling distribution of the original pair of estimators $(\hat{\phi}_1, \hat{\phi}_2)$ into account. In this case we estimate the parameters of the regression model by minimizing the squared error between the corrected and true partial autocorrelations. This is computationally beneficial to avoid the triangular constraints on ϕ_1 and ϕ_2. Also, the correlation between $\hat{\psi}_1$ and $\hat{\psi}_2$ is much smaller than the corresponding correlation between the estimates of the ϕ-coefficients. Naturally, this only makes a difference for the first coefficient as $\phi_2 = \psi_2$. The algorithm extending the weighted polynomial regression model to two dimensions can be summarized as follows:

1. Using the logit transformation in (2), the underlying true partial autocorrelations are modeled by

$$\psi_i = f(\hat{\psi}_1, \hat{\psi}_2, \boldsymbol{\beta}_i) = g^{-1}\left(\sum_{k=0}^{K}\sum_{q=0}^{K-k} \beta_{k,q,i} h_{k,q}(g(\hat{\psi}_1), g(\hat{\psi}_2))\right), \quad i = 1, 2 \quad (11)$$

where $h_{k,q}(g(\hat{\psi}_1), g(\hat{\psi}_2)) = h_k(g(\hat{\psi}_1))h_q(g(\hat{\psi}_2))$ denotes the product of Hermite polynomials of order k and q. Notice that the two partial autocorrelations are modeled separately, giving separate sets of regression coefficients $\boldsymbol{\beta}_i = \{\beta_{k,q,i}\}$ for $i = 1, 2$. However, each of the true partial autocorrelations need to be modeled in terms of the estimated pair $(\hat{\psi}_1, \hat{\psi}_2)$ as these parameters are not independent.

2. Due to the dependence, the regression coefficients $\boldsymbol{\beta} = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\}$ of the given predictors for ψ_1 and ψ_2 are found simultaneously. This is achieved by solving the following optimization problem:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{r=1}^{l}\sum_{i=1}^{2} \frac{1}{s_{ri}^2}\left(\frac{1}{m}\sum_{j=1}^{m} g^{-1}\left(\sum_{k=0}^{K}\sum_{q=0}^{K-k}\beta_{k,q,i}h_{k,q}(g(\hat{\psi}_{rj1}), g(\hat{\psi}_{rj2}))\right) - \psi_{ri}\right)^2$$

$$= \arg\min_{\boldsymbol{\beta}} \sum_{r=1}^{l}\sum_{i=1}^{2} \frac{1}{s_{ri}^2}\left(\frac{1}{m}\sum_{j=1}^{m} f(\hat{\psi}_{rj1}, \hat{\psi}_{rj2}, \boldsymbol{\beta}_i) - \psi_{ri}\right)^2. \quad (12)$$

The values $(\hat{\psi}_{rj1}, \hat{\psi}_{rj2})$ denote the original estimates for the rth pair of the partial autocorrelations in simulation j while s_{ri}^2 denotes the sample variances for the m simulations in each case. The value l denotes the total number of pairs of the partial autocorrelations that are included in the minimization.

In solving (12), we needed to generate time series for a fine two-dimensional grid of the parameter values (ψ_1, ψ_2) within the square defining the stationary area. The estimate $\hat{\boldsymbol{\beta}}$ is based on using the grid $\psi_i \in (-0.95, -0.925, \ldots, 0.95)$, $i = 1, 2$. This gives a total of $l = 77^2 = 5929$ different combinations of the partial autocorrelations. For each pair (ψ_1, ψ_2), we generated $m = 10,000$ times series of a specific length n implying that the regression coefficients are estimated based on approximately 59 million time series. This was repeated for all sample sizes $n = 10, 11, \ldots, 50$, such that the total number of generated time series is equal to $77^2 \cdot 41 \cdot 10,000 = 2,430,890,000$ or approximately 2.43 billion time series for a given value of K. We then saved the regression coefficients for each sample size and for each of the original estimation methods providing bias-corrected estimators

$$\hat{\psi}_{c,i} = f(\hat{\psi}_1, \hat{\psi}_2, \hat{\boldsymbol{\beta}}_i), \quad i = 1, 2$$

which are then transformed by (10) to give estimates for the AR coefficients. As in the AR(1) case, the estimated coefficients might fall at the border of the stationary area but not outside.

