



**MONASH** University

**Australia**

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**Bias Correction of Persistence Measures in  
Fractionally Integrated Models**

**Simone D. Grose, Gael M. Martin and Donald S. Poskitt,**

**December 2013**

**Working Paper 29/13**

# Bias Correction of Persistence Measures in Fractionally Integrated Models\*

Simone D. Grose, Gael M. Martin<sup>†</sup> and Donald S. Poskitt

*Department of Econometrics & Business Statistics, Monash University*

December 12, 2013

## Abstract

This paper investigates the accuracy of bootstrap-based bias correction of persistence measures for long memory fractionally integrated processes. The bootstrap method is based on the semi-parametric sieve approach, with the dynamics in the long memory process captured by an autoregressive approximation. With a view to improving accuracy, the sieve method is also applied to data pre-filtered by a semi-parametric estimate of the long memory parameter. Both versions of the bootstrap technique are used to estimate the finite sample distributions of the sample autocorrelation coefficients and the impulse response coefficients and, in turn, to bias-adjust these statistics. The accuracy of the resultant estimators in the case of the autocorrelation coefficients is also compared with that yielded by analytical bias adjustment methods when available.

*Keywords:* Long memory, ARFIMA, sieve bootstrap, bootstrap-based bias correction, sample autocorrelation function, impulse response function.

*JEL Classification:* C18, C22, C52

## 1 Introduction

Measuring the degree of persistence, or memory, in an economic or financial time series is crucial for understanding the response of the variable to shocks, in particular to policy-induced shocks. Traditionally, discussion of persistence has taken place in the

---

\*This research has been supported by Australian Research Council (ARC) Discovery Grant DP120102344 and ARC Future Fellowship FT0991045.

<sup>†</sup>Corresponding author: Gael Martin, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia. Tel.: +61-3-9905-1189; fax: +61-3-9905-5474; email: gael.martin@monash.edu.

context of models that are either integrated of order zero ( $I(0)$ ) or of order one ( $I(1)$ ), with the most commonly applied measures in this context being the impulse response and autocorrelation functions. The focus of this paper is on measuring persistence in the class of fractionally integrated ( $I(d)$ ) processes introduced by [Granger and Joyeux \(1980\)](#) and [Hosking \(1980\)](#) – a key class of models used to capture long memory, or *strong* dependence, in a wide range of empirical applications.

Long memory  $I(d)$  processes can be characterized by the specification

$$y(t) = \sum_{j=0}^{\infty} \psi(j)\varepsilon(t-j) = \frac{\kappa(z)}{(1-z)^d} \varepsilon(t), \quad (1.1)$$

where  $\varepsilon(t)$ ,  $t \in \mathbb{Z}$ , is a zero mean white noise process with variance  $\sigma^2$ ,  $z$  denotes the lag operator, and the ‘short-memory’ component,  $\kappa(z) = \sum_{j \geq 0} \kappa(j)z^j$ , is assumed to satisfy  $\sum_{j \geq 0} |\kappa(j)| < \infty$ , the transfer function of a stable, invertible autoregressive moving average (ARMA) process, for example. The long-run behaviour of this process depends on the fractional integration parameter  $d$ . Specifically, for any  $d \neq 0$  the impulse response coefficients  $\psi(j)$  in (1.1), as well as the autocovariances of the process, decline at a hyperbolic rate, rather than the exponential rate typical of a short-memory ARMA process. For the empirically relevant values of  $d > 0$  the rate of decline is slow enough to preclude (absolute) summability for both measures of persistence, leading to the characterization of  $y(t)$  as a ‘long-memory’ process in this case.

While the literature dealing with inference in the context of autoregressive fractionally integrated moving average (ARFIMA) models is well-developed<sup>1</sup>, some open problems remain, including with respect to inference about the two persistence measures. Most notable here is the well-known (downward) bias of estimates of the autocorrelation function (ACF) under long memory ([Hosking, 1996](#)), and the asymptotic non-Gaussianity of the sample autocorrelations for  $d \geq 0.25$ . While [Hosking \(1996\)](#) provides an asymptotically valid representation of the bias of the general  $k^{\text{th}}$ -order sample autocorrelation, it is dependent on unknown parameters. The same point holds for the higher-order bias adjustment for the first-order sample autocorrelation coefficient derived by [Lee and Ko \(2009\)](#). Both results yield feasible bias-adjustment methods only via the insertion of estimates of unknown population values, and the sampling properties of any resultant bias-adjusted estimators remain unknown<sup>2</sup>. Similarly, the problem of producing accurate confidence intervals for the impulse response function

<sup>1</sup> See [Beran \(1994\)](#), [Doukhan, Oppenheim and Taqqu \(2003\)](#) and [Robinson \(2003\)](#) for textbook expositions

<sup>2</sup> Lee and Ko exploit their bias expression to construct a bias-adjusted estimate of the long memory parameter. They achieve this via a method of moments approach that links an estimate of  $d$  to a bias-adjusted estimate of the first-order autocorrelation coefficient. This procedure is, in turn, only undertaken for the case of fractional noise, in which case the bias-adjustment of the first-order autocorrelation coefficient is a function of  $d$  only, with short memory coefficients playing no role in the calculations.

(IRF) under long memory has prompted investigation (e.g., [Baillie and Kapetanios, 2012](#)), but with no fully satisfactory solution having been found to date.

In the spirit of recent work in [Poskitt \(2008\)](#), [Baillie and Kapetanios \(2012\)](#) and [Poskitt, Grose and Martin \(2013\)](#) we use the semi-parametric sieve bootstrap (SBS) method to estimate the unknown sampling distributions of all autocorrelation and impulse response coefficients. The sieve works by ‘whitening’ the data using an autoregressive (AR) approximation and capturing the dynamics of the process in the fitted AR (the order of which increases at a suitable rate with the sample size). Results presented by [Poskitt \(2008\)](#), building on earlier results in [Poskitt \(2007\)](#), demonstrate that the sieve method produces error rates that are superior to those of the block bootstrap of [Künsch \(1989\)](#). Subsequently, [Poskitt et al. \(2013\)](#) have strengthened these results considerably, with the higher-order improvement yielded by the sieve method demonstrated using an Edgeworth expansion for a broad class of statistics that includes both forms of statistics investigated here. Furthermore, the authors have shown that the rate of convergence of a modified version of the technique – in which a consistent semi-parametric estimator of  $d$  is used to ‘pre-filter’ the data prior to the application of the sieve algorithm – is equivalent to that associated with the application of the sieve method to intermediate and short memory processes (see [Choi and Hall, 2000](#))<sup>3</sup>.

The results in [Poskitt et al. \(2013\)](#), relevant as they are to the IRF, go well beyond the results of [Baillie and Kapetanios \(2012\)](#), which (building in part on [Poskitt, 2008](#)) demonstrate convergence (using the Mallows metric) of the conventional sieve bootstrap to the true sampling distribution(s) of the impulse response coefficient(s). Moreover, in addition to the sample ACF being covered by the general theoretical results derived in [Poskitt et al. \(2013\)](#), the latter work also presents detailed numerical results pertaining to selected autocorrelation coefficients. Specifically, the accuracy with which the sieve methods reproduce both Monte Carlo-based estimates of the sampling distributions of these coefficients, and the (empirically infeasible) Edgeworth expansions, is documented. Most notably, the pre-filtering is shown to produce marked improvements over the ‘raw’ sieve method when the long-range dependence is strong. In the current paper the focus is explicitly on bias correction of both persistence measures. Exploiting the theoretical (and numerical) accuracy of the sieve-based distributional estimates established in [Poskitt et al. \(2013\)](#), we extract from those estimated distributions an appropriate estimate of bias (for any given statistic) and bias-adjust the statistic in the usual way. Consistent with the semi-parametric spirit of the exercise, the impulse response coefficients are produced as the inversion of a long AR fitted to the data, rather than as non-linear functions of the parameters of some parametrically specified ARFIMA model. The sample autocorrelation coefficients are calculated

---

<sup>3</sup> This rate is, in turn, arbitrarily close to the bootstrap rate of convergence attained for independent data.

using the standard Pearson formula.

In order for the bootstrap estimate of the bias to be a valid representation of the true but unknown bias, the ‘true’ reference value used in the bias computation needs to be consistent with the manner in which the bootstrap samples are generated. For example, when the raw sieve method is used, and the bootstrap samples thereby generated via an AR fitted to the empirical data, the reference values of all impulse response and sample autocorrelation coefficients need to be produced via this fitted AR process. In the case of the IRF this occurs via the inversion of the estimated AR model, whilst the reference ACF is deduced from the fitted AR coefficients using the Yule-Walker method. When the pre-filtered sieve method is used, the reference value for the IRF is produced by inverting the formula for an ARFIMA model with the given pre-filtered value used for  $d$  and the fitted AR model used for the autoregressive component. The reference value for the sample ACF in this case is produced using Sowell’s (1992) algorithm (as modified by Doornik and Ooms, 2001) for calculating the variance-covariance matrix of an ARFIMA model. Two alternative methods are, in turn, used to determine the length of the AR approximation in both the sieve method itself and in the computation of the reference values. For both persistence functions the pre-filtered sieve is based on the ‘semi-parametric Gaussian’ estimator of  $d$  examined by Robinson (1995), here referred to as the ‘semi-parametric local Whittle’ (SPLW) estimator. This estimator is shown in Poskitt et al. (2013) to satisfy the necessary conditions for the higher-order convergence properties of the pre-filtered sieve to obtain. As a proof-of-concept exercise, we also document results based on the use of the true (unknown) value of  $d$  as the pre-filter.

Accuracy of the bootstrap bias-corrected autocorrelation estimators is assessed relative to that of estimators adjusted using both the asymptotic bias of Hosking (1996) and, in the case of first-order autocorrelation, the analytic bias expression of Lee and Ko (2009). Given that the former expression for bias depends on the true (unknown) values of the parameters, it is empirically infeasible (in the absence of suitable estimators of those parameters) as a bias adjustment procedure; however, for a large enough sample size it represents the ideal toward which the bootstrap-based methods should converge. The latter bias expression is a higher order approximation than the former, and we render it feasible by inserting the empirical values of the Yule-Walker-based estimates of the autocorrelation coefficients upon which the expression depends. In so doing one is producing an appropriate expression for the bias (and subsequent bias-adjusted statistic) that the raw sieve bootstrap, in particular, is attempting to replicate numerically, making the comparison particularly apt.

The paper proceeds as follows. Section 2 briefly outlines the methodology underlying the sieve bootstrap and its use in estimating the sampling distribution and finite-sample bias of selected persistence measures. For conciseness we present the

more general pre-filtered methodology in detail, with this technique nesting the ‘raw’ sieve technique when the pre-filtering step is omitted. Selected results from [Poskitt et al. \(2013\)](#) that highlight the theoretical convergence rates of the bootstrap-based estimates of the sampling distributions on which the subsequent bias-adjustment rests, are also reproduced for convenience. In Section 3 we outline the key theoretical properties of the two persistence measures to be bias-adjusted, whilst in Section 4 the finite sample performance of the bias-corrected estimators is assessed via an extensive simulation exercise.

## 2 Long-memory processes, autoregressive approximation, and the pre-filtered sieve bootstrap

We assume that  $y(t)$  is a linearly regular, covariance-stationary process with representation as in (1.1) where the stochastic disturbance and the impulse response coefficients satisfy the following conditions:

**Assumption 1** *The process  $\varepsilon(t)$  is ergodic and,*

$$E[\varepsilon(t) \mid \mathcal{E}_{t-1}] = 0 \quad \text{and} \quad E[\varepsilon(t)^2 \mid \mathcal{E}_{t-1}] = \sigma^2, \quad (2.1)$$

where  $\mathcal{E}_t$  denotes the  $\sigma$ -algebra of events determined by  $\varepsilon(s)$ ,  $s \leq t$ . Furthermore,  $E[\varepsilon(t)^4] < \infty$ .

**Assumption 2** *The transfer function in the representation of the process  $y(t)$ , namely  $k(z) = \sum_{j \geq 0} \psi(j)z^j$ , is given by  $\psi(z) = \kappa(z)/(1-z)^d$  where  $|d| < 0.5$  and  $\kappa(z)$  satisfies  $\kappa(z) \neq 0$ ,  $|z| \leq 1$ , and  $\sum_{j \geq 0} j|\kappa(j)| < \infty$ .*

Assumption 1 imposes a classical martingale difference structure on the stochastic disturbance process; the key property of such a process that underlies the asymptotic results being that a martingale difference is uncorrelated with any measurable function of its own past. Assumptions 1 and 2, taken together, incorporate a wide class of linear processes, including the ARFIMA family of models that are the focus of this work.

Under the martingale difference structure for  $\varepsilon(t)$  imposed by Assumption 1, the linear predictor

$$\bar{y}(t) = \sum_{j=1}^{\infty} \pi(j)y(t-j) \quad (2.2)$$

is the minimum mean squared error predictor (MMSEP) of  $y(t)$ . The MMSEP of  $y(t)$  based only on the finite past is then

$$\bar{y}_h(t) = \sum_{j=1}^h \pi_h(j)y(t-j) \equiv - \sum_{j=1}^h \phi_h(j)y(t-j), \quad (2.3)$$

where we adopt the minor reparameterization from  $\pi_h$  to  $\phi_h$  in order to allow us, on also defining  $\phi_h(0) = 1$ , to write the corresponding prediction error as

$$\varepsilon_h(t) = \sum_{j=0}^h \phi_h(j)y(t-j). \quad (2.4)$$

The finite-order autoregressive coefficients  $\phi_h(1), \dots, \phi_h(h)$  can, in turn, be deduced from the Yule-Walker equations

$$\sum_{j=0}^h \phi_h(j)\gamma(j-k) = \delta_0(k)\sigma_h^2, \quad k = 0, 1, \dots, h, \quad (2.5)$$

where  $\gamma(\tau) = \gamma(-\tau) = E[y(t)y(t-\tau)]$ ,  $\tau = 0, 1, \dots$  is the autocovariance function of the process  $y(t)$ ,  $\delta_0(k)$  is Kronecker's delta (i.e.,  $\delta_0(k) = 0 \forall k \neq 0$ ;  $\delta_0(0) = 1$ ), and

$$\sigma_h^2 = E[\varepsilon_h(t)^2] \quad (2.6)$$

is the prediction error variance associated with  $\bar{y}_h(t)$  in (2.3).

