Misspecification and Bootstrap-based Inference in Non-Invertible and Fractionally Integrated Processes[†]

S. D. Grose and D. S. Poskitt[*]

Department of Econometrics and Business Statistics, Monash University

Abstract

This paper investigates the effect of model misspecification on the empirical properties of the sieve bootstrap when applied to the non-invertible and fractionally integrated processes considered in Poskitt (2006a). The behaviour of the sieve bootstrap is compared to that of equivalent "model-based" boostraps and the exact Edgeworth expansion where this is known, in the presence of various forms of misspecification.

AMS subject classifications: Primary 62M15: Secondary 62G05, 60G18

Key words and phrases: Misspecification, autoregressive approximation, fractional process, non-invertibility, sieve bootstrap.

Version: July 2, 2007

[†]Preliminary Draft: Not to be quoted without permission

[*] Corresponding author: Simone Grose, Department of Econometrics and Business Statistics, Monash University, Victoria 3800, Australia Tel.:+61-3-9905-9378; fax:+61-3-9905-5474.
E-mail address: Simone.Grose@Buseco.monash.edu.au

1 Introduction

The so-called "long memory", or strongly dependent, processes have come to play an important role in time series analysis. Long-range dependence, observed in a very wide range of empirical applications, is characterized by an autocovariance structure that decays too slowly to be summable. Specifically, rather than the autocovariance function declining at the exponential rate characteristic of a stable and invertible $ARMA$ process, it declines at a hyperbolic rate dependent on a "long memory" parameter $\alpha \in (0, 1)$; i.e.,

$$\gamma(\tau) \sim C\tau^{-\alpha}, C \neq 0, \text{ as } \tau \to \infty \, .$$

A detailed description of the properties of such processes can be found in Beran (1994).

Statistical procedures for analyzing such processes have ranged from the likelihood-based methods studied in Fox and Taqqu (1986), Sowell (1992) and Beran (1995), to the non-parametric and semi-parametric techniques advanced by Robinson (1995), among others. These techniques typically focus on obtaining an accurate estimate of the parameter (or parameters) governing the long-term behaviour of the process, and while maximum likelihood is asymptotically efficient in this context, that result as always depends on the correct specification of a parametric model.

The asymptotic theory for Maximum Likelihood estimation of the parameters of such processes is well established, at least under the assumption of Gaussian errors – see, for instance, Fox and Taqqu (1986), also Dahlhaus (1989). In small samples, however, the normal approximation can be quite (or very) poor, as numerous simulation studies (cite ...) have shown. Poskitt (2006b) proposed the use of the sieve bootstrap as a means of estimating the finite sample distribution in such cases, and briefly compared its properties with those of the Edgeworth expansion of Lieberman, Rousseau and Zucker (2003).

The paper proceeds as follows. Section 2 summarizes the statistical properties of long memory processes, and the properties of the sieve bootstrap in this context. In Section 3 we discuss the Edgeworth expansion of Lieberman et al. (2003). Details of the simulation study are given in Section 4, followed by the results presented in Section 5. Section 6 closes the paper.

2 Long-memory processes, Autoregressive approximation, and the Sieve Bootstrap

Perhaps the most popular model of such a process is the fractionally integrated ($I(d)$) process introduced by Granger and Joyeux (1980) and Hosking (1980). This class of processes can be characterized by the specification

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

where $\varepsilon(t)$ is zero-mean white noise, z is here interpreted as the lag operator ($z^j y(t) = y(t-j)$), and $\kappa(z) = \sum_{j\geq 0} \kappa(j) z^j$.

The behaviour of this process naturally depends on the fractional integration parameter d; for instance, if $d \geq 1/2$ the process is no longer stationary, although it may be made so by differencing. More pertinently, the impulse response coefficients of the Wold representation (2.2) characterized by $k(z)$ are now not absolutely summable for any $d > 0$; and the auto-covariances decline at the rate $\gamma(\tau) \sim C\tau^{2d-1}$ (i.e., $\alpha = 1 - 2d$). Such processes have been found to exhibit dynamic behaviour very similar to that observed in many empirical time series.