If run sequentially on an ordinary single-core laptop, the given brute-force simulation approach to compute the bias correction would be computationally infeasible as it would take approximately 5 years of CPU time. The computations were therefore done on the Ibex cluster at KAUST (https://www.hpc.kaust.edu.sa/ibex), which reduced the time down to about 2 days. The computation time could have been further reduced by lowering the number of generated times series from $m = 10,000$ to e.g $m = 3000$ giving approximately the same results. However, as the given calculations were done only once we chose to use a large number of generated time series.
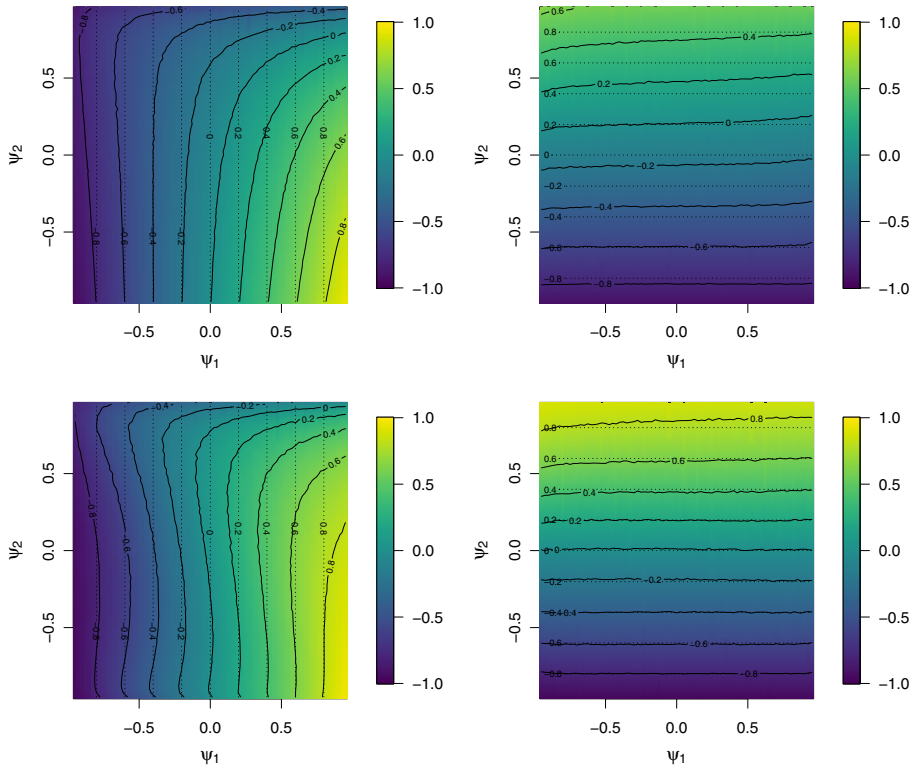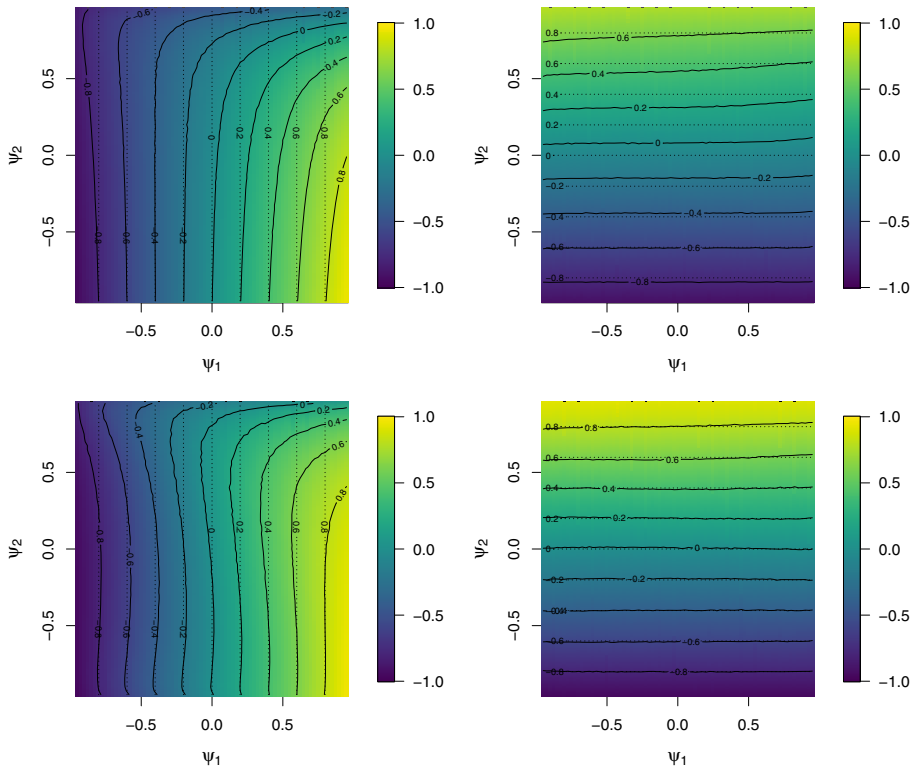
Fig. 8 The mean estimated values of ψ_1 (left) and ψ_2 (right) when $n = 15$, using the exact MLE (upper) and the bias-corrected estimator with $K = 3$ (lower)

3.2 Properties of the bias-corrected estimators

By using the partial autocorrelations in (12), the resulting corrected estimators for the AR(2) coefficients are not completely unbiased. However, in addition to the numerical advantages, we have noticed that this approach also gives smaller variance and RMSE compared to performing the optimization with respect to the ϕ-coefficients. As in the AR(1) case, the order K of the Hermite polynomials can be increased to slightly reduce the overall bias but the variance and the RMSE then increase. In the subsequent analysis we have therefore chosen to use $K = 3$ as this gave the best overall results. The number of regression coefficients in (12) is then equal to 10 for each of the parameters ψ_1 and ψ_2. Using for example $K = 7$, the model in (12) would have a total of 36 terms for each of the parameters.