The use of the optimal predictor  $\bar{y}_h(t)$  determined from the autoregressive model of finite order  $h$  is appropriate only if it is a good approximation to the ‘infinite-order’ predictor  $\bar{y}(t)$  for sufficiently large  $h$ . Poskitt (2007) addresses this very issue under regularity conditions that admit non-summable processes, proving the asymptotic validity, and properties, of finite-order AR models when  $h \rightarrow \infty$  with the sample size  $T$  at a suitable rate. In brief, the order- $h$  prediction error  $\varepsilon_h(t)$  converges to  $\varepsilon(t)$  in mean-square, the estimated sample-based covariances converge to their population counterparts – albeit at a slower rate than for a conventionally stationary process – and the least squares and Yule-Walker estimators of the coefficients of the approximating autoregression are asymptotically equivalent and consistent. Order selection by AIC is also shown to be asymptotically efficient in the sense of being equivalent to minimizing Shibata's (1980) figure of merit, something addressed in more detail in Section 4.1 below. It thus follows (see Poskitt, 2008), that the sieve bootstrap, which uses an estimated AR approximation to capture the dynamics of the process, is a plausible semi-parametric bootstrap technique for long-memory processes.

Motivated by the theoretical results in Poskitt et al. (2013) we, in turn, modify this ‘raw’ sieve approach by applying the sieve to ‘data’ pre-filtered via a suitable  $\sqrt{N}$ -consistent semi-parametric estimator, where  $N$  increases with  $T$  such that  $N/T \rightarrow 0$  as  $T \rightarrow \infty$ . Details of the both the raw and pre-filtered sieve bootstrap, including their relevant orders of accuracy are, as noted earlier, given in Poskitt et al. (2013). For convenience, we reproduce here the basic steps needed to implement the pre-filtered sieve bootstrap, denoted hereafter by PFSBS. A brief summary of the relevant convergence results then follows.

## 2.1 The PFSBS algorithm

Suppose that a value  $\hat{d}$  is available such that  $\hat{d} - d \in N_\delta = \{x : |x| < \delta\}$  where  $0 < \delta < 0.5$ . For any  $d > -1$  let  $\alpha_j^{(d)}$ ,  $j = 0, 1, 2, \dots$ , denote the coefficients of the binomial expansion of the fractional difference operator,

$$\begin{aligned} (1 - z)^d &= \sum_{j=0}^{\infty} \alpha_j^{(d)} z^j = 1 + \sum_{j=1}^{\infty} \left( \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} \right) z^j \\ &= 1 + \sum_{j=1}^{\infty} \left( \prod_{0 < k \leq j} \frac{k-1-d}{k} \right) z^j. \end{aligned}$$

Setting

$$w(t) = \sum_{j=0}^{t-1} \alpha_j^{(d)} y(t-j), \quad t = 1, \dots, T,$$

and using the preliminary estimate  $\hat{d}$ , pre-filtered sieve bootstrap realizations of  $y(t)$  are generated as follows:

**PFSBS1.** Calculate the coefficients of the filter  $(1 - z)^{\hat{d}}$  and from the empirical data generate the filtered values

$$\hat{w}(t) = \sum_{j=0}^{t-1} \alpha_j^{(\hat{d})} y(t-j), \quad t = 1, \dots, T.$$

**PFSBS2.** Fit an AR approximation to  $\hat{w}(t)$  and generate a sieve bootstrap sample  $\hat{w}^*(t)$ ,  $t = 1, \dots, T$ , of the filtered data as follows:

**SBS1.** Given the filtered series  $\hat{w}(t)$ ,  $t = 1, \dots, T$ , calculate the parameter estimates of the  $AR(h)$  approximation, denoted by  $\hat{\phi}_h(1), \dots, \hat{\phi}_h(h)$  and  $\hat{\sigma}_h^2$ , and evaluate the residuals,

$$\hat{\varepsilon}_h(t) = \sum_{j=0}^h \hat{\phi}_h(j) \hat{w}(t-j), \quad t = 1, \dots, T,$$

using  $\hat{w}(1-j) = \hat{w}(T-j+1)$ ,  $j = 1, \dots, h$ , as initial values. From  $\hat{\varepsilon}_h(t)$ ,  $t = 1, \dots, T$ , construct the standardized residuals  $\tilde{\varepsilon}_h(t) = (\hat{\varepsilon}_h(t) - \bar{\varepsilon}_h) / s_{\hat{\varepsilon}_h}^2$ , where  $\bar{\varepsilon}_h = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_h(t)$  and  $s_{\hat{\varepsilon}_h}^2 = T^{-1} \sum_{t=1}^T (\hat{\varepsilon}_h(t) - \bar{\varepsilon}_h)^2$ .

**SBS2.** Let  $\varepsilon_h^+(t)$ ,  $t = 1, \dots, T$ , denote a simple random sample of *i.i.d.* values drawn from

$$U_{\tilde{\varepsilon}_h, T}(e) = T^{-1} \sum_{t=1}^T \mathbf{1}\{\tilde{\varepsilon}_h(t) \leq e\},$$

the probability distribution function that places a probability mass of  $1/T$  at each of  $\tilde{\varepsilon}_h(t)$ ,  $t = 1, \dots, T$ . Set  $\varepsilon_h^*(t) = \hat{\sigma}_h \varepsilon_h^+(t)$ ,  $t = 1, \dots, T$ .



**SBS3.** Construct the sieve bootstrap realization  $\widehat{w}^*(1), \dots, \widehat{w}^*(T)$  where  $\widehat{w}^*(t)$  is generated from the autoregressive process

$$\sum_{j=0}^h \widehat{\phi}_h(j) \widehat{w}^*(t-j) = \varepsilon_h^*(t), \quad t = 1, \dots, T,$$

initiated at  $\widehat{w}^*(1-j) = \widehat{w}(\tau-j+1)$ ,  $j = 1, \dots, h$ , where  $\tau$  has the discrete uniform distribution on the integers  $h, \dots, T$ .

**PFSBS3.** Using the coefficients of the (inverse) filter  $(1-z)^{-\widehat{d}}$ , construct, for  $y(t)$ , the corresponding pre-filtered sieve bootstrap draw:

$$\widehat{y}^*(t) = \sum_{j=0}^{t-1} \alpha_j^{(-\widehat{d})} \widehat{w}^*(t-j), \quad t = 1, \dots, T.$$

Note that the process

$$(1-z)^{\widehat{d}} y(t) = \frac{\kappa(z)}{(1-z)^{d-\widehat{d}}} \varepsilon(t)$$

has fractional index  $d - \widehat{d}$ , where by assumption  $|d - \widehat{d}| < \delta$ . By [Poskitt et al. \(2013, Theorem 3\)](#) the error in the AR approximation to  $\widehat{w}(t)$  will accordingly be of order  $O(h(\ln T/T)^{1-2\delta})$  or smaller. That this level of accuracy is transferred to the pre-filtered sieve bootstrap realizations  $\widehat{y}^*(t)$  of  $y(t)$ , via the sieve bootstrap draws  $\widehat{w}^*(t)$  of  $\widehat{w}(t)$ , and hence to the pre-filtered sieve bootstrap approximation to the sampling distribution of any given statistic in a suitable class, rests upon [Poskitt et al. \(2013, Proposition 9\)](#). Specifically, for the broad class of statistics that satisfy Assumptions 4 and 5 in [Poskitt et al. \(2013\)](#), a class which includes the sample autocorrelations and impulse response functions considered in this paper, Theorem 11 in [Poskitt et al. \(2013\)](#) proves that the rate of convergence of the PFSBS algorithm is  $O(T^{-1+\beta})$  for all  $\beta > 0$  (and, hence, arbitrarily close to the rate achieved with simple random samples) for any pre-filtering estimate  $\widehat{d}$  such that  $|\widehat{d} - d| \ln T \rightarrow 0$  *a.s.* as  $T \rightarrow \infty$ . The raw bootstrap, on the other hand, involves the omission of steps PFSBS1 and PFSBS3, and the application of step PFSBS2 to the raw data  $y(t)$  rather than the pre-filtered series  $\widehat{w}(t)$ . As shown in Theorem 6 of [Poskitt et al. \(2013\)](#), the consequence of omitting the pre-filtering step is to reduce the rate of convergence of the algorithm to  $O(T^{-(1-d')+\beta})$  for all  $\beta > 0$ , where  $d' = \max\{0, d\}$ .

Clearly, the more accurate the preliminary estimate of  $d$ , the more useful the pre-filtering, in terms of yielding a filtered process for which the AR approximation is accurate. Given the non-parametric flavour of our approach, in the simulation exercise that follows we apply a PFSBS algorithm based on a pre-filtering value of  $d$  equivalent to the semi-parametric estimator of [Robinson \(1995\)](#), where the estimator

is constrained to lie in the stationary region. As a corollary of [Giraitis and Robinson \(2003, Lemma 5.8\)](#) this estimator satisfies  $P(|\hat{d} - d| \ln T > \epsilon) = o(N^{-p})$ , where  $p > 1/\epsilon$  and  $N$ , the bandwidth, satisfies  $T^\epsilon < N < T^{1-\epsilon}$  for some  $\epsilon > 0$ . As such, the almost sure limiting criterion required for the pre-filtering value holds and the  $O(T^{-1+\beta})$  convergence rate for the sieve method is attainable.

The exercise to be undertaken here is to use both the raw and pre-filtered methods to estimate the sampling distributions of the sample autocorrelation and impulse response coefficients and, in so doing, obtain an estimate of the relevant expectations. Those (estimated) expectations are, in turn, used to define an appropriate measure of bias and the relevant bias-adjusted statistic defined by subtracting the bias from the statistic in the usual way. As noted in the Introduction and discussed in more detail below, the reference value used in the formula for the bias, in any particular case, is produced via the particular model that is implicit in the (raw or pre-filtered) sieve method used to generate the bootstrap draws.

### 3 Properties of persistence measures for a fractional process

Prior to conducting the bias adjustment of the relevant persistence measures using the sieve-based techniques, we provide, in the following two sections, a precise definition of both measures and a brief summary of their known properties in the long memory setting. In the case of the sample autocorrelation function, we also include details of the analytical bias-adjustment methods that we use as comparators of our simulation-based approach.

#### 3.1 The sample autocorrelation function

Following [Hosking \(1996\)](#), we define the  $k^{\text{th}}$  sample autocorrelation coefficient as

$$\hat{\rho}(k) = \frac{\sum_{t=1}^{T-k} (y(t) - \bar{y}_T)(y(t+k) - \bar{y}_T)}{\sum_{t=1}^T (y(t) - \bar{y}_T)^2}, \quad (3.1)$$

where  $\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y(t)$ . [Hosking](#) derives the following results regarding the asymptotic distribution of  $\hat{\rho}(k)$ :

1. If  $-0.5 < d < 0.25$  then any finite subset of  $T^{1/2}(\hat{\rho}(k) - \rho(k))$ ,  $k = 0, 1, \dots, T-1$ , has a limiting multivariate Normal distribution with mean zero and covariances given by Bartlett's formula ([Hosking, 1996, Theorem 4\(iii\)](#)).

2. If  $d = 0.25$  then any pair of  $\frac{(T/\ln T)^{1/2}}{(1-\rho(k))}(\widehat{\rho}(k) - \rho(k))$ ,  $k = 0, 1, \dots, T-1$ , converge in probability and have a common  $N(0, 4(\frac{\lambda}{\gamma(0)})^2)$  limiting distribution where

$$\lambda = \{\sigma\kappa(1)\}^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}. \quad (3.2)$$

3. If  $0.25 < d < 0.5$  then  $\frac{T^{1-2d}}{(1-\rho(k))}(\widehat{\rho}(k) - \rho(k))$ ,  $k = 0, 1, \dots, T-1$  converges to the ‘modified Rosenblatt’ distribution. The cumulants of this distribution, for different values of  $d$ , are documented in Table 2 of [Hosking \(1996\)](#). Most notably, the mean of this limiting distribution is shown to be both substantially less than zero, for all  $d > 0.25$ , and larger in magnitude than the standard deviation for  $d > 0.35$ .<sup>4</sup>

In summary then, the characteristics of the sampling distribution of the sample autocorrelations vary throughout the parameter space for  $d$ . In particular, in cases where the true persistence in the process is high (Case 3 above), it is to be anticipated that the sample autocorrelation function will substantially underestimate the extent of this persistence. Further, in this case, an approximating normal distribution is inappropriate in terms of capturing sampling variation in the estimated autocorrelation coefficients. [Hosking \(1996\)](#) supplements these general distributional results with the following approximation to the asymptotic bias in  $\widehat{\rho}(k)$ ,

$$\text{Bias}[\widehat{\rho}(k)] \sim \frac{-\lambda}{d(1+2d)} \left\{ \frac{1-\rho(k)}{\gamma(0)} \right\} T^{2d-1}, \quad (3.3)$$

where  $\lambda$  is given in (3.2). This is seen to be negative for any  $-0.5 < d < 0.5$ .