Nonetheless, if the "non-fractional" component $\kappa(z)$ is absolutely summable (i.e., $\kappa(z)$ is the transfer function of a stable, invertible ARMA process) and $|d| < 0.5$, then the coefficients of $k(z)$ are square-summable ($\sum_{j\geq 0} |k(j)|^2 < \infty$), in which case $y(t)$ is well-defined as the limit in mean square of a covariance-stationary process. The model is now essentially a generalization of the classic Box-Jenkins ARIMA model (Box and Jenkins, 1970),

$$(1-z)^d \Phi(z) y(t) = \Theta(z)\varepsilon(t) \tag{2.1}$$

in that we now allow non-integer values of the integrating parameter d.

The asymptotic theory for Maximum Likelihood estimation of the parameters of such processes is well established, at least under the assumption of Gaussian errors – see, for instance, Fox and Taqqu (1986), also Dahlhaus (1989). In particular, we have consistency, asymptotic efficiency, and asymptotic normality for the MLE of the fractional differencing parameter, so providing a basis for large sample inference in the usual manner.

However, a natural alternative to using large sample asymptotics is to bootstrap the finite sample distribution of the statistic of interest. Bootstrap methodology may be particularly appropriate in the fractional context because the maximum likelihood estimates are not even asymptotically pivotal.

Bootstrap methodology may be thought of as coming in two "flavours" – the model-based, or parametric, bootstrap, and a variety of non- or semi-parametric schemes. The parametric bootstrap resamples the residuals from estimation of the (assumed) parametric model, then uses these resampled residuals, and the parameter estimates associated with them, to simulate the process being estimated. The generated "data" is used to re-estimate the model parameters; repeatedly generating these "bootstrap" parameters builds a bootstrap distribution for them.

The main drawback of this procedure is that it relies on a correct parametric specification of the process. A less model-dependent approach nonetheless requires a re-sampling scheme that is able to capture the salient features of the data generating process, with, in the time series context, the dependence structure of the process being of prime importance. While the block bootstrap of Künsch (1989) has traditionally been employed for this purpose, blocking has been found to suffer from relatively poor convergence rates; for instance, the error in the coverage probability of a one-sided confidence interval derived from the block bootstrap is

$O(T^{-3/4})$, compared to the $O(T^{-1})$ rate achieved with simple random samples, where T is the sample size.

An attractive alternative is the "sieve" bootstrap of Bühlmann (1997). This works by "whitening", or filtering, the data using a finite-order autoregression, with the dynamics of the process captured in the fitted autoregression. Provided the order, h, of the autoregression increases at a suitable rate with T, the convergence rates for the sieve bootstrap are much closer (in fact arbitrarily close) to those for simple random samples. Choi and Hall (2000) showed that the error in the coverage probability of a one-sided confidence interval derived from the sieve bootstrap is $O(T^{\beta-1})$ for any $\beta > 0$, and they accordingly argue that for a linear process the sieve bootstrap is to be preferred.

One problem with application of the sieve bootstrap in the fractionally-integrated context is that its asymptotic properties have been derived assuming a certain rate of decay of the coefficients of the infinite-order autoregressive representation; see, for instance, Hannan and Deistler (1988). Non-invertible processes do not meet the necessary criteria, and neither do fractionally-integrated ones. However, Poskitt (2006a) derives asymptotic results for autoregressive modeling under regularity conditions that allow for both non-invertible and fractionally integrated processes, including uniform convergence rates for the sample autocovariances, and corresponding convergence rates for Yule-Walker and Least Squares estimates of the autoregressive coefficients. This immediately suggests the use of the sieve bootstrap in these "non-standard" contexts; and indeed Poskitt (2006b) shows that for an $I(d)$ process the sieve bootstrap achieves an error rate of $O(T^{\beta+d-1})$ for any $\beta > 0$ for a general class of statistics admitting an Edgeworth expansion.

2.1 Autoregressive approximation in the fractional context

Let $y(t)$ for $t \in \mathcal{Z}$ denote a linearly regular, covariance-stationary process,

$$y(t) = \sum_{j=0}^{\infty} k(j)\varepsilon(t-j) \tag{2.2}$$

where $\varepsilon(t)$, $t \in \mathcal{Z}$, is a zero mean white noise process with variance σ^2 and the impulse response coefficients satisfy the conditions $k(0) = 1$ and $\sum_{j\geq 0} k(j)^2 < \infty$.