The estimated partial autocorrelations using the exact MLE and the corresponding bias-corrected estimates are shown for sample sizes $n = 15$ and $n = 30$ in Figs. 8 and 9, respectively. Visually, the corrected estimator is very accurate in estimating ψ_2 but show some bias in estimating ψ_1, especially for the upper part of the square where the value of ψ_1 is either underestimated or overestimated depending on the value of ψ_2. The corresponding bias for ϕ_1 is much smaller than for ψ_1 as the given squared domain translates into a triangular area.

To further study finite-sample properties, we have calculated the overall average bias, variance and the RMSE of both the original and the bias-corrected versions. The calculations

Fig. 9 The mean estimated values of ψ_1 (left) and ψ_2 (right) when $n = 30$, using the exact MLE (upper) and the bias-corrected estimator with $K = 3$ (lower)

are based on taking the averages for the m simulations for each of the $l = 5929$ combinations of (ψ_1, ψ_2), which are easily transformed to give estimates for the AR coefficients by (10). In calculating the overall averages we have added the results for both parameters. For example, RMSE for the bias-corrected estimator is given by

$$\text{RMSE}((\hat{\phi}_{c,1}, \hat{\phi}_{c,2})) = \sqrt{\frac{1}{2l} \sum_{r=1}^{l} \sum_{i=1}^{2} \frac{1}{m} \sum_{j=1}^{m} (\hat{\phi}_{c,rji} - \phi_{ri})^2}.$$

The bias and the variance are calculated correspondingly. Table 3 summarizes the overall average bias, variance and the RMSE for the original and bias-corrected estimators using $K = 3$. The bias-corrected estimators do have a smaller bias and larger variance than the original estimators, but the improved bias properties do not appear at the expense of any significant increase in RMSE. When $n = 15$, the bias-corrected estimator only have a slightly larger RMSE for the MLE's, avoiding the rather large bias of the original estimators. When $n = 30$, the RMSE is actually smaller for the bias-corrected versions versus all of the original estimators.

Table 3 Bias, variance and root mean square error for the original estimators of (ϕ_1, ϕ_2) and the corrected estimator for $n = 15$ and $n = 30$

	Method	Original estimator			Bias-corrected estimator		
		Bias	Variance	RMSE	Bias	Variance	RMSE
$n = 15$	Exact MLE	-0.127	0.068	0.311	-0.017	0.099	0.317
	Conditional MLE	-0.126	0.068	0.312	-0.017	0.099	0.318
	Burg's method	-0.119	0.064	0.306	-0.018	0.102	0.323
	Yule–Walker	-0.124	0.054	0.334	-0.019	0.121	0.352
$n = 30$	Exact MLE	-0.061	0.029	0.186	-0.005	0.034	0.182
	Conditional MLE	-0.061	0.029	0.186	-0.005	0.034	0.183
	Burg's method	-0.061	0.029	0.188	-0.005	0.035	0.184
	Yule–Walker	-0.064	0.028	0.207	-0.006	0.040	0.199

These results are calculated using all the 10,000 simulations for each of the selected pairs in (ϕ_1, ϕ_2) within a fine grid of the stationary area

3.3 Sampling distributions using a Gaussian copula representation

Similar to the AR(1) case, we move on to find approximate sampling distributions for the estimators of the parameters of the AR(2) model. For each method and each sample size, we have fitted skew-normal densities to $m = 10,000$ generated samples of $g(\hat{\psi}_{r,1})$ and $g(\hat{\psi}_{r,2})$ where $r = 1, \ldots, l$. Figure 10 illustrates the skew-normal approximation for a few selected pairs of the transformed original estimators where the true underlying partial autocorrelations are given by $\boldsymbol{\psi} \in \{(-0.6, -0.6), (0.6, -0.6), (0.6, 0.6)\}$. The skew-normal densities are seen to give accurate approximations of the sampling distributions also in this case.