The definition in (3.1) is, of course, only one of several closely-related estimators of  $\rho(k)$ . [Lee and Ko \(2009\)](#) instead define

$$r(k) = \frac{C(k)}{C(0)} = \frac{\frac{1}{T-k} \sum_{t=1}^{T-k} (y(t) - \bar{y}_{[1:T-k]})(y(t+k) - \bar{y}_{[k+1:T]})}{\frac{1}{T} \sum_{t=1}^T (y(t) - \bar{y}_T)^2}, \quad (3.4)$$

where  $\bar{y}_{[1:T-k]} = \sum_{t=1}^{T-k} y(t)/(T-k)$ , and  $\bar{y}_{[k+1:T]} = \sum_{t=k+1}^T y(t)/(T-k)$ ; and proceed to derive a closed-form expression for the bias of  $r(1)$  based on the much earlier work of [Marriott and Pope \(1954\)](#), in which, up to  $O(T^{-1})$ , the bias in the sample autocorrelation coefficients is shown to be

$$E(r(k)) = \frac{E[C(k)]}{E[C(0)]} \left[ 1 - \frac{\text{cov}[C(k), C(0)]}{E[C(k)] E[C(0)]} + \frac{\text{var}[C(0)]}{E^2[C(0)]} \right]. \quad (3.5)$$

[Newbold and Agiakloglou \(1993\)](#), in turn, evaluate (3.5) under a fractional noise process (produced by setting  $\kappa(z) = 1$  in (1.1)), as well as a corresponding expression for the standard deviation of  $\widehat{\rho}(k)$ , and tabulate the results for different values

<sup>4</sup> Note that Hosking’s symbol  $\alpha$  corresponds to  $1-2d$  in the terminology used in the current paper.

of  $d$ . The results demonstrate a distinct negative bias in the  $\widehat{\rho}(k)$ , for all values of  $k$  considered, and are consistent with the expectation – given the relevant distributional results of [Hosking \(1996\)](#) – that this bias is more pronounced the larger is  $d$ . The bias is also found to be more pronounced in the empirically relevant case considered here, in which the sample mean is used in the calculation of the sample autocorrelations, rather than in the artificial scenario in which the (unknown) zero mean is imposed. (See [Newbold and Agiakloglou](#) for more details.)

In [Lee and Ko \(2009\)](#), the expression in (3.5) is used to produce a closed-form representation of the bias of  $r(1)$  in terms of the true  $\rho(1) \dots \rho(T-1)$  which is exact to  $O(T^{-1})$ , compared with the  $O(T^{2d-1})$  accuracy of the [Hosking \(1996\)](#) result in (3.3).<sup>5</sup> [Lee and Ko](#) plot the ratio of the asymptotic bias (3.3) to their ‘first-order’ bias for different values of  $T$  and  $d$ , under the assumption of fractional noise, in which case the asymptotic measure is shown to underestimate the first-order measure for any  $d > 0$ , with the extent of this underestimation increasing rapidly with  $d$ . The authors proceed to use their expression, evaluated at a preliminary estimate of  $d$  (upon which this expression naturally depends) to bias correct  $r(1)$  and so produce a simple method of moments estimator of  $d$  based on that bias-corrected estimate of  $\rho(1)$ . As noted in footnote 2, no further use is made of the bias-corrected estimate of  $\rho(1)$ , nor are the sampling properties of this estimator itself examined.

Finally, the asymptotic normality that pertains in Case 1 enables, in principle, the derivation of an Edgeworth expansion. [Lieberman, Rousseau and Zucker \(2001\)](#) prove the validity of such an expansion for the normalized quantity,  $\sqrt{T}(\widehat{\rho}_0(k) - \rho(k))$ , where

$$\widehat{\rho}_0(k) = \frac{\sum_{t=1}^{T-k} y(t)y(t+k)}{\sum_{t=1}^T y(t)^2} \quad (3.6)$$

(i.e. the true zero mean is imposed), and where  $d < 0.1$  is required for convergence of the expansion. This second-order approximation is, in turn, evaluated in [Poskitt et al. \(2013\)](#) and used as a comparator for the sieve-based bootstrap estimates of the sampling distributions of selected autocorrelation coefficients. The sieve bootstrap method is shown to produce estimated sampling distributions for  $\widehat{\rho}_0(k)$  that are visually indistinguishable from those produced by the second-order Edgeworth expansion, in that part of the parameter space in which the latter is valid.

### 3.2 The impulse response function

As noted above, our focus is on the production and bias-correction of a semi-parametric estimate of the  $k^{\text{th}}$  impulse response coefficient  $\psi(k)$  defined in (1.1). The estimation procedure involves fitting an AR model of order  $h$  (to be determined) to  $y(t)$  and

<sup>5</sup>  $r(k)$  and  $\widehat{\rho}(k)$  are of course asymptotically equivalent; i.e.,  $r(k) = \widehat{\rho}(k) + O(T^{-1})$ , and hence an  $O(T^{-1})$  bias result applies in either case.

inverting, to produce  $\widehat{\psi}(k)$  as the  $k^{\text{th}}$  term in the expansion

$$\widehat{\psi}(z) = \widehat{\Phi}_h^{-1}(z) = \sum_{k=1}^{\infty} \widehat{\psi}(k)z^k, \quad (3.7)$$

where  $\widehat{\Phi}_h(z) = 1 + \widehat{\phi}_h(1)z + \widehat{\phi}_h(2)z^2 + \dots + \widehat{\phi}_h(h)z^h$ , and the  $\phi_h(j)$ ,  $j = 1, 2, \dots, h$  are estimated as described in Section 2. As documented in Baillie and Kapetanios (2012), use of this approach in the long memory setting yields more accurate estimates of the true impulse response coefficients than do certain mis-specified parametric methods, and may even compete well with correctly specified parametric methods for some parameter designs. However, a marked negative bias is still a characteristic of the semi-parametric estimates, as we also document below. Baillie and Kapetanios (2012) produce a bias-adjusted estimate of the IRF by using the bootstrap technique of Kilian (1998) to bias-adjust the estimated AR coefficients prior to inverting to them to produce the  $\widehat{\psi}(k)$ . In contrast, we bias correct the  $\widehat{\psi}(k)$  directly, as described in detail in the computational section that follows. As portent to our results, we find that the use of the *pre-filtered* sieve produces bias-adjusted statistics that are very accurate, and at smaller sample sizes than those documented in Baillie and Kapetanios (2012).

## 4 Simulation Exercise

In this section we examine the performance of the raw and pre-filtered sieve algorithms via a simulation experiment. Specifically, we investigate the finite sample accuracy of the (PF)SBS-based bias-adjusted estimates of the autocorrelation and impulse response coefficients, documenting both the (remaining) bias and mean squared error across Monte Carlo replications, as well as plotting selected sampling distributions. The results for the raw (unadjusted) statistics are also documented, in order to demonstrate the extent of the improvement yielded by the bias-adjustment techniques. As noted above, in the first-order autocorrelation case, we are able to compare the finite sample performance of the bootstrap-based method with that of an estimator based on the analytic bias expression of Lee and Ko (2009). For interest we also document selected results regarding the performance of the estimator of  $\rho(k)$  based on the asymptotically valid bias expression of Hosking (1996), as given in (3.3).

Finally, for reference we also consider the accuracy with which the bootstrap algorithms reproduce the ‘true’ (Monte Carlo) sampling distribution of the raw persistence statistics, in selected cases, as it is these bootstrap distributions that underlie the subsequent bias-adjustment.

## 4.1 Simulation design and computational details

Data are simulated from a zero mean Gaussian  $ARFIMA(1, d, 0)$  process,

$$(1 - L)^d \Phi(z)y(t) = \varepsilon(t), \quad 0 < d < 0.5, \quad (4.1)$$

with  $\Phi(z) = 1 - \phi z$  being the operator for a stationary AR(1) component and  $\varepsilon(t)$  is zero-mean Gaussian white noise. The process in (4.1) is simulated  $R = 1000$  times for  $d = 0.2$  and  $0.4$ ;  $\phi = 0.6$  and  $0.9$ , and sample sizes  $T = 100$  and  $500$ , via Levinson recursion applied to the autocovariance function of the desired  $ARFIMA(1, d, 0)$  process and the generated pseudo-random  $\varepsilon(t)$  (see, for instance, Brockwell and Davis, 1991, §5.2). The autocovariance function for given  $T$ ,  $\phi$  and  $d$  is calculated using Sowell's (1992) algorithm as modified by Doornik and Ooms (2001). Parameter settings are chosen that yield, respectively, moderate and large bias in both the estimated IRF and the estimated ACF.

For each realization  $r$  of the process we compute the relevant scalar statistic,  $s_{T,r}$ , plus  $B = 1000$  bootstrap estimates  $s_{T,r}^*(b)$ , constructed using  $b = 1, \dots, B$  bootstrap re-samples obtained via the (PF)SBS algorithm. Each realized value  $s_{T,r}$  thus has associated with it a 'bootstrap distribution' based on the  $B$  bootstrap resamples  $s_{T,r}^*(b)$ ,  $b = 1, \dots, B$ , with each such distribution serving as an estimate of the sampling distribution of  $s_T$ . In order to compare the  $R$  bootstrap distributions with the finite sample distribution estimated from the Monte Carlo draws, we first compute an 'average' bootstrap distribution by sorting the  $B$  bootstrap draws for each MC replication into ascending order, then average these ordered bootstrap values across the Monte Carlo draws. The  $B$  averaged draws are then used to produce a kernel density estimate, which we refer to as the average bootstrap distribution.

Our focus is on two types of statistic:  $s_T = \hat{\rho}(k)$ , computed as per (3.1), and  $s_T = \hat{\psi}(k)$ , computed as per (3.7), for  $k = 1, 2, \dots, 99$ ; and on using the sieve bootstrap techniques to bias-adjust each. Specifically, for any given realization  $r$ , the bootstrap distribution (computed from the  $B$  bootstrap resamples) is used to produce an estimate of  $E(s_T)$ ,  $\hat{E}(s_T)$ , and a bias-adjusted statistic,

$$s_{T,r}^{(BA)} = s_{T,r} - \hat{E}(s_T), \quad (4.2)$$

is thereby constructed. The sampling distribution of this statistic is then estimated from the  $R$  Monte Carlo draws using kernel density methods and the finite sample performance of the statistic as an estimator of the true parameter (function) summarized via bias and root mean square error (RMSE) computations. Each different form of sieve bootstrap algorithm produces a different estimate  $\hat{E}(s_T)$  and, hence, a different bias-adjusted statistic in (4.2).

With reference to the sieve method, as is common practice (Politis, 2003, §3), we begin by specifying the order of the autoregressive approximation as  $h = \hat{h}_T = \operatorname{argmin}_{h=0,1,\dots,M_T} \ln(\hat{\sigma}_h^2) + 2h/T$ , where  $\hat{\sigma}_h^2$  denotes the residual mean square obtained from an  $AR(h)$  model and  $M_T = \lceil (\ln T)^2 \rceil$ . Let  $\bar{h}_T = \operatorname{argmin}_{h=0,1,\dots,M_T} L_T(h)$  where  $L_T(h) = (\sigma_h^2 - \sigma^2) + h\sigma^2/T$  and  $\sigma^2$  and  $\sigma_h^2$  are as defined in 1 and (2.6) respectively. The function  $L_T(h)$  was introduced by Shibata (1980) as a figure of merit and the  $AR(\hat{h}_T)$  model is asymptotically efficient in the sense that  $L_T(\hat{h}_T) = L_T(\bar{h}_T)\{1 + o(1)\}$  *a.s.* as  $T \rightarrow \infty$  (Poskitt, 2007, Theorem 9). It follows that  $\hat{h}_T/\bar{h}_T \rightarrow 1$  *a.s.* as  $T \rightarrow \infty$ , so as  $T$  increases  $\hat{h}_T$  behaves almost surely like a deterministic sequence that satisfies the technical requirements of the theoretical convergence results in Poskitt et al. (2013).

For comparison, however, we also employ the fixed (for given  $T$ ) value of  $h = h_T = (\ln T)^2$  used in Baillie and Kapetanios (2012). Note that in the case of the IRF the alternative values for  $h$  are relevant not only in defining the order of the fitted AR in the sieve, and hence the bootstrap ‘data generating process’ from which the reference values used in the bias calculations (for both the IRF and the ACF) are backed out (about which more below);  $h$  also defines the order of the autoregression used to define the sample impulse response coefficients themselves (i.e., the impulse response statistics being bootstrapped and bias-adjusted).

As noted in the Introduction, the appropriate reference value for bias-adjustment of estimated persistence measures (IRF and/or ACF) is essentially that implied by the ‘bootstrap model’. Accordingly, the reference value for bias-adjustment in the case of the conventional SBS algorithm is simply that implied by the  $AR(h)$  sieve (where  $h$  may be  $\hat{h}_T$ ,  $h_T$ , or any other value that increases at the appropriate rate in  $T$ ) fitted to the raw data in step SBS1. Denoting this by

$$\bar{\Phi}_h(z) = 1 + \bar{\phi}_h(1)z + \dots + \bar{\phi}_h(h)z^h,$$

the reference IRF  $\tilde{\psi}(k)$  appropriate to the SBS is accordingly produced, for any  $k = 1, 2, \dots$ , by the inversion of  $\bar{\Phi}_h(z)$ , whilst the corresponding reference ACF  $\tilde{\rho}(k)$ ,  $k = 1, \dots, T-1$ , follows via the Yule-Walker equations.