Under Assumption 1 of Poskitt (2006a) the innovations $\varepsilon(t)$ conform to a classical martingale difference structure, and the linear predictor

$$\bar{y}(t) = \sum_{j=1}^{\infty} \varphi(j)y(t-j) \tag{2.3}$$

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

$$\bar{y}_h(t) = \sum_{j=1}^{h} \varphi_h(j)y(t-j) \equiv -\sum_{j=1}^{h} \phi_h(j)y(t-j); \tag{2.4}$$

where the minor reparameterization from φ_h to ϕ_h allows us, on also defining $\phi_h(0) = 1$, to conveniently write the corresponding prediction error as

$$\epsilon_h(t) = \sum_{j=0}^{h} \phi_h(j) y(t-j). \tag{2.5}$$

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

$$\sum_{j=0}^{h} \phi_h(j)\gamma(j-k) = \delta_0(k)\sigma_h^2\,, \quad k = 0, 1, \ldots, h, \tag{2.6}$$

in which $\gamma(\tau) = \gamma(-\tau) = E[y(t)y(t-\tau)]$, $\tau = 0, 1, \ldots$ 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

$$\sigma_h^2 = E\big[\epsilon_h(t)^2\big] \tag{2.7}$$

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

The use of finite-order AR models to approximate an unknown (but suitably regular) process therefore requires that the optimal predictor $\bar{y}_h(t)$ determined from the autoregressive model of order h be a good approximation to the "infinite-order" predictor $\bar{y}(t)$ for sufficiently large h.

The asymptotic validity, and properties, of finite order autoregressive models when $h \to \infty$ with the sample size T under regularity conditions which admit non-summable processes was proved in Poskitt (2006a). Briefly, the order-h prediction error $\epsilon_h(t)$ converges to $\varepsilon(t)$ in mean-square, the estimated sample-based covariances converge to their population counterparts, though 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. Furthermore, order selection by AIC is asymptotically efficient in the sense of being equivalent to minimizing Shibata's (1980) figure of merit, discussed in more detail in Grose and Poskitt (2006). The sieve bootstrap, with order selected via an asymptotically efficient criterion, is accordingly a plausible "non-parametric" bootstrap technique for long-memory processes.

3 The Edgeworth Expansion

Suppose we possess an estimator $\hat{\theta}_T$ of an unknown parameter θ_0 such that

$$\zeta_T = \sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

The Edgeworth expansion of the distribution function of ζ_T, if it exists, is the power series

$$P(\zeta_T \leq z) = \Phi(z) + T^{-1/2}p_1(z)\phi(z) + \cdots + T^{-j/2}p_j(z)\phi(z) + \cdots \tag{3.1}$$

where $\Phi(z)$, $\phi(z)$ are the standard Normal distribution function and density respectively, and the coefficients $p_j(z)$ are polynomials in z with coefficients depending on the cumulants of ζ_T – see, for instance, Hall (1992).

(3.1) is typically available only as an asymptotic expansion, with a remainder after truncation of order smaller than the last included term (i.e., $o(T^{-j/2})$), subject to certain regularity conditions. Establishing the "validity" of an Edgeworth expansion is accordingly a matter of proving the necessary conditions apply.

Lieberman et al. (2003) establish the validity of an Edgeworth expansion for the distribution of the MLE of the parameters of a stationary long memory Gaussian model with zero mean. The unknown mean case is covered in Andrews, Lieberman and Marmer (2006) for the estimator defined by maximizing the log-likelihood with the unknown mean replaced by a consistent estimator, such as the sample mean (referred to as the "plug-in" MLE, or PMLE). However, the results given only hold for $s = 3$ (i.e., the first-order expansion) in ARFIMA(p, d, q) models with $p > 0$ (Andrews et al., 2006, §4.3).

Results for the Whittle MLE (WMLE) are less restricted – Andrews and Lieberman (2005) prove the validity of an Edgeworth expansion for the distribution of the WMLE of the parameters of a stationary long memory Gaussian model with unknown mean μ and unit variance, and show that this applies for the general ARFIMA(p, d, q) to any order.

Unfortunately, while the expansions are known to be valid, computation of the terms of an expansion for estimates of the parameters of a general model is no trivial task; although, Lieberman and Phillips (2004) do derive a second order expansion for the distribution of the MLE of the fractional differencing parameter in the ARFIMA$(0, d, 0)$ model ("fractional noise") with zero mean, unit variance and $d > 0$. Terms in the first two derivatives of the log-likelihood (LLD's), required for computation of the coefficients of the expansion, are given in Lieberman, Rousseau and Zucker (2000).