The next step is to find approximations of the sampling distributions for the original non-transformed estimators $(\hat{\phi}_1, \hat{\phi}_2)$ and the bias-corrected estimators $(\hat{\phi}_{c,1}, \hat{\phi}_{c,2})$. Similar to the AR(1) case, we first need to assess the relationship between the parameter estimates of the skew-normal approximations to $g(\hat{\psi}_1)$ and $g(\hat{\psi}_2)$ as a function of the true underlying parameters. To sample from the resulting bivariate distribution, we also need to be able to predict the correlation $\rho = \mathrm{Cor}(g(\hat{\psi}_1), g(\hat{\psi}_2))$. The full set of estimated parameters is then given by

$$\hat{\boldsymbol{\theta}}_r = \{\hat{\theta}_{1,r}, \hat{\theta}_{2,r}, \ldots, \hat{\theta}_{7,r}\} = \{\hat{\mu}_{1,r}, \hat{\sigma}_{1,r}, \ln(\hat{\xi}_{1,r}), \hat{\mu}_{2,r}, \hat{\sigma}_{2,r}, \ln(\hat{\xi}_{2,r}), \mathrm{logit}(\hat{\rho}_r)\}$$

for $r = 1, \ldots, l$. Using the same approach as in the AR(1) case, each of the seven parameters $\theta_{s,r} \in \hat{\boldsymbol{\theta}}_r$ is used as response variable in separate orthogonal polynomial regression models given by

$$\hat{\theta}_{s,r} = \sum_{k=0}^{K} \sum_{q=0}^{K-k} b_{k,q,s} h_{k,q}(g(\hat{\psi}_{r,1}), g(\hat{\psi}_{r,2})), \quad s = 1, \ldots, 7, \quad r = 1, \ldots, l \qquad (13)$$

The estimated coefficients $\{\hat{b}_{k,q,s}\}$ are again found by ordinary least squares. Using $K = 3$, this gives a set of 10 regression coefficients for each of the seven parameters, stored for all methods and sample sizes.
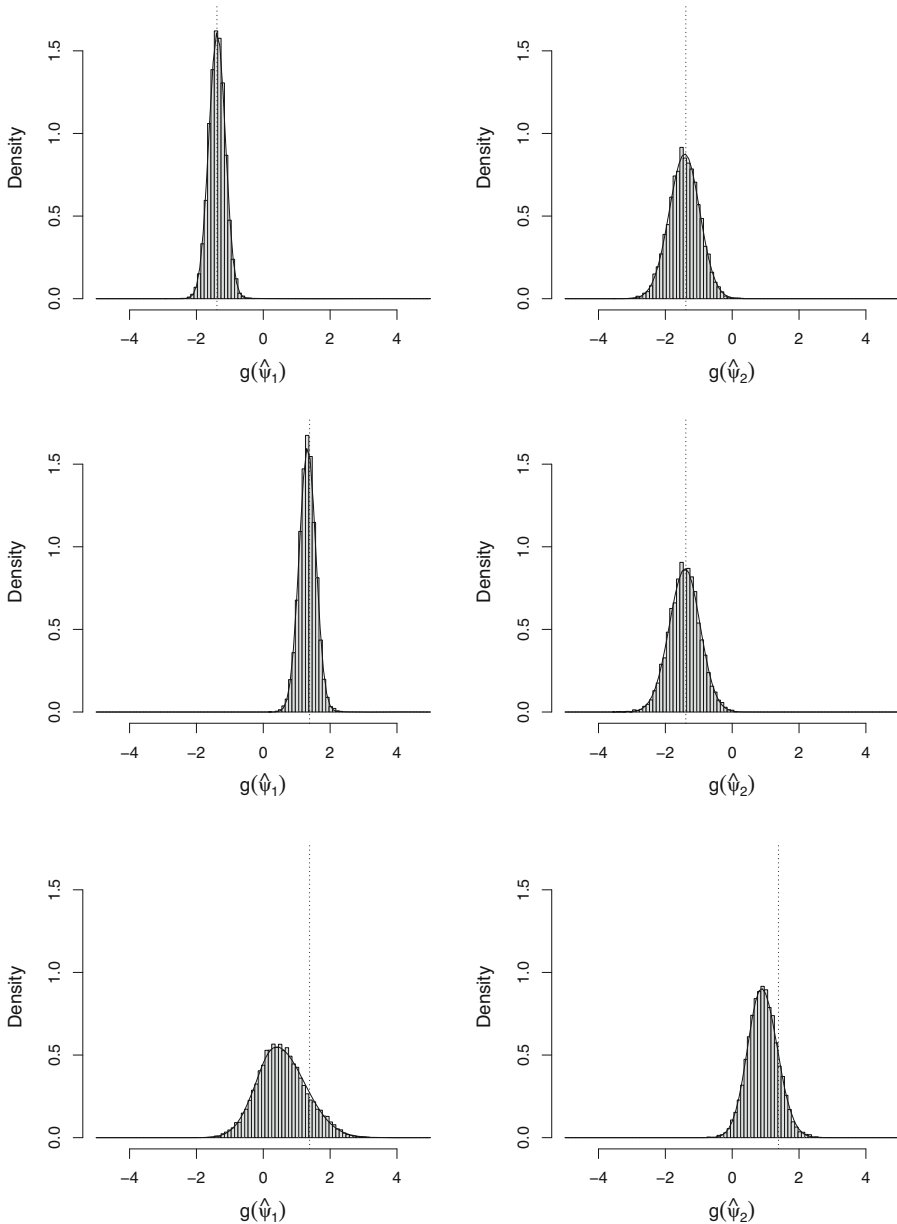
Fig. 10 The marginal sampling distributions for $g(\hat{\psi}_1)$ and $g(\hat{\psi}_2)$ for different combinations of (ψ_1, ψ_2), where the AR coefficients were estimated using the exact MLE. The different combinations of true partial autocorrelations include the pairs $(-0.6, -0.6)$ (upper), $(0.6, -0.6)$ (middle) and $(0.6, 0.6)$ (lower). The dotted vertical lines give the true values for $g(\psi_1)$ and $g(\psi_2)$