The pre-filtered SBS method, on the other hand, implies an ARFIMA( $h, \hat{d}, 0$ ) bootstrap model, with  $\hat{d}$  the pre-filtering fractional integration parameter, and autoregressive coefficients  $\hat{\phi}_h(1), \dots, \hat{\phi}_h(h)$  produced by fitting an  $AR(h)$  to the filtered data  $(1-z)^{\hat{d}}y(t)$ . The reference IRF is therefore now obtained by inverting the implied ARFIMA( $h, \hat{d}, 0$ ) polynomial; i.e.,

$$\tilde{\psi}(z) = \hat{\Phi}_h^{-1}(z)(1-z)^{-\hat{d}},$$

(cf. 3.7) while the reference ACF is calculated by applying the Sowell/Doornik and Ooms algorithm to the implied ARFIMA( $h, \hat{d}, 0$ ) model.

Finally, we note that in order to produce bias-corrected estimates of  $\rho(k)$  that necessarily lie between zero and one we perform our bias-correction in terms of the so-called ‘‘Fisher- $z$ ’’ transformation, which maps from any  $r \in (-1, 1)$  to  $\zeta \in \mathbb{R}$  via

$$\zeta = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) = \operatorname{artanh}(r).$$

That is, while the statistic of interest is still  $\widehat{\rho}(k)$ , the bootstrap bias correction (and therefore the bootstrapping itself) is done in terms of  $\zeta(\widehat{\rho}(k))$ , with the bootstrap-bias-adjusted estimate of  $\rho(\cdot)$  produced via the reverse mapping

$$r = \frac{e^{2\zeta} - 1}{e^{2\zeta} + 1} = \tanh(\zeta).$$

It is the reverse mapping  $\zeta \rightarrow r$  that ensures that the bias-corrected result is within  $(-1, 1)$ .

For the ACF we also plot results for two additional bias-corrected quantities: one based on the subtraction of (3.3) from  $\widehat{\rho}(k)$ , with all unknown parameters in (3.3) assigned their true values from the data generating process; and the second (for the case of  $k = 1$  only) based on the subtraction of an estimate of the Lee and Ko (2009)  $O(T^{-1})$  bias expression from  $\widehat{\rho}(1)$ . The former (theoretical asymptotic-bias-adjusted ACF) is denoted by  $\widehat{\rho}^{(ASY)}(k)$ ; the latter (estimated  $O(T^{-1})$  bias-adjusted  $\widehat{\rho}(1)$ ) by  $\widehat{\rho}^{(LK)}(1)$ . The Lee and Ko bias is estimated by replacing the unknown population autocorrelations in their bias formula by the  $\tilde{\rho}(k)$ ’s implied by an  $AR((\ln T)^2)$  fitted to the (unfiltered) data. In other words, the Lee and Ko bias is calculated using the reference ACF corresponding to the raw SBS as described above, with  $h = (\ln T)^2$ .

## 4.2 Simulation Results

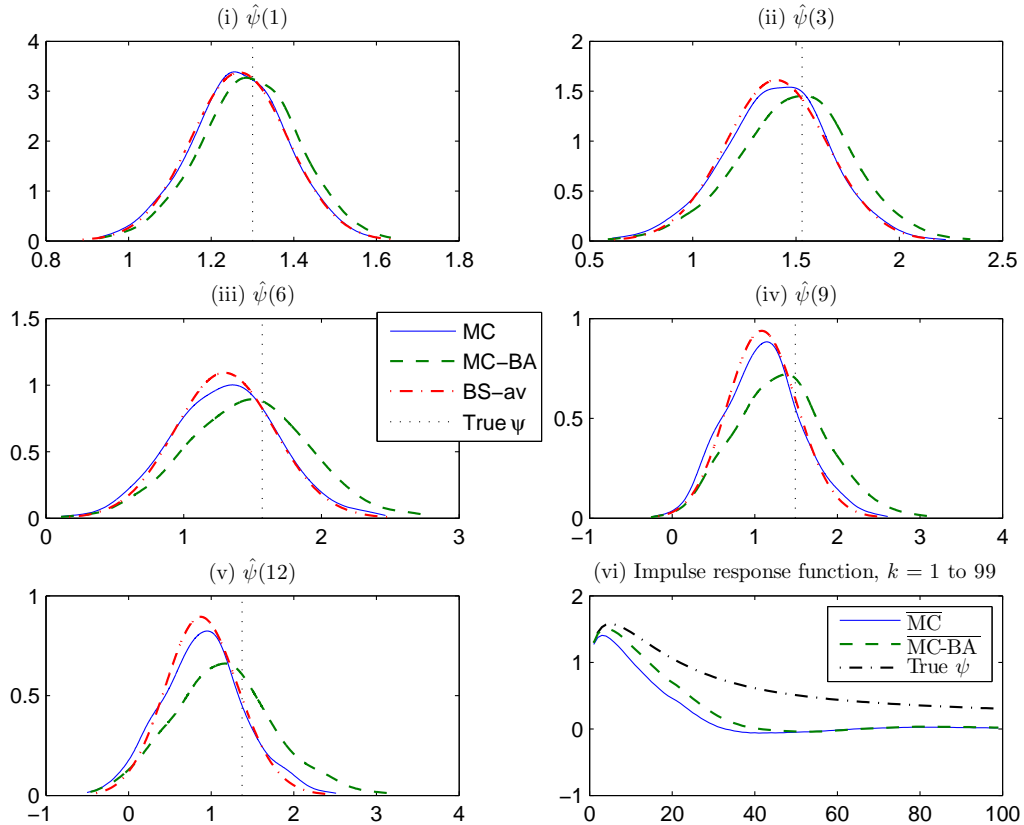
### 4.2.1 Bias-correction of the sample IRF

We begin by plotting selected distributional results for the sample IRF, where the bias adjustment occurs via the raw SBS algorithm. Figures 1 and 2 display results for  $T = 100$  and 500 respectively, with  $d = 0.4$  and  $\phi = 0.9$ . Both sets of results are for  $h = (\ln T)^2$ , with results for  $h$  chosen via the AIC criterion reproduced in the subsequent tables only.

Panels (i) to (v) in each figure plot the Monte Carlo distribution of the unadjusted statistic  $\widehat{\psi}(k)$ ; the averaged bootstrap estimate of the distribution of  $\widehat{\psi}(k)$ ; and the Monte Carlo distribution of the bootstrap-bias-adjusted statistic  $\widehat{\psi}^{(BA)}(k)$ . (These are indicated by the legend entries ‘‘MC’’, ‘‘BS-av’’, and ‘‘MC-BA’’ respectively). The vertical dotted line in each panel indicates the position of the true value of  $\psi(k)$  for each  $k = 1, 3, 6, 9, 12$ . Panel (vi) plots, for lags  $k = 1, 2, \dots, 99$ , the true IRF  $\psi(k)$  (based on the parameters of the true data generating process); the mean of the Monte

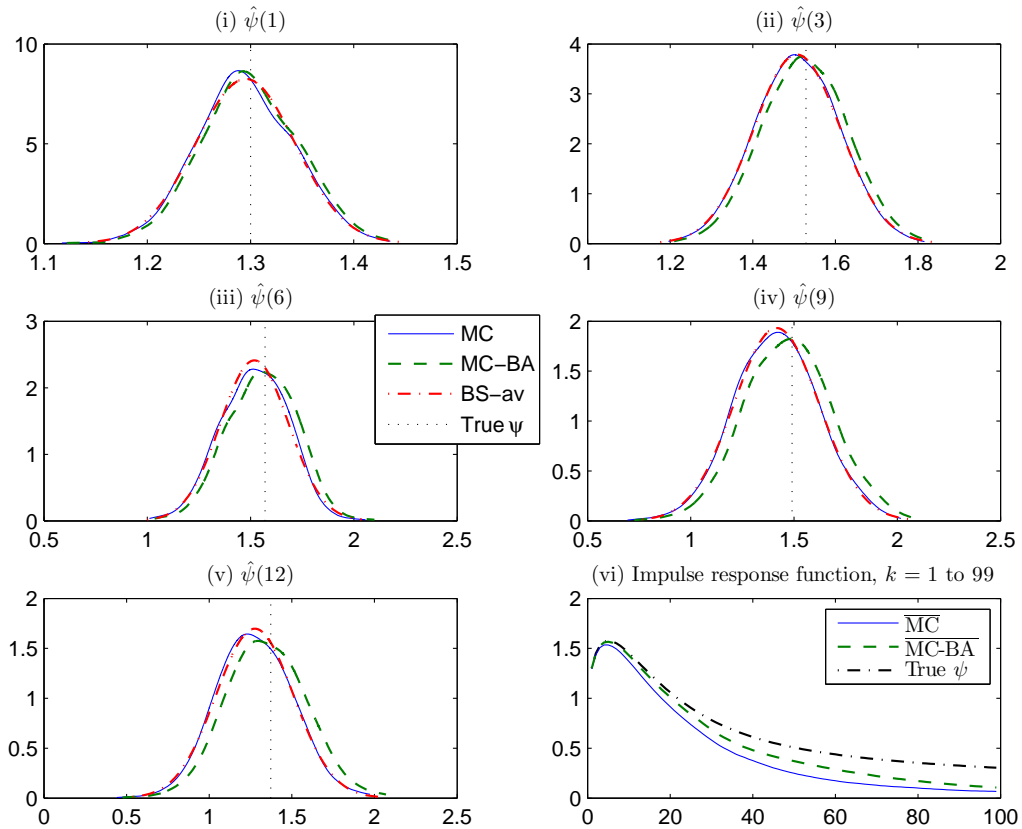


Carlo distribution of  $\hat{\psi}(k)$ ; and the mean of the Monte Carlo distribution of  $\hat{\psi}^{(BA)}(k)$  (designated “True  $\psi$ ”, “ $\overline{MC}$ ”, and “ $\overline{MC-BA}$ ” respectively).



**Figure 1.** Distributional results for estimates of the IRF under ARFIMA(1,  $d$ , 0) with  $T = 100$ ,  $d = 0.4$ ,  $\phi = 0.9$ . The BS-av and MC-BA distributions are here based on the (raw) sieve bootstrap.

The first thing to note from both Figures 1 and 2, and something that will be a feature of all graphs reproduced, is the accuracy with which the sieve (and, to an even greater extent, the pre-filtered sieve) technique reproduces the true sampling distribution of the statistic to be bias adjusted. This result is consistent with the supporting theoretical convergence results produced in Poskitt et al. (2013), and provides further justification for using the bootstrap-based estimate of the sampling distribution as a basis for estimating the bias of any given statistic, and bias-adjusting subsequently. As is clear from Figure 1 (for  $T = 100$ ), the small sample (negative) bias of  $\hat{\psi}^{(BA)}(k)$  that has been documented in Baillie and Kapetanios (2012) is in evidence here, for all lags  $k$ , with the extent of the bias increasing (in magnitude) with  $k$ , up to approximately  $k = 30$ , then levelling out thereafter to a fairly constant value. A qualitatively similar pattern is in evidence in Figure 2 (for  $T = 500$ ), but with the extent of the bias much less, for any particular value of  $k$ . The SBS-based bias adjustment is seen to produce quite effective results for the smaller lag lengths, but with no gains yielded after approximately  $k = 50$ , for the smaller sample size. For  $T = 500$ , in Figure 2, the

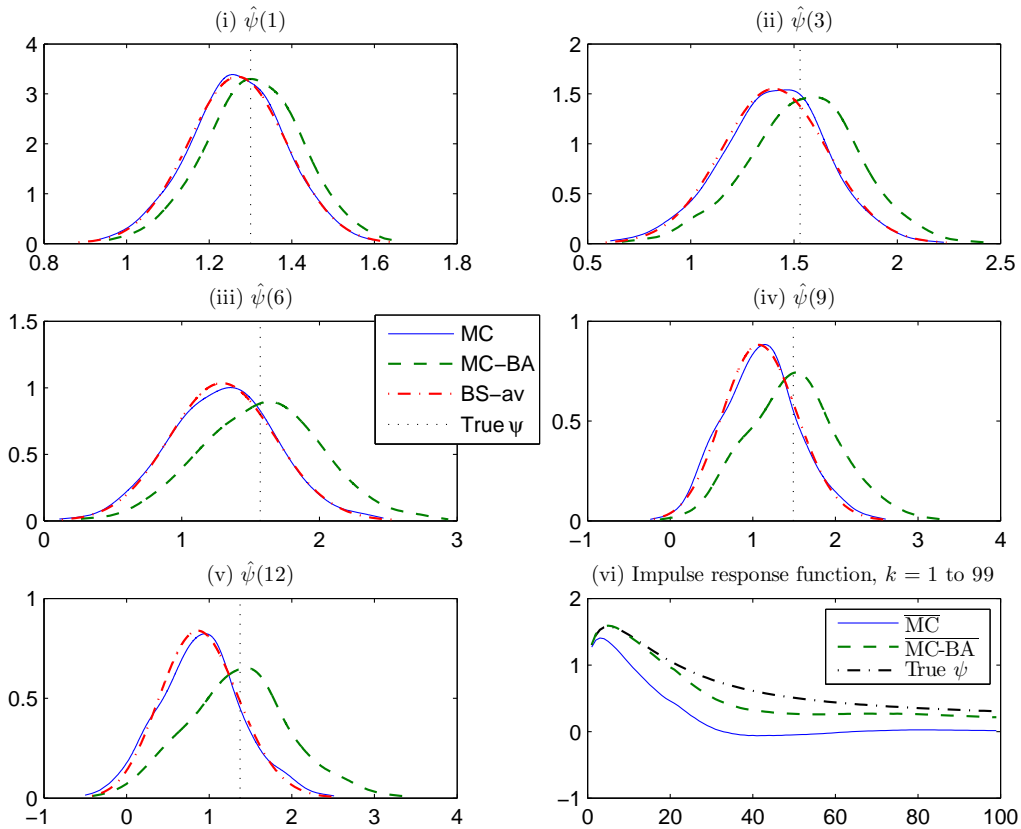


**Figure 2.** Distributional results for estimates of the IRF under ARFIMA(1,  $d$ , 0) with  $T = 500$ ,  $d = 0.4$ ,  $\phi = 0.9$ . The BS-av and MC-BA distributions are here based on the (raw) sieve bootstrap.

bootstrap technique is shown to produce a very accurate bias-adjusted estimator for low values of  $k$ , and to continue to yield improvements over the non-adjusted statistic for all values of  $k$  considered.