Table 4 Coverage for 95% confidence intervals for ϕ_1 and ϕ_2 found by sampling using a Gaussian copula and transformations of skew-normal densities

n	Method	Original estimator		Bias-corrected estimator	
		ϕ_1	ϕ_2	ϕ_1	ϕ_2
15	Exact MLE	0.8965	0.7454	0.9398	0.9637
	Conditional MLE	0.8967	0.7474	0.9413	0.9656
	Burg's method	0.8940	0.7268	0.9485	0.9687
	Yule–Walker	0.7896	0.5485	0.9507	0.9778
30	Exact MLE	0.9138	0.7815	0.9459	0.9443
	Conditional MLE	0.9139	0.7819	0.9459	0.9459
	Burg's method	0.9100	0.7609	0.9500	0.9438
	Yule–Walker	0.8335	0.6133	0.9360	0.9217

The results are based on 10,000 simulations where ψ_1 and ψ_2 are randomly generated from $(-1, 1)$

In generating samples for the original and bias-corrected estimators, we need to preserve the correlation between the transformed partial autocorrelations. This can be done by constructing a two-dimensional Gaussian copula by

$$C(u_1, u_2) = \Phi_{\hat{\rho}}(F_{\text{sn}}^{-1}(u_1; \hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1)), F_{\text{sn}}^{-1}(u_2; \hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2))).$$

The function $\Phi_{\hat{\rho}}(.)$ denotes the joint cumulative distribution of a bivariate standard normal vector with correlation between the components being equal to $\hat{\rho}$. The functions $F_{sn}(.)$ represent the skew-normal cumulative distribution functions (cdf) of $g(\hat{\psi}_1)$ and $g(\hat{\psi}_2)$. By the probability integral transform,

$$\boldsymbol{x} = (F_{\text{sn}}^{-1}(u_1; \hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1), F_{\text{sn}}^{-1}(u_2; \hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2))$$

represents samples from the given skew-normal densities where the uniformly distributed variables u_1 and u_2 are generated from the inverse cdf of the standard normal marginals of the bivariate distribution.

The resulting samples for $(\hat{\phi}_1, \hat{\phi}_2)$ and $(\hat{\phi}_{c,1}, \hat{\phi}_{c,2})$ can be used to find confidence intervals for (ϕ_1, ϕ_2). We performed a similar simulation study as in the AR(1) case, generating 10,000 AR(2) processes where the partial autocorrelation coefficients are drawn randomly from $(-1, 1)$. The coverage probabilities of the estimated 95% confidence intervals are given in Table 4 for sample sizes $n = 15$ and $n = 30$. We do notice that the coverage probabilities using the original estimators are smaller than the nominal level in all cases. In particular, the coverage is very low for ϕ_2 giving values below 0.80 also when $n = 30$. The confidence intervals using the bias-corrected estimators have coverage probabilities quite close to the nominal levels being within the interval 0.95 ± 0.03.

4 Application in population ecology

Wildlife ecological research studies are often characterized by small sample sizes (Bissonette 1999). This can be due to sparse distributions of the animal species of interest, and also the research design in which data are collected by fieldwork. As expressed by Ives et al. (2010): " *While a time series covering 40 years might represent an ecologist's entire career, such time series are short for statistical purposes"*. The population density of specific animal species
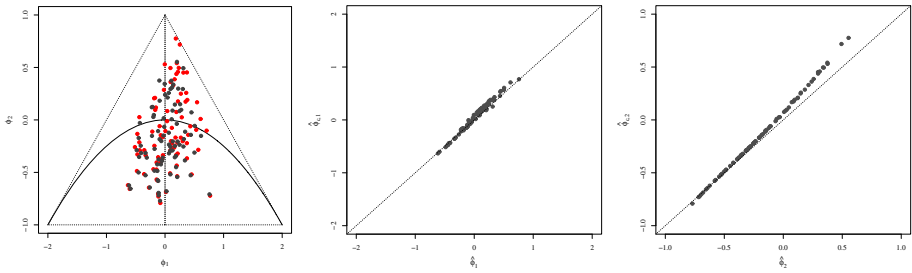
Fig. 11 Coefficient estimates fitting AR(2) processes to annual log-density estimates of Hokkaido vole populations observed at 85 different sites. The left panel shows the estimates of (ϕ_1, ϕ_2) using the exact MLE (black) and the bias-corrected estimates (red). The middle and right panels show the individual scatter plots of the bias-corrected versus original estimates

often exhibit cyclical fluctuations which, to a great extent, are driven by the relationship between the density and the carrying capacity of the environment. This is referred to as density dependence as the population density itself regulates the growth rates of the species, see e.g. Sinclair and Pech (1996). Both AR(1) and AR(2) processes have commonly been used to estimate this type of intra-specific population dynamics of a species, among others in modeling the population cycles of small rodents (Bjørnstad et al. 1995; Hansen et al. 1999; Stenseth et al. 2003; Nicolau et al. 2020). This type of analysis calls for accurate estimation of the AR coefficients as these are interpreted directly in terms of the strength of direct and delayed intra-specific dependence.