In Figures 3 and 4 we then plot the corresponding results, but now based on the pre-filtered bootstrap technique, with the *true* value of  $d$  used in the pre-filtering. The results provide resounding proof-of-concept support for the pre-filtering technique, with the TPFBS-based bias-adjusted estimator seen to be very accurate, even for the very small sample size, and to have a mean value (across Monte Carlo replications) that is almost visually indistinguishable from the true  $\psi(k)$  for all values of  $k$  considered when  $T = 500$ .

An empirically feasible version of the pre-filtering technique requires the substitution of an estimate of  $d$  in the PFSBS algorithm, with the (constrained) SPLW estimator of Robinson (1995) used for this purpose. As highlighted in Figures 5 and 6, we observe excellent bias-correction for the lower values of  $k$ , with the sampling distributions of the adjusted statistic (MC-BA in the graphs) located quite precisely with respect to the true value of the IRC in each case, and with very little cost in terms of additional dispersion. Note that, although we haven't included these figures

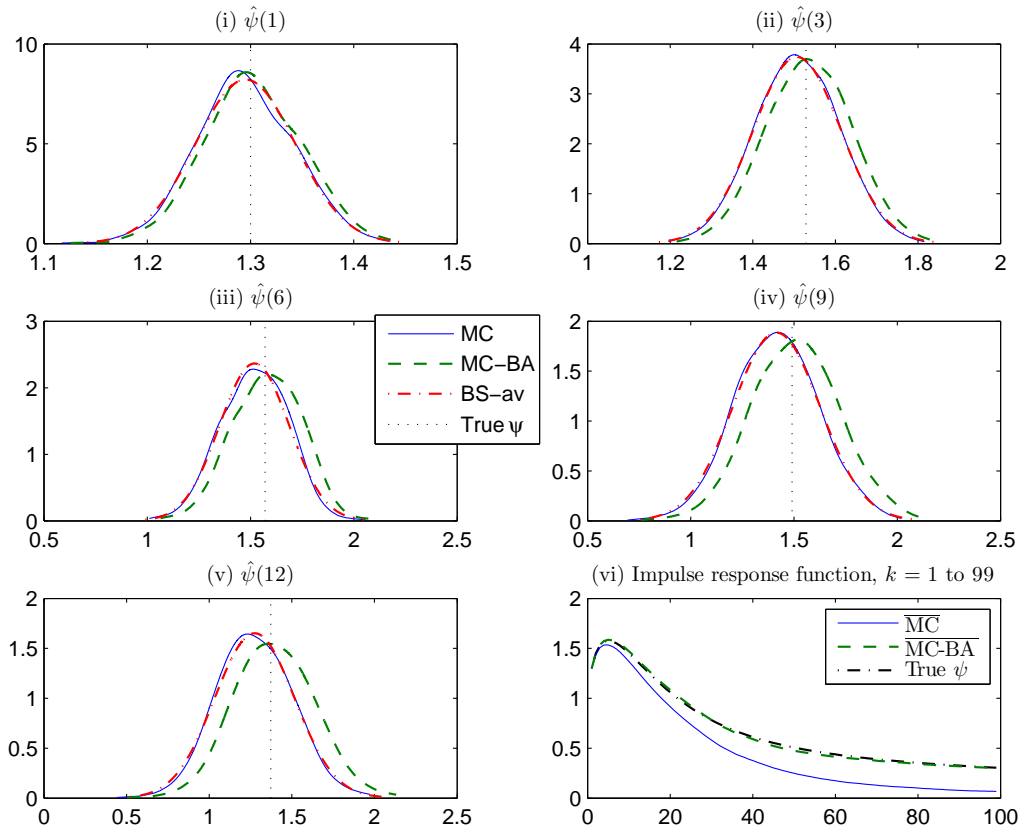


**Figure 3.** Distributional results for estimates of the IRF under ARFIMA(1,  $d$ , 0) with  $T = 100$ ,  $d = 0.4$ ,  $\phi = 0.9$ . The BS-av and MC-BA distributions are here based on the pre-filtered sieve bootstrap using the true value of  $d$  as the pre-filter (TPFBS).

here, for the medium persistence design ( $d = 0.2$ ,  $\phi = 0.6$ ) the SPLW-based PFSBS technique does tend to very slightly ‘over-correct’ for the longer lag lengths ( $k > 25$  for  $T = 100$ ;  $k > 45$  for  $T = 500$ ), where the TPFBS does not. Overall, however, very little accuracy is lost via the substitution of  $\hat{d}$  for  $d$ , with the bias-adjusted estimator remaining remarkably accurate, in particular for the larger sample size.

These selected graphical results are supplemented by the bias and root mean square error (RMSE) results recorded in Panel B of Tables 1 (for  $T = 100$ ) and 2 (for  $T = 500$ ), in which (Monte Carlo estimates of) these quantities (for  $d = 0.2, 0.4$ ;  $\phi = 0.6; 0.9$ ) are recorded for the raw, SBS-based bias adjusted, and PFSBS-based bias-adjusted statistics. We have not reported numerical results for the TPFBS-based estimator. Results are reported for  $k = 1, 6$  and  $12$ , with the average of all results (bias or RMSE) for  $k = 1, 3, 6, 9, 12$  also recorded under the column headed ‘av’.

Note that when bootstrapping the IRF we have set the order of the sieve approximation to be consistent with the order of the autoregression used to produce the IRF estimator being examined. That is, when  $\hat{\psi}(k)$  is produced via an autoregression with fixed order  $h = (\ln T)^2$ , the order of the sieve used in the bootstrap, whether raw or pre-filtered, is also set to  $h = (\ln T)^2$ . Similarly, when  $\hat{\psi}(k)$  is produced via an

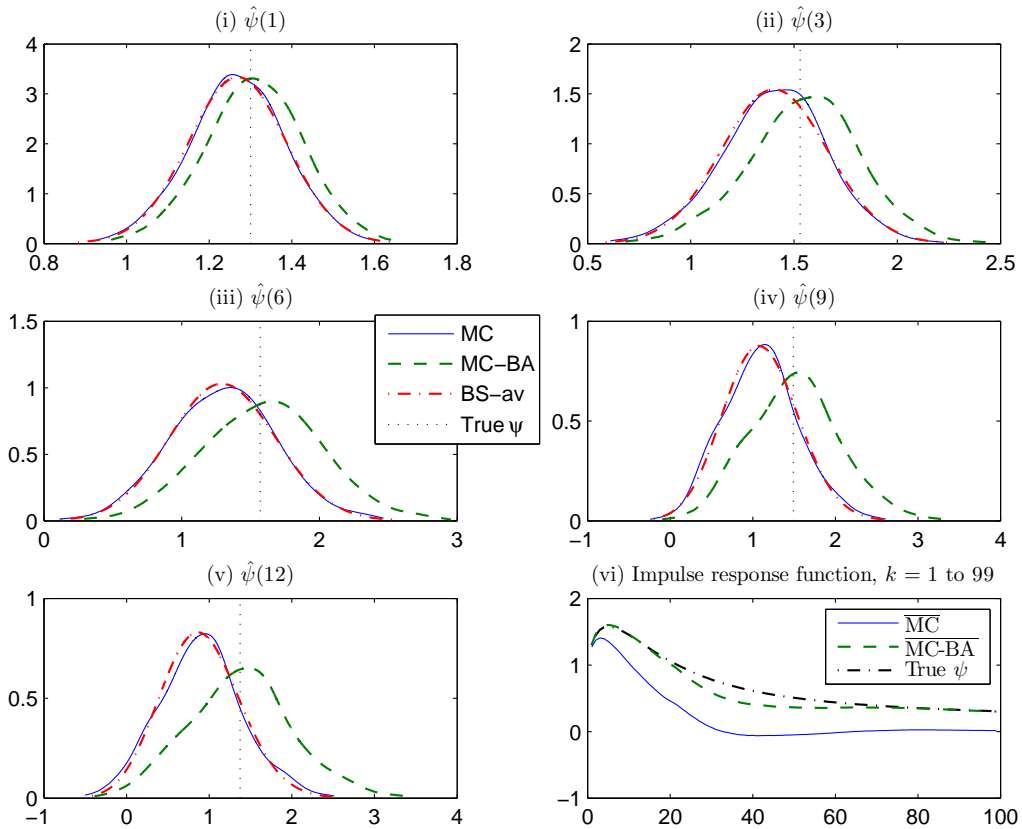


**Figure 4.** Distributional results for estimates of the IRF under ARFIMA(1,  $d$ , 0) with  $T = 500$ ,  $d = 0.4$ ,  $\phi = 0.9$ . The BS-av and MC-BA distributions are here based on the pre-filtered sieve bootstrap using the true value of  $d$  as the pre-filter (TPFBS).

autoregression with order selected by AIC, the order of the sieve used in the bootstrap is also selected by AIC. When using the raw sieve this naturally means that the sieve and estimating AR are exactly the same. However, this last is not the case when we switch to the PFSBS.

Beginning with the small sample size  $T = 100$ , and the ‘long AR’ ( $h = (\ln T)^2$ )-based estimator and bootstrap (Panel B of Table 1), we find that, relative to the unadjusted estimator the SBS-based bias-adjusted estimator is invariably superior in terms of bias for all values of  $d$  and  $\phi$  here considered, with bias reductions of 50-90%. The RMSE, however, is virtually identical, indicating the increased dispersion that generally accompanies bias correction based on an estimated measure of the bias. The bias of the PFSBS-based estimator is also invariably less than that of the unadjusted, and PFSBS tends to do better than SBS as  $k$  and/or the degree of persistence increases. Interestingly, PFSBS does rather worse than SBS in bias-adjusting  $\hat{\psi}(1)$ , perhaps because it so happens the SBS-based  $\hat{\psi}^{(BA)}(1)$  is essentially unbiased, but still improves on the unadjusted  $\hat{\psi}(1)$ .

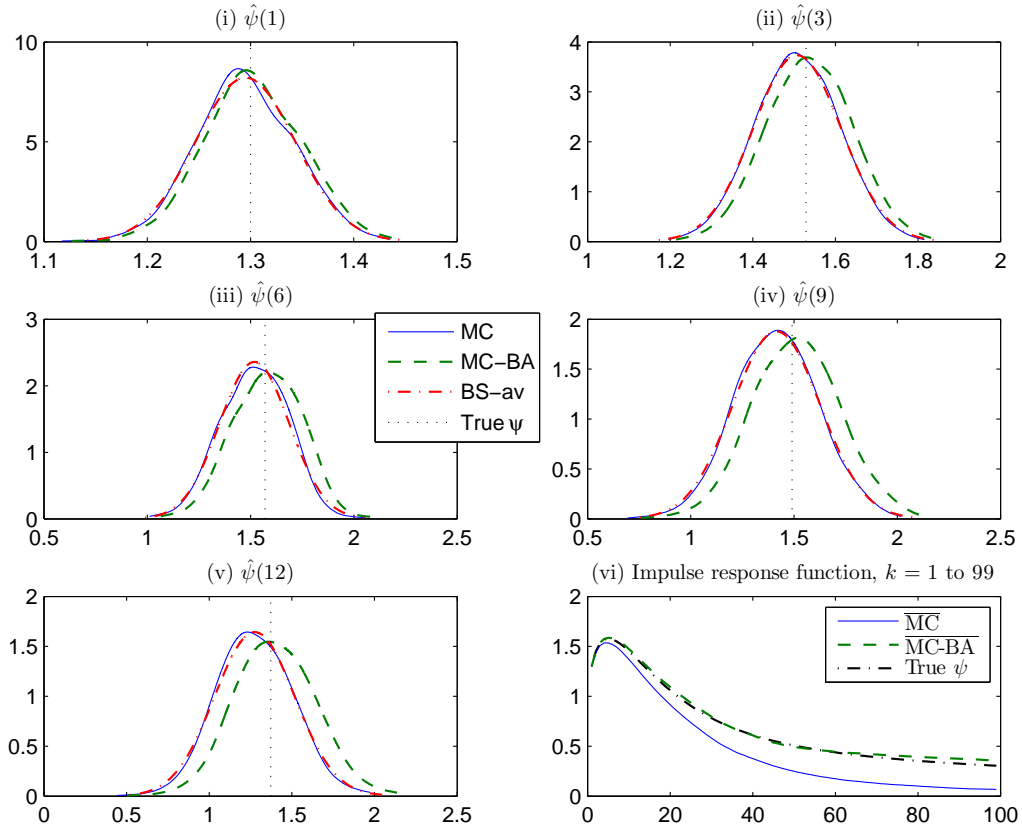
Turning to the larger sample size  $T = 500$ , the bias in  $\hat{\psi}(k)$  is of course much smaller to begin with, particularly for  $\hat{\psi}(1)$ , so very small effects have a large relative



**Figure 5.** Distributional results for estimates of the IRF under ARFIMA(1,  $d$ , 0) with  $T = 100$ ,  $d = 0.4$ ,  $\phi = 0.9$ . The BS-av and MC-BA distributions are here based on the pre-filtered sieve bootstrap using the SPLW estimate of  $d$  as the pre-filter (PFSBS).

impact. Nonetheless, for the estimator and sieve based on  $h = (\ln T)^2$  (Panel B of Table 2) we see that use of the SBS to bias adjust results in across-the-board bias reductions, essentially to zero, for the values of  $d$ ,  $\phi$ , and  $k$  considered, with the RMSE remaining virtually unchanged. Whilst the performance of the ‘feasible’ PFSBS for these relatively low lag values is slightly mixed relative to the ‘raw’ SBS, it generally results in an improvement relative to the unadjusted statistic, again doing worst for  $\hat{\psi}(1)$ , where there was almost no bias to correct in the first place. The PFSBS is evidently most advantageous relative to the SBS as the lag length  $k$  increases, with the result that the PFSBS produces a vast reduction in bias overall, relative to the raw SBS, when an extended spectrum of values for  $k$  is considered, as the graphical results recorded in Figures 1, 2, 5 and 6 highlight.

Results for the IRF estimator based on an autoregression with order  $h$  selected via AIC (i.e.,  $h = \hat{h}_{AIC}$ ) (Panel A of Tables 1 and 2) tell a qualitatively similar story. The unadjusted bias in the AIC-based estimator is comparable to that for  $\hat{\psi}(k)$  based on the longer (and fixed) value of  $h$  discussed above, tending to be slightly higher for  $\phi = 0.9$ ; though its RMSE is generally lower. For  $T = 100$ , relative to the unadjusted estimator the SBS-based bias-adjusted estimator is again uniformly superior in terms



**Figure 6.** Distributional results for estimates of the IRF under ARFIMA(1,  $d$ , 0) with  $T = 500$ ,  $d = 0.4$ ,  $\phi = 0.9$ . The BS-av and MC-BA distributions are here based on the pre-filtered sieve bootstrap using the SPLW estimate of  $d$  as the pre-filter (PFSBS).

of bias, with bias reductions of between 50 and 70%; the SBS-based estimator typically also has lower RMSE than the unadjusted  $\hat{\psi}(k)$ . For medium and high persistence the PFSBS does even better than SBS, with bias now typically around 20% of the original, and RMSE lower again. PFSBS-based adjustment can however do worse than the unadjusted estimator for the low-persistence design, for particular values of  $k$ .

With reference to the  $AR(\hat{h}_{AIC})$ -based results for the larger sample size (Panel A of Table 2), we find that the raw sieve generally still performs well, with two exceptions, both of which occur for  $d, \phi, k$  combinations for which the unadjusted estimator happens to be already effectively unbiased. The PFSBS does better as the lag-length increases, and best for high persistence ( $d = 0.4, \phi = 0.9$ ). The bias-adjusted RMSE, as before, is either comparable to the unadjusted, or somewhat improved; with the PFSBS resulting in a reduction of up to 17% in the higher persistence case. Indeed, for this high persistence setting, results (not reported) for the full set of  $k$  values  $1, \dots, 99$  and both sample sizes demonstrate a considerable reduction overall in bias for the PFSBS-based bias-adjusted estimator relative to the estimator based on SBS.

**TABLE 1**  
Estimates of impulse response: bias and RMSE,  $T = 100$

		$\widehat{\psi}(k)$			$\widehat{\psi}^{(BA)}(k)$ (SBS)			$\widehat{\psi}^{(BA)}(k)$ (PFSBS)					
		$k = 1$	$k = 6$	$k = 12$	av.	$k = 1$	$k = 6$	$k = 12$	av.	$k = 1$	$k = 6$	$k = 12$	av.
Panel A: $T = 100$ ; $h$ based on AIC selection													
$d$	$\phi$	Bias											
0.2	0.6	-0.0378	-0.0467	-0.0503	-0.0463	-0.0094	-0.0182	-0.0437	-0.0223	0.0049	0.0522	0.0973	0.0496
	0.9	-0.0406	-0.1974	-0.235	-0.1643	-0.0119	-0.0564	-0.0836	-0.0514	0.0260	0.0363	0.0343	0.0344
0.4	0.6	-0.0589	-0.1047	-0.1242	-0.0991	-0.0294	-0.0275	-0.0781	-0.0427	0.0179	0.0196	0.0423	0.0246
	0.9	-0.0143	-0.2984	-0.5071	-0.2734	0.0090	-0.0979	-0.2122	-0.0972	0.0092	0.0405	-0.0153	0.0175
RMSE													
0.2	0.6	0.1075	0.1201	0.0862	0.1107	0.1006	0.1292	0.0972	0.1157	0.1133	0.139	0.1328	0.1338
	0.9	0.1293	0.2998	0.3325	0.2635	0.1170	0.2604	0.3207	0.2391	0.1051	0.2826	0.3173	0.2448
0.4	0.6	0.1327	0.1982	0.1868	0.1792	0.1181	0.1981	0.1974	0.1772	0.1076	0.1926	0.1648	0.1657
	0.9	0.1153	0.4577	0.6471	0.4134	0.1138	0.3844	0.5536	0.3558	0.1065	0.3843	0.5374	0.3497
Panel B: $T = 100$ ; $h = (\ln T)^2$													
$d$	$\phi$	Bias											
0.2	0.6	-0.0216	-0.0673	-0.0627	-0.0548	-0.0005	-0.0238	-0.029	-0.0191	0.0127	0.0245	0.0335	0.0236
	0.9	-0.0264	-0.1667	-0.2416	-0.1482	-0.0024	-0.0538	-0.102	-0.0522	0.0132	0.0449	0.0627	0.0410
0.4	0.6	-0.0281	-0.1239	-0.1472	-0.1046	-0.0046	-0.0515	-0.0774	-0.0455	0.0095	0.0171	0.0191	0.0156
	0.9	-0.0318	-0.2682	-0.4737	-0.2571	-0.0061	-0.1028	-0.2288	-0.109	0.0096	0.0244	0.0063	0.0157
RMSE													
0.2	0.6	0.1165	0.1759	0.1709	0.1608	0.1172	0.1814	0.1871	0.1677	0.1181	0.186	0.1926	0.1711
	0.9	0.1173	0.3264	0.4039	0.2905	0.1173	0.3216	0.4072	0.2889	0.1174	0.322	0.4143	0.2907
0.4	0.6	0.1180	0.2481	0.2644	0.219	0.1175	0.2464	0.2729	0.2202	0.1176	0.2453	0.2711	0.2189
	0.9	0.1191	0.4675	0.6863	0.4298	0.1182	0.4472	0.6551	0.4118	0.1174	0.4329	0.6233	0.3963

TABLE 2

Estimates of impulse response: bias and RMSE,  $T = 500$ 

		$\widehat{\psi}(k)$			$\widehat{\psi}^{(BA)}(k)$ (SBS)			$\widehat{\psi}^{(BA)}(k)$ (PFSBS)					
		$k = 1$	$k = 6$	$k = 12$	av.	$k = 1$	$k = 6$	$k = 12$	av.	$k = 1$	$k = 6$	$k = 12$	av.
Panel A: $T = 500$ ; $h$ based on AIC selection													
$d$	$\phi$	Bias											
0.2	0.6	-0.0111	-0.0026	-0.0310	-0.0126	-0.0057	0.0056	-0.0282	-0.0063	-0.0190	-0.0140	0.0414	-0.0032
	0.9	-0.0020	-0.0614	-0.0610	-0.0447	0.0023	-0.0283	-0.0109	-0.0151	0.0027	-0.0106	-0.0224	-0.0091
0.4	0.6	-0.0089	-0.0011	-0.0219	-0.0092	-0.0043	0.0196	0.0002	0.0075	0.0016	-0.0278	0.0050	-0.0108
	0.9	0.0004	-0.0944	-0.1617	-0.0841	0.0048	-0.0521	-0.0728	-0.0399	0.0007	0.0184	-0.0097	0.0056
RMSE													
0.2	0.6	0.0453	0.0612	0.0532	0.0548	0.0438	0.0634	0.055	0.0561	0.0545	0.0543	0.0575	0.0582
	0.9	0.0472	0.1190	0.1304	0.1027	0.0471	0.1064	0.1246	0.0956	0.0443	0.1223	0.1388	0.1055
0.4	0.6	0.0525	0.0963	0.0995	0.0846	0.0513	0.1017	0.1048	0.088	0.0442	0.0991	0.0885	0.0817
	0.9	0.0464	0.1874	0.2524	0.1645	0.0467	0.1692	0.2148	0.1465	0.0438	0.1554	0.2280	0.1442
Panel B: $T = 500$ ; $h = (\ln T)^2$													
Bias													
0.2	0.6	-0.0030	-0.0130	-0.0139	-0.0105	0.0009	-0.0031	-0.0032	-0.0018	0.0034	0.0058	0.0085	0.0061
	0.9	-0.0036	-0.0296	-0.0469	-0.0266	0.0010	-0.0040	-0.0073	-0.0030	0.0039	0.0153	0.0280	0.0160
0.4	0.6	-0.0043	-0.0245	-0.0324	-0.0208	0.0001	-0.0078	-0.0108	-0.0061	0.0029	0.0061	0.0100	0.0067
	0.9	-0.0050	-0.0512	-0.0969	-0.0499	0.0001	-0.0128	-0.0265	-0.0122	0.0028	0.0122	0.0256	0.0137
RMSE													
0.2	0.6	0.0465	0.0725	0.0726	0.0664	0.0465	0.0727	0.0737	0.0668	0.0466	0.0737	0.0752	0.0677
	0.9	0.0465	0.1235	0.1558	0.1112	0.0464	0.1224	0.1548	0.1105	0.0466	0.1247	0.1603	0.1131
0.4	0.6	0.0466	0.0964	0.1051	0.0859	0.0465	0.0955	0.1044	0.0854	0.0466	0.0964	0.1060	0.0862
	0.9	0.0466	0.1684	0.2478	0.1559	0.0464	0.1642	0.2392	0.1517	0.0466	0.1658	0.2432	0.1536

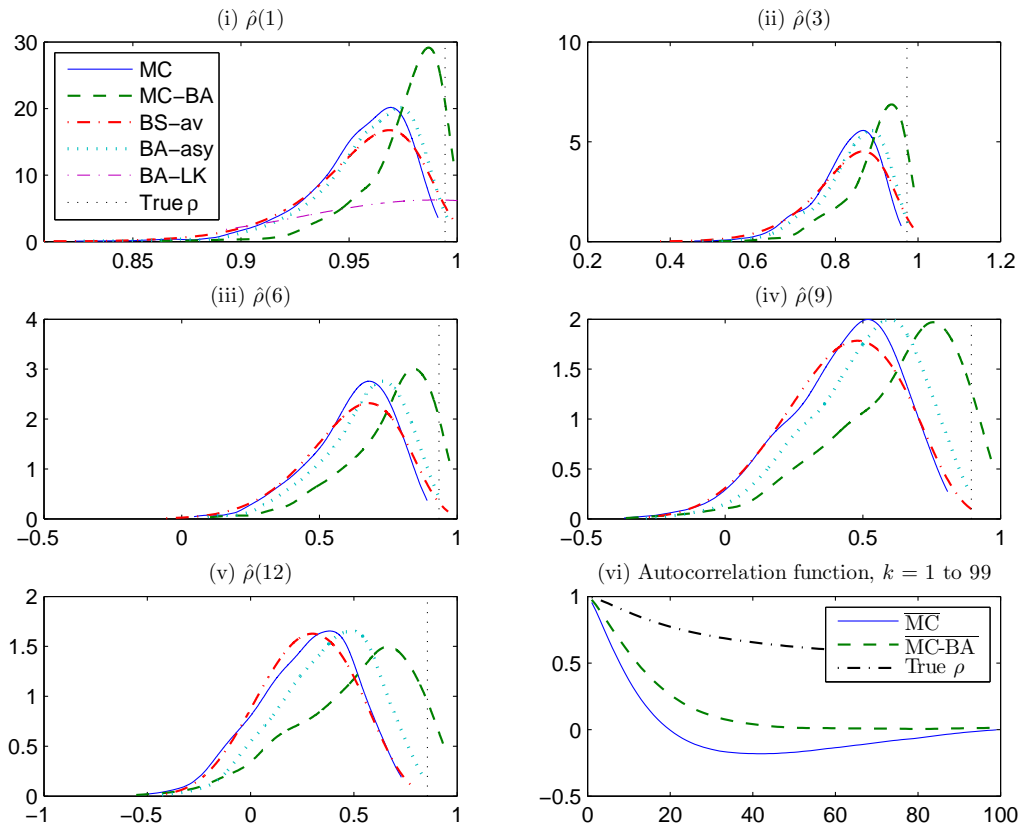


### 4.2.2 Bias-correction of the sample ACF

As in the previous section, we begin by plotting selected distributional results for the sample ACF, where the bias-adjustment occurs via the raw SBS algorithm. Figures 7 and 8 display results for  $T = 100$  and 500 respectively, with  $d = 0.4$  and  $\phi = 0.9$ . As was the case with IRF estimation we find that the qualitative results for ACF estimation are robust to the method by which  $h$  is selected, with there being no clear superiority of one set of results over the other. In this case we choose to present graphical results for the more conventional choice of  $h$ , based on AIC, with results for  $h = (\ln T)^2$  reproduced in the subsequent tables only.