To demonstrate the new bias correction for a real world example, we consider a dataset on gray-sided voles (*Myodes rufocanus*), collected by the Japanese Forestry Agency at 85 different sites in the Hokkaido island, Japan (Saitoh et al. 1998). The dataset includes time series for the raw counts of the voles at each site, collected during both spring and fall for a total of 31 years (1962–1992). These data have been extensively studied in the literature and AR(2) model approximations have been used to assess the strength of density dependence, periodicities and synchrony of the vole populations (Stenseth et al. 2003; Hugueny 2006; Cohen and Saitoh 2016). Here, we fit the AR(2) model to the log of the density estimates for the annual fall time series derived by Cohen and Saitoh (2016) which account for differences in sampling effort. The data were downloaded from the Supporting Information of Cohen and Saitoh (2016), available at https://doi.org/10.1002/ecy.1575. Specifically, we used the log of the density estimates given in the file *App3BayesCountsParameterEstimates.csv* in their Zip archive.

The left panel of Fig. 11 illustrates the estimated AR(2) coefficients for the 85 time series using the exact MLE (black) and the corresponding bias-corrected estimates (red). About two-thirds (57/85) of the original pairs of estimates are within the pseudo-periodic area which implies cyclic population dynamics where shorter periods imply stronger density dependence (Stenseth et al. 2003). As expected, the difference between the original and bias-corrected estimates are not very large for this dataset as the time series length is $n = 31$ years. However, the bias correction is systematic in the sense that the estimates of ϕ_1 and ϕ_2 are mainly shifted right and upwards, respectively. This is further illustrated in the middle and right panels of Fig. 11 showing scatterplots of the corrected versus the original estimates. In correspondence with our simulation results, the bias correction of ϕ_1 is quite small for estimates that are not too close to the borders of the triangle, while ϕ_2 is increasingly underestimated for larger values of the parameter. The given bias-correction implies slightly weaker estimated direct density

dependence implying longer periods and two of the series can no longer be considered to be cyclic as the pairs of autoregressive coefficients fall outside of the pseudo-periodic area. Overestimation of the strength of density dependence have been associated with ignoring sampling variance (Stenseth et al. 2003) and ignoring the estimation bias could thus add to this overestimation.

Further use of the given bias correction for AR processes in population dynamics could include estimation of other ecologically interesting characteristics like the spectral density function, measures of cyclical frequency and measures of spatial synchrony. A simple example is the average length of the stochastic cycles for AR(2) processes

$$l = \frac{2\pi}{\cos^{-1}\left(\phi_1/(2\sqrt{-\phi_2})\right)}.$$

This is a non-linear function of the coefficients, but for simplicity we can use the original and bias-corrected estimates as plug-in estimates in this formula. For the 55 time series where both the original and bias-corrected ML estimates were within the pseudo-periodic area, the mean average cycle length increased slightly from $\bar{l}_{\text{original}} = 4.1$ to $\bar{l}_{\text{corrected}} = 4.4$ using the bias-corrected estimates. Naturally, this analysis is quite simplistic as we have neither taken sampling error nor spatial correlation between sites into account. It is well-known that spatial heterogeneity in population densities might have important implications for the population dynamics (Thorson et al. 2015), and a proper ecologically valid analysis of these time series should take this into account. Also, a more thorough study of the implications of using the bias-corrected estimator in estimating various characteristics of population dynamics represents interesting topics for future research.

5 Concluding remarks

The simplicity and parsimonious parameterization of first and second order AR processes make them attractive as plausible models for short time series. The AR(1) model reflects the Markov property, providing an important extension to a temporal independence assumption. The AR(2) process is more flexible and is particularly useful in modeling pseudo-periodic dynamics. The bias involved in estimating the coefficients of short AR processes has been well-known for decades but this still remains a practical problem even for the simple first and second order models. The choice of estimator does make a difference for small sample sizes and incautious use of commonly applied estimators might give misleading results. The default choice using the `ar`-function in R gives the Yule–Walker estimates which are clearly not optimal, neither for short nor long time series. As stated by Tjøstheim and Paulsen (1983), "*uncritical use of Yule–Walker estimates may be hazardous*".

The main goal of this paper was to investigate the finite-sample properties of commonly used estimators for the coefficients of AR(1) and AR(2) processes and provided bias-corrected versions of these estimators. This was achieved by modeling the true parameter values in terms of estimated values for a huge number of simulations, accounting for the sampling distribution of the original estimators and different sample sizes. The model fitting step was computationally expensive but needed to be done only once for each of the original estimators and each sample size, providing the regression coefficients which are used to compute the bias-corrected estimates.