Panels (i) to (v) in each figure plot respectively: the Monte Carlo distribution of the unadjusted statistic  $\hat{\rho}(k)$ ; the average bootstrap estimate of this ‘exact’ sampling distribution; the Monte Carlo distribution of the SBS-based bias-adjusted statistic (referred to as  $\hat{\rho}^{(BA)}(k)$ ); and the estimator adjusted using the (infeasible) asymptotic bias formula of Hosking (1996) (referred to as  $\hat{\rho}^{(ASY)}(k)$ ). As previously noted, the bootstrapping is performed in terms of the Fisher  $z$  transform of the ACF so as to restrict the bias-adjusted ACF to the  $(-1, 1)$  interval. The four plots are indicated by the legend entries “MC”, “BS-av”, “MC-BA” and “BA-asy” respectively. The vertical dotted line indicates the position of the true value of  $\rho(k)$  for each  $k = 1, 3, 6, 9, 12$ . Panel (vi) plots, for lags  $k = 1, 2, \dots, 99$ , the true ACF  $\rho(k)$  (based on the parameters of the true data generating process); the mean of the Monte Carlo distribution of  $\hat{\rho}(k)$ ; and the mean of the Monte Carlo distribution of  $\hat{\rho}^{(BA)}(k)$  (designated “True  $\rho$ ”, “ $\overline{MC}$ ”, and “ $\overline{MC-BA}$ ” respectively). In Panel (i) we also plot the sampling distribution of the feasible Lee and Ko bias-adjusted estimator (referred to as  $\hat{\rho}^{(LK)}(1)$ , and designated “MC-LK” on the figure).

Largely mimicking the results pertaining to the estimation of the IRF, the sieve-based technique reproduces quite accurately the ‘true’ Monte Carlo distribution of the statistic to be bias-adjusted. However, as Figure 7 demonstrates, and as has been documented elsewhere (see, for example, Hosking, 1996 and Poskitt et al., 2013), the conventional autocorrelation coefficient  $\hat{\rho}(k)$  is extremely biased, and none of the techniques considered here manage to completely eradicate that bias. The SBS bias-adjustment technique does, nevertheless, succeed in producing a statistic  $\hat{\rho}^{(BA)}(k)$  that is notably *less* biased, in particular for values of  $k$  in the medium lag range (20-60). Notably, the SBS-based technique produces an estimate of  $\rho(k)$  that is more accurate (for the recorded values of  $k$ ) than the analytically adjusted (using the known data generating parameters!) estimator,  $\hat{\rho}^{(ASY)}(k)$ , with  $\hat{\rho}^{(BA)}(k)$  being both less biased and having a much smaller RMSE than  $\hat{\rho}^{(ASY)}(k)$  on average. For the larger sample size, documented in Figure 8,  $\hat{\rho}^{(BA)}(k)$  matches the performance  $\hat{\rho}^{(ASY)}(k)$  quite closely, for all  $k$  values considered.

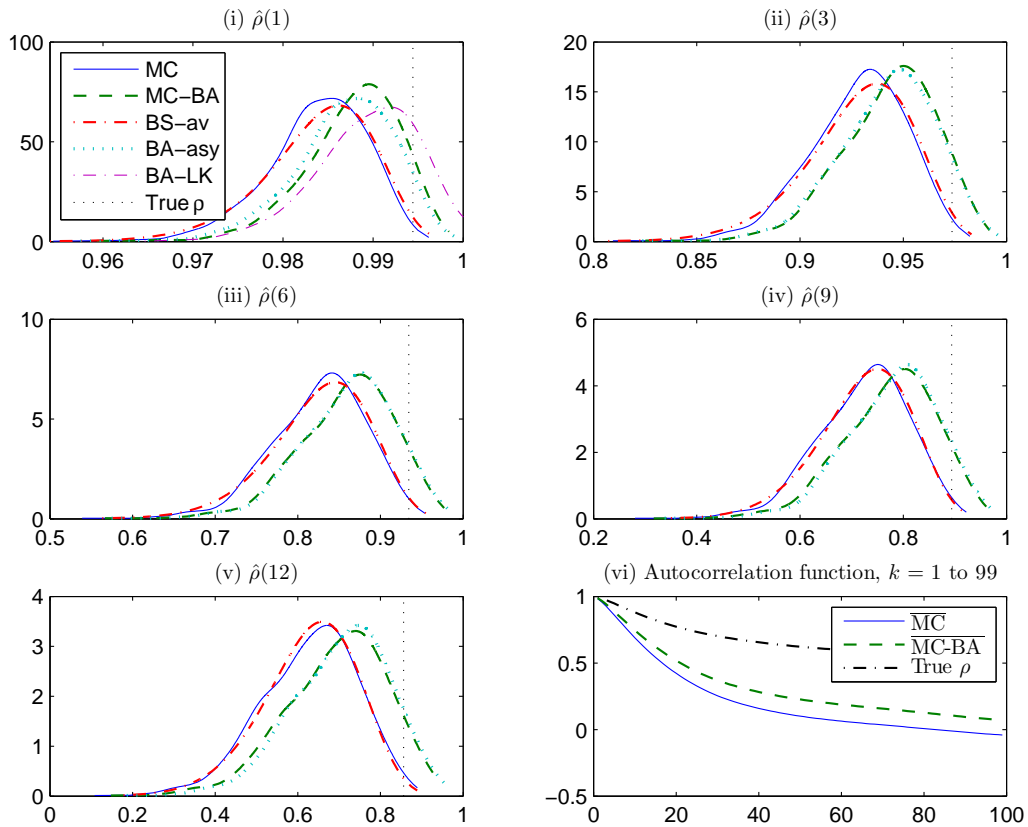


**Figure 7.** Distributional results for estimates of the ACF under ARFIMA(1,  $d$ , 0) with  $T = 100$ ,  $d = 0.4$ ,  $\phi = 0.9$ . The BS-av and MC-BA distributions are here based on the (raw) sieve bootstrap (SBS).

Comparing  $\hat{\rho}^{(BA)}(1)$  with the bias-adjusted estimator  $\hat{\rho}^{(LK)}(1)$  based on the estimated Lee and Ko bias, we see that our ‘plug-in’ estimate of the latter results in an estimator with slightly less bias than that of  $\hat{\rho}^{(BA)}(1)$ , but at the cost of a much larger RMSE. Results not reported here indicate that, depending on the precise data generating process, the sampling distribution of  $\hat{\rho}^{(LK)}(1)$  can exhibit very poor properties, including very large *positive* biases, due to the cumulative error in estimating the entire ACF, which we recall is required for implementation of any Lee and Ko-adjusted estimator.

In Figures 9 and 10 we plot the corresponding results based on the pre-filtered bootstrap technique, with the true value of  $d$  used in the pre-filtering. The results confirm, once again, the remarkable accuracy of this approach, with the TPFBS-based bias-adjusted estimator seen to be essentially unbiased, even for  $T = 100$ , and to have a mean value (across Monte Carlo replications) that is almost visually indistinguishable from the true  $\rho(k)$  for all values of  $k$  considered for  $T = 500$ .

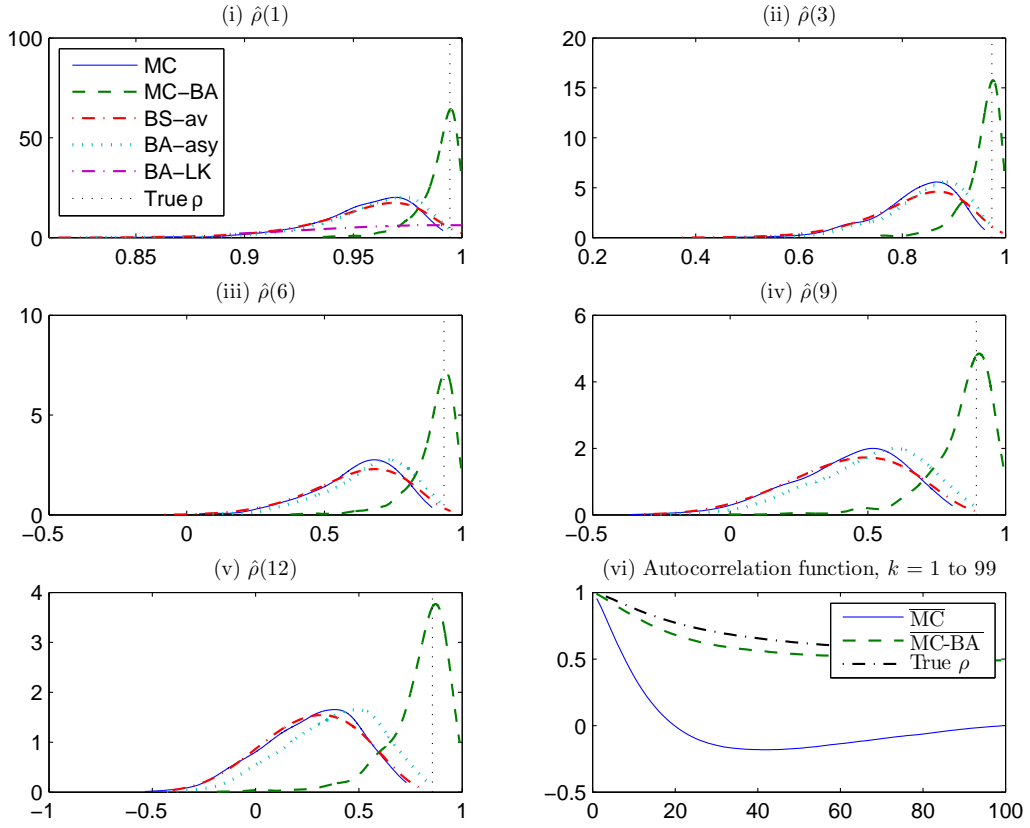
However, in contrast to the case for the IRF, rendering the pre-filtered technique feasible via the substitution of the SPLW estimate for  $d$  in the pre-filtering algorithm does not produce a bias-adjusted estimator whose performance mimics that of the



**Figure 8.** Distributional results for estimates of the ACF under ARFIMA(1,  $d$ , 0) with  $T = 500$ ,  $d = 0.4$ ,  $\phi = 0.9$ . The BS-av and MC-BA distributions are here based on the (raw) sieve bootstrap (SBS).

TPFBS-based version. Rather than reproducing graphs to highlight this fact, we simply alert the reader to the summary results in Tables 3 and 4. Careful investigation of the underlying outcomes indicates that the SPLW estimator is itself biased upwards, and that the bias in the SPLW estimator of  $d$  skews the reference value of  $\rho(k)$  in such a way that its use as a basis for calculating the bootstrap estimate of bias is severely compromised. Thus, despite the accuracy of the estimate of the sampling distribution of  $\hat{\rho}(k)$  as produced by the TPFBS, inaccuracy in the estimate of  $d$  can produce a reference value for use in the bias-correction that is itself an inaccurate representation of the true but unknown value of  $\rho(k)$  that underlies the data generating process. Hence, the PFSBS-based measure of bias is not an accurate estimate of the true unknown bias in  $\hat{\rho}(k)$ .

To understand this point it is worth remembering the situation that obtains for bias-adjustment in a parametric bootstrap setting. In that case, an unknown parameter  $\theta$  that characterizes the data generating process is estimated as  $\hat{\theta}$ . Repeated bootstrap samples are then generated from the estimated model (based on  $\hat{\theta}$ ), producing repeated bootstrap values,  $\hat{\theta}_b, b = 1, 2, \dots, B$ , and the bias of  $\hat{\theta}$ , defined as



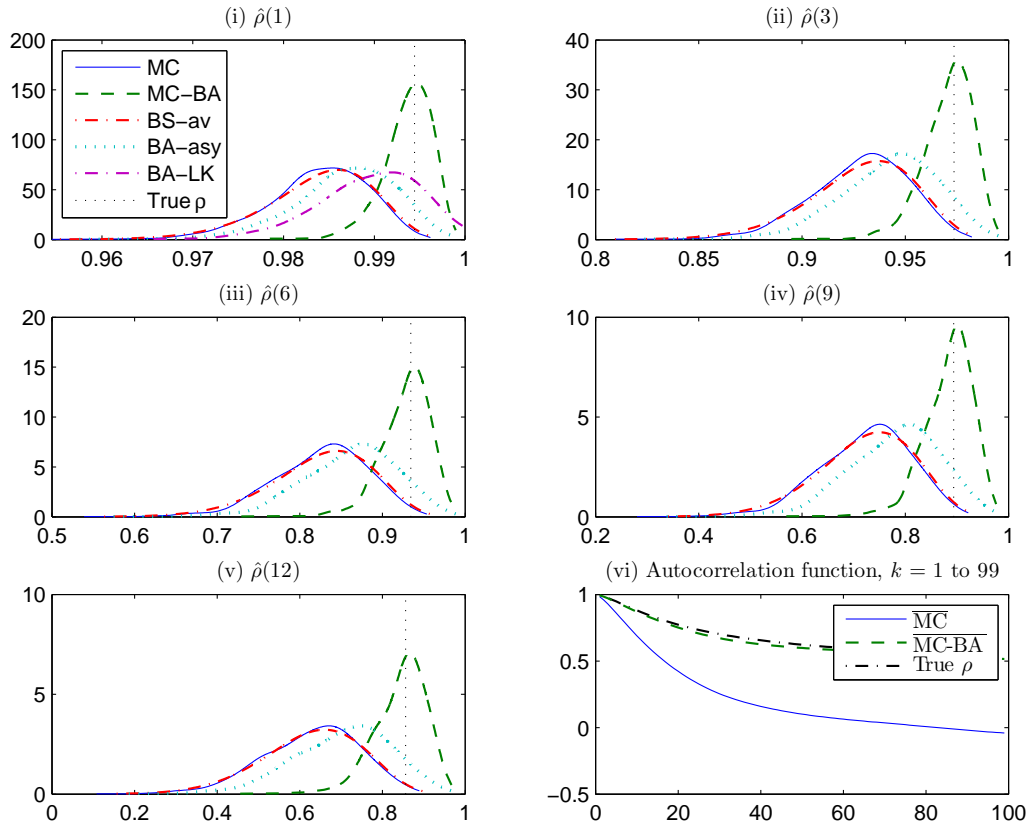
**Figure 9.** Distributional results for estimates of the ACF under ARFIMA(1,  $d$ , 0) with  $T = 100$ ,  $d = 0.4$ ,  $\phi = 0.9$ . The BS-av and MC-BA distributions are here based on the pre-filtered sieve bootstrap using the true value of  $d$  as the pre-filter (TPFBS).

$bias = E(\hat{\theta} - \theta)$ , is estimated by

$$\widehat{bias} = \sum_{b=1}^B \hat{\theta}_b - \hat{\theta}.$$

The key here is that  $\hat{\theta}$  plays exactly the same role in generating the bootstrap samples as does  $\theta$  in generating the empirical sample. In the case of bootstrapping the IRF or ACF using the (pre-filtered) sieve, however, the true data generating process is (by the very nature of the exercise) *not* estimated but, rather, approximated via the combination of an estimate of  $d$  and the long AR. The requisite parameter reference values to use in the bootstrap bias calculations therefore need to be backed out from the approximating model.

Now, whereas inaccuracies in the estimate  $\hat{d}$  appear to be compensated for by changes in the autoregressive estimates  $\hat{\phi}_h(1), \dots, \hat{\phi}_h(h)$  of the  $AR(h)$  approximation fitted to the filtered data  $(1-z)^{\hat{d}}y(t)$  in such a way that the reference IRF  $\tilde{\psi}(k)$  implicit in the  $ARFMA(h, \hat{d}, 0)$  approximating model provides a clear reflection of the true IRF coefficients; the same is not true of the ACF. A small amount of inaccuracy in the estimate of  $d$  produces an implied reference value  $\tilde{\rho}(k)$  that is sufficiently different



**Figure 10.** Distributional results for estimates of the ACF under ARFIMA(1,  $d$ , 0) with  $T = 500$ ,  $d = 0.4$ ,  $\phi = 0.9$ . The BS-av and MC-BA distributions are here based on the pre-filtered sieve bootstrap using the true value of  $d$  as the pre-filter (TPFBS).

from what would be produced by using the true (unknown) value of  $d$  to ultimately produce an inaccurate estimate of the true bias of  $\hat{\rho}(k)$ . The reason for this difference in sensitivity presumably lies in the fact that for any given values of  $\hat{d}$  and  $\hat{\phi}_h(1), \dots, \hat{\phi}_h(h)$  the reference values for the two different statistics are related via the expression  $\tilde{\rho}(k) = \sum_{s \geq k} \tilde{\psi}(s)\tilde{\psi}(k-s) / \sum_{s \geq 0} \tilde{\psi}^2(s)$ . This suggests that small perturbations in the  $\tilde{\psi}(k)$ , that are immaterial for the PFSBS bias correction of  $\hat{\psi}(k)$ , multiply and accumulate so as to result in a change in the value of  $\tilde{\rho}(k)$  that is sufficiently large to distort the PFSBS bias correction of  $\hat{\rho}(k)$ . The implication is that use of the PFSBS to bias correct the ACF requires a greater degree of precision in the preliminary estimate  $\hat{d}$  in order to achieve the high level of accuracy seen when employing the PFSBS to bias correct the IRF. Whilst it is beyond the scope of this paper to investigate this point further, we note that in related work (Poskitt, Martin and Grose, 2012) the authors are investigating the use of sieve-based techniques to bias-adjust  $d$  itself. It could be hoped that such a procedure may produce estimates of  $d$  that are accurate enough to alleviate the sensitivity problem observed here in the bias-adjustment of  $\hat{\rho}(k)$ . We leave that investigation for a later date.

**TABLE 3**

Estimates of autocorrelation: bias and RMSE,  $T = 100$

		$\widehat{\rho}(k)$			$\widehat{\rho}^{(B,A)}(k)$ (SBS)			$\widehat{\rho}^{(B,A)}(k)$ (PFSBS)					
		$k = 1$	$k = 6$	$k = 12$	av.	$k = 1$	$k = 6$	$k = 12$	av.	$k = 1$	$k = 6$	$k = 12$	av.
Panel A: $T = 100$ ; $h$ based on AIC selection													
$d$	$\phi$	Bias											
0.2	0.6	-0.0591	-0.1570	-0.1518	-0.1309	-0.0299	-0.0862	-0.0885	-0.0724	0.1727	0.6146	0.7291	0.5269
	0.9	-0.0453	-0.2570	-0.3835	-0.2324	-0.0198	-0.1144	-0.1774	-0.1051	0.0322	0.2403	0.4408	0.2356
0.4	0.6	-0.0959	-0.4249	-0.5256	-0.3629	-0.0673	-0.3150	-0.4072	-0.2724	0.0566	0.2899	0.3932	0.2537
	0.9	-0.0392	-0.3186	-0.5851	-0.3115	-0.0191	-0.1803	-0.3546	-0.1814	0.0055	0.0650	0.1424	0.0688
RMSE													
0.2	0.6	0.0919	0.2241	0.2143	0.1890	0.0780	0.1997	0.1891	0.1665	0.1948	0.6583	0.7672	0.5649
	0.9	0.0575	0.3079	0.4409	0.2747	0.0383	0.2181	0.3288	0.1983	0.0322	0.2403	0.4408	0.2356
0.4	0.6	0.1099	0.4638	0.5587	0.3940	0.0864	0.3809	0.4676	0.3246	0.0602	0.2972	0.3981	0.2596
	0.9	0.0454	0.3537	0.6285	0.3408	0.0270	0.2406	0.4540	0.2375	0.0055	0.0650	0.1424	0.0688
Panel B: $T = 100$ ; $h = (\ln T)^2$													
$d$	$\phi$	Bias											
0.2	0.6	-0.0591	-0.1570	-0.1518	-0.1309	-0.0314	-0.0969	-0.1062	-0.0824	0.1583	0.5749	0.6839	0.4920
	0.9	-0.0453	-0.2570	-0.3835	-0.2324	-0.0232	-0.1461	-0.2521	-0.1403	0.0319	0.2384	0.4373	0.2338
0.4	0.6	-0.0959	-0.4249	-0.5256	-0.3629	-0.0690	-0.3286	-0.4364	-0.2869	0.0551	0.2837	0.3855	0.2484
	0.9	-0.0392	-0.3186	-0.5851	-0.3115	-0.0209	-0.2056	-0.4238	-0.2116	0.0054	0.0637	0.1398	0.0675
RMSE													
0.2	0.6	0.0919	0.2241	0.2143	0.189	0.0796	0.2164	0.2270	0.1845	0.1928	0.6452	0.7491	0.5536
	0.9	0.0575	0.3079	0.4409	0.2747	0.0404	0.2358	0.3806	0.2206	0.0319	0.2384	0.4374	0.2338
0.4	0.6	0.1099	0.4638	0.5587	0.3940	0.0880	0.3963	0.5074	0.3426	0.0608	0.2955	0.3933	0.2578
	0.9	0.0454	0.3537	0.6285	0.3408	0.0286	0.2614	0.5145	0.2633	0.0054	0.0638	0.1400	0.0675

**TABLE 4**  
Estimates of autocorrelation: bias and RMSE,  $T = 500$

		$\widehat{\rho}(k)$			$\widehat{\rho}^{(B,A)}(k)$ (SBS)			$\widehat{\rho}^{(B,A)}(k)$ (PFSBS)					
		$k = 1$	$k = 6$	$k = 12$	av.	$k = 1$	$k = 6$	$k = 12$	av.	$k = 1$	$k = 6$	$k = 12$	av.
Panel A: $T = 500$ ; $h$ based on AIC selection													
$d$	$\phi$	Bias											
0.2	0.6	-0.0166	-0.0509	-0.057	-0.0437	-0.0100	-0.0327	-0.0394	-0.0286	0.0993	0.3311	0.3964	0.2875
	0.9	-0.0100	-0.0667	-0.1143	-0.0636	-0.0045	-0.0315	-0.0555	-0.0303	0.0322	0.2404	0.4409	0.2357
0.4	0.6	-0.0445	-0.2191	-0.2927	-0.1912	-0.0377	-0.1889	-0.2549	-0.1653	0.0577	0.2913	0.3935	0.2548
	0.9	-0.0106	-0.1085	-0.2283	-0.1128	-0.0066	-0.0752	-0.1623	-0.079	0.0055	0.0652	0.1428	0.0690
RMSE													
0.2	0.6	0.0336	0.1007	0.1065	0.0849	0.0311	0.0957	0.1011	0.0803	0.1249	0.4174	0.483	0.3578
	0.9	0.0146	0.0975	0.1661	0.0927	0.0114	0.0796	0.1381	0.0762	0.0322	0.2404	0.4409	0.2357
0.4	0.6	0.0497	0.2429	0.3211	0.2112	0.0439	0.2186	0.2912	0.1904	0.0583	0.294	0.3962	0.2570
	0.9	0.012	0.1224	0.2564	0.1270	0.0085	0.0946	0.2028	0.0991	0.0055	0.0652	0.1428	0.0690
Panel B: $T = 500$ ; $h = (\ln T)^2$													
$d$	$\phi$	Bias											
0.2	0.6	-0.0166	-0.0509	-0.0570	-0.0437	-0.0085	-0.0271	-0.0323	-0.0237	0.0741	0.2560	0.3096	0.2221
	0.9	-0.0100	-0.0667	-0.1143	-0.0636	-0.0047	-0.0335	-0.0603	-0.0326	0.0321	0.2397	0.4396	0.2350
0.4	0.6	-0.0445	-0.2191	-0.2927	-0.1912	-0.0354	-0.1760	-0.2370	-0.1540	0.0570	0.2882	0.3898	0.2522
	0.9	-0.0106	-0.1085	-0.2283	-0.1128	-0.0067	-0.0767	-0.1664	-0.0808	0.0055	0.065	0.1423	0.0688
RMSE													
0.2	0.6	0.0336	0.1007	0.1065	0.0849	0.0313	0.0978	0.1061	0.0827	0.1106	0.3752	0.4339	0.3209
	0.9	0.0146	0.0975	0.1661	0.0927	0.0116	0.0815	0.1446	0.0788	0.0321	0.2397	0.4396	0.235
0.4	0.6	0.0497	0.2429	0.3211	0.2112	0.0425	0.211	0.2818	0.184	0.058	0.2925	0.3939	0.2556
	0.9	0.012	0.1224	0.2564	0.127	0.0086	0.0964	0.2083	0.1014	0.0055	0.065	0.1423	0.0688

## References

- BAILLIE, R. and KAPETANIOS, G. (2012). Estimation and inference for impulse response functions from univariate strongly persistent processes. *Econometrics Journal*, **1** 1–26.
- BERAN, J. (1994). *Statistics for long-memory processes*, vol. 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York.
- BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*. 2nd ed. Springer Series in Statistics, Springer-Verlag, New York.
- CHOI, E. and HALL, P. G. (2000). Bootstrap confidence regions from autoregressions of arbitrary order. *Journal of the Royal Statistical Society*, **B 62** 461–477.
- DOORNIK, J. A. and OOMS, M. (2001). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis*, **42** 333–348. Also a 2001 Nuffield discussion paper.
- DOUKHAN, P., OPPENHEIM, G. and TAQQU, M. S. (eds.) (2003). *Theory and applications of long-range dependence*. Birkhäuser Boston Inc., Boston, MA.
- GIRAITIS, L. and ROBINSON, P. M. (2003). Edgeworth expansions for semiparametric Whittle estimation of long memory. *Annals of Statistics*, **31** 1325–1375.
- GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1** 15–29.
- HOSKING, J. R. M. (1980). Fractional differencing. *Biometrika*, **68** 165–176.
- HOSKING, J. R. M. (1996). Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long memory time series. *Journal of Econometrics*, **73** 261–284.
- KILIAN, L. (1998). Small-sample confidence intervals for impulse response functions. *The Review of Economics and Statistics*, **80** 218–230. URL <http://ideas.repec.org/a/tpr/restat/v80y1998i2p218-230.html>.
- KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, **17** 1217–1241.
- LEE, J. and KO, K. (2009). First-order bias correction for fractionally integrated time series. *Canadian Journal of Statistics*, **37** 476–493.



- LIEBERMAN, O., ROUSSEAU, J. and ZUCKER, D. M. (2001). Valid Edgeworth expansion for the sample autocorrelation function under long range dependence. *Econometric Theory*, **17** 257–275.
- MARRIOTT, F. H. C. and POPE, J. A. (1954). Bias in the estimation of autocorrelations. *Biometrika*, **41** pp. 390–402. URL <http://www.jstor.org/stable/2332719>.
- NEWBOLD, P. and AGIAKLOGLOU, C. (1993). Bias in the sample autocorrelations of fractional noise. *Biometrika*, **80** pp. 698–702. URL <http://www.jstor.org/stable/2337190>.
- POLITIS, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*, **18** 219–230.
- POSKITT, D. S. (2007). Autoregressive approximation in nonstandard situations: The fractionally integrated and non-invertible cases. *Annals of Institute of Statistical Mathematics*, **59** 697–725.
- POSKITT, D. S. (2008). Properties of the sieve bootstrap for fractionally integrated and non-invertible processes. *Journal of Time Series Analysis*, **29** 224–250.
- POSKITT, D. S., GROSE, S. D. and MARTIN, G. M. (2013). Higher order improvements of the sieve bootstrap for fractionally integrated processes. Tech. Rep. arXiv:1311.0096 [stat.ME], Monash University. URL <http://arxiv.org/abs/1311.0096>.
- POSKITT, D. S., MARTIN, G. M. and GROSE, S. G. (2012). Bias reduction of long memory parameter estimators via the pre-filtered sieve bootstrap. Econometrics & Business Statistics Working Paper WP 08/12, Monash University.
- ROBINSON, P. (ed.) (2003). *Time series with long memory*. Advanced texts in econometrics, Oxford University Press, Oxford [u.a.].
- ROBINSON, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, **23** 1630–1661.
- SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics*, **8** 147–164.
- SOWELL, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, **53** 165–188.