The asymptotic behavior of the estimators of AR coefficients have been thoroughly studied in literature, while less effort has been made in studying their finite-sample properties. The given simulation-based approach makes it possible to find skew-normal approximations

to transformations of the coefficient estimates, in which the parameters of the skew-normal distributions are modelled in terms of the estimated AR coefficients. This provides approximate finite-sample distributions for both the original and bias-corrected estimators. These distributions are then used to give approximate confidence intervals for the AR coefficients.

The given simulation-based approach cannot easily be generalized to higher order AR processes, but this was never our intention. For short time series, it does make sense to apply simple and well interpretable models like the AR(1) and AR(2) processes, avoiding potential overfitting. Possible extensions of the given approach include studying finite-sample properties of estimators for other parsimoniously parameterized models like moving average (MA) processes of order 1 or 2 or the combined AR and MA model, ARMA(1,1). Naturally, this requires that the simulations and the regression model fitting steps have to be repeated. Even though this is computationally expensive, it is feasible with access to supercomputers, especially as the optimization step only needs to be done once for each method and sample size.

Data Availability The data analysed in Sect. 4 were downloaded from the Supporting Information of Cohen and Saitoh (2016), available at https://doi.org/10.1002/ecy.1575. Specifically, we used the log of the density estimates given in the file *App3BayesCountsParameterEstimates.csv* in their Zip archive.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Appendix: R-package

To make the presented methodology easily available, we include the R-package ARbiascorrect which can be used to obtain the bias-corrected estimates and the resulting approximate 95% confidence intervals, in addition to 95% confidence intervals using the original estimates. The package consists of only one function as defined below. The user can either use the relevant time series as input or the parameter estimates found by one of the original methods including the exact and conditional MLE, Burg's algorithm or the Yule–Walker solution. The R-package can be downloaded and installed from https://github.com/pedrognicolau/ARbiascorrect, or installed directly in R through the devtools package by running:

```
devtools::install_github("pedrognicolau/ARbiascorrect-v1")
```

The specifications of this package are as follows:

    Description
    Gives bias-corrected estimates and 95% confidence intervals for autoregressive coefficients of AR(1) and AR(2) processes, for sample sizes $n = 10, 11, \ldots, 50$.

Usage

```
biascorrect(phi = NULL, n = NULL,
method = c("mle", "cmle", "burg", "yw"),
order = NULL, x = NULL)
```

Arguments

phi Single value (AR1) or two-dimensional vector (AR2) containing the coefficient estimates subject to bias correction. Not required if the time series x is used as input.

n Integer for the length of the time series. Needs to be between 10 and 50. Not required if the time series x is used as input.

method Character string specifying the method used to estimate the autoregressive coefficients. Needs to be either mle, cmle, burg or yw, specifying the exact MLE, the conditional MLE, Burg's method and the Yule–Walker solution, respectively.

order Order of the estimated autoregressive process. Needs to be provided if x is used as input.

x The time series to be fitted by AR(1) or AR(2) in which order needs to be specified. The default estimation method is the exact MLE if none is specified.

Value

phi.hat The original estimates of the AR coefficients

phi.correct The bias-corrected estimates

ci.hat The 95% confidence interval using the original estimates

ci.correct The 95% confidence interval using the bias-corrected estimates

References

Andrews DWK (1993) Exactly median-unbiased estimation of first order autoregressive/unit root models. Econometrica 61:139–165

Andrews DWK, Chen HY (1994) Approximately median-unbiased estimation of autoregressive models. J Bus Econ Stat 12:187–204

Arnau J, Bono R (2001) Autocorrelation and bias in short time series: an alternative estimator. Qual Quant 35:365–387

Bence JR (1995) Analysis of short time series: correcting for autocorrelation. Ecology 76:628–639

Bissonette JA (1999) Small sample size problems in wildlife ecology: a contingent analytical approach. Wildl Biol 5:65–71

Bjørnstad ON, Falck W, Stenseth NC (1995) A geographic gradient in small rodent density fluctuations: a statistical modelling approach. Proc R Soc Lond B 262:127–133

Box GEP, Luceno A (1997) Time series analysis: forecasting and control. Wiley, New York

Box GEP, Jenkins GM, Reinsel GC (2008) Time series analysis: forecasting and control. Wiley, Hoboken

Brockwell PJ, Davis RA (2002) Introduction to time series and forecasting, 2nd edn. Springer, New York

Burg JP (1967) Maximum entropy spectral analysis. In: Proceedings of 37th meeting of the society of exploration geophysics, Oklahoma City

Cheang WK, Reinsel GC (2000) Bias reduction of autoregressive estimates in time series regression model through restricted maximum likelihood. J Am Stat Assoc 95:1173–1184

Cohen JE, Saitoh T (2016) Population dynamics, synchrony, and environmental quality of Hokkaido voles lead to temporal and spatial Taylors laws. Ecology 97:3402–3413

Cordeiro GM, Klein R (1994) Bias correction in ARMA models. Statist Probab Lett 19:169–176

DeCarlo LT, Tryon WW (1993) Estimating and testing autocorrelation with small samples: a comparison of the c-statistic to a modified estimator. Behav Res Ther 31:781–788

Fernandez C, Steel MFJ (1998) On Bayesian modeling of fat tails and skewness. J Am Stat Assoc 93:359–371

Hannan EJ (1970) Multiple time series. Wiley, New York

Hansen TF, Stenseth NC, Henttonen H, Tast J (1999) Interspecific and intraspecific competition as causes of direct and delayed density dependence in a fluctuating vole population. Proc Natl Acad Sci USA 96:986–991

Hugueny B (2006) Spatial synchrony in population fluctuations: extending the Moran theorem to cope with spatially heterogeneous dynamics. Oikos 115:3–14

Huitema BE, McKean JW (1991) Autocorrelation estimation and inference with small samples. Psycol Bull 110:291–304

Ives AR, Abbott KC, Ziebarth NL (2010) Analysis of ecological time series with ARMA(p, q) models. Ecology 91:858–871

Kendall MG (1954) Note on bias in the estimation of autocorrelation. Biometrika 41:403–404

Kim JH (2003) Forecasting autoregressive time series with bias-corrected parameter estimators. Int J Forecast 19:493–502

Kim JH (2014) BootPR: bootstrap prediction intervals and bias-corrected forecasting. https://CRAN.R-project.org/package=BootPR, R package version 0.60

Krone T, Albers CJ, Timmerman ME (2017) A comparative simulation study of AR(1) estimators in short time series. Qual Quant 51:1–21

Kruse R, Kaufmann H, Wegener C (2018) Bias-corrected estimation for speculative bubbles in stock prices. Econ Model 73:354–364

Liu S, Maharaj EA (2013) A hypothesis test using bias-adjusted AR estimators for classifying time series in small samples. Comput Stat Data Anal 60:32–49

Marriott FHC, Pope JA (1954) Bias in the estimation of autocorrelations. Biometrika 41:390–402

McLeod AI, Zhang Y (2006) Partial autocorrelation parameterization for subset autoregression. J Time Ser Anal 27:599–612

McLeod AI, Zhang Y (2008) Improved subset autoregression: with R package. J Stat Softw 28:1–28

Moran P (1953) The statistical analysis of the Canadian lynx cycle. Aust J Zool 1(3):291

Nicolau PG, Sørbye SH, Yoccoz NG (2020) Incorporating capture heterogeneity in the estimation of autoregressive coefficients of animal population dynamics using capture-recapture data. Ecol Evol 10:12710–12726

Patterson KD (2000) Bias reduction in autoregressive models. Econ Lett 68:135–141

R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Roy A, Fuller WA (2001) Estimation for autoregressive time series with a root near one. J Bus Econ Stat 19:482–493

Saitoh T, Stenseth NC, Bjørnstad ON (1998) The population dynamics of the vole clethrionomys rufocanus in Hokkaido, Japan. Res Popul Ecol 40:61–76

Shaman P, Stine RA (1988) The bias of autoregressive coefficient estimators. J Am Stat Assoc 83:842–848

Shumway RH, Stoffer DS (2017) Time series analysis and its applications with r examples. Springer, New York

Sinclair A, Pech RP (1996) Density dependence, stochasticity, compensation and predator regulation. Oikos 75:164–173

Stenseth NC, Viljugrein H, Saitoh T, Hansen TF, Kittilsen MO, Bølviken E, Glockner F (2003) Seasonality, density dependence, and population cycles in Hokkaido voles. Proc Natl Acad Sci 100:11478–11483

Stine RA (1987) Estimating properties of autoregressive forecasts. J Am Stat Assoc 82:1072–1078

Tanaka K (1984) An asymptotic expansion associated with the maximum likelihood estimators in ARMA models. J Roy Stat Soc B 46:58–67

Thombs LA, Schucany WR (1990) Bootstrap prediction intervals for autoregression. J Am Stat Assoc 85:486–492

Thorson JT, Skaug HJ, Kristensen K, Shelton AO, Ward EJ, Harms JH, Benante JA (2015) The importance of spatial models for estimating the strength of density dependence. Ecology 96:1202–1212

Tjøstheim D, Paulsen J (1983) Bias of some commonly-used time series estimates. Biometrika 70:389–399

Walker G (1931) On periodicity in series of related terms. Proc Roy Soc Ser A 131:518–532

Wuertz D, Setz T, Chalabi Y, Boudt C, Chausse P, Miklovac M (2020) fGarch: Rmetrics-autoregressive conditional heteroskedastic modelling. https://CRAN.R-project.org/package=fGarch, R package version 3042.83.2

Yule GU (1927) On a method of investigating periodicities in disturbed series, with special reference to Wolfers sunspot numbers. Philos Trans Roy Soc A Math Phys Eng Sci 226:267–298