4 Bootstrap procedures

4.1 Model-based Bootstrap

We assume a parametric specification of the process thought to be generating the data, here referred to as the "model", $\mathcal{M}_\theta$. The observed data is then supposed to be the result of applying the model to an unobserved innovations process, e, which may or may not be completely specified. We also assume an estimation procedure, yielding an estimator $\hat{\theta}$, and a statistic, or statistics, of interest, $s(\hat{\theta})$. The model-based bootstrap procedure is then: for given data vector y:

1. The model is estimated, returning estimates $\hat{\theta}$ and residuals $\hat{e}$.
2. The residuals are repeatedly resampled (B times); each T-vector of resampled residuals being combined with $\hat{\theta}$ and the model to simulate a corresponding "bootstrap" data vector, y_b. (i.e., $y_b^{(i)}$, $i = 1, \ldots, B$).

3. Each bootstrap data vector y_b is passed through the estimation procedure to obtain a corresponding bootstrap $\hat{\theta}_b$, and hence B realisations of $s(\hat{\theta}_b)$. The bootstrap distribution of $s(\hat{\theta})$ follows.

Note that if we are prepared to assume a known distribution for the innovations then Step 2 might be modified accordingly. For instance, if it is assumed that the innovations are Gaussian, then a "fully" parametric bootstrap would use the estimated innovations variance $\hat{\sigma}^2$ to generate T-vectors of "errors" according to $\mathrm{N}(0, \hat{\sigma}^2)$. These would take the place of the resampled residuals in step 2 above.

4.2 Sieve Bootstrap

The sieve bootstrap procedure is: for given data vector y:

1. The best approximating autoregression is selected and estimated, returning the autoregressive coefficients, $\hat{\phi}_h$, the MSE $\hat{\sigma}_h^2$, and the residuals, $\hat{e} = \hat{\Phi}(L)y$.
2. The filtered data is standardized using its own mean and standard deviation, then rescaled by $\hat{\sigma}_h$.
3. This rescaled residual vector is then repeatedly resampled; with each set of resampled residuals being passed back through the estimated AR(h) filter to obtain the "bootstrap" data y_b.
4. 'B' realization of the statistic of interest are obtained as described above.

Selection of "the best approximating autoregression" is according to some criterion, such as AIC; that is, all autoregressions of order 1 to some suitably large H_T are estimated, and the one resulting in the smallest AIC taken to be "optimal".

The other choice to be made in implementing the sieve bootstrap is that of the autoregressive estimator. The autoregressive filter used in step 1 must be stable, which would preclude the Least Squares estimator and suggest Yule-Walker; however Yule-Walker has been shown to exhibit substantial finite sample bias which might be expected to adversely affect its performance as a "whitening" tool. Poskitt (2006b) suggests the use of Burg's (1968) algorithm; this also guarantees a stable filter, but was shown in Grose and Poskitt (2006) to have finite sample bias very similar to that of Least Squares.

5 Simulation

We will initially focus our attention on the simplest of fractionally integrated processes, namely the fractional noise process

$$y(t) = \varepsilon(t)/(1-L)^d\,,\ 0 < d < 0.5\,, \tag{5.1}$$

with $\varepsilon(t)$ taken to be Gaussian white noise with unit variance.

The theoretical ACF for this processes is very simply computed, for $k \geq 1$, via the recursion

$$\gamma(k) = \gamma(k-1)\,\frac{k-1+d}{k-d},$$

initialized at $\gamma(0) = \Gamma(1-2d)/\Gamma^2(1-d)$; see (for instance) Brockwell and Davis (1991, §13.2). Knowledge of the ACF allows both simulation of the process itself and computation of the coefficients of the h-step ahead linear filter via Levinson recursion.

The process (5.1) is simulated $R = 1000$ times for $d = 0.1, 0.2, 0.3, 0.4$ and sample sizes $T = 40, 80, 160, 320$. The statistic of interest is the MLE ($\hat{d}$) of the fractional integration parameter d, estimated for each realisation by minimizing the negative of the usual Gaussian log-likelihood:

$$\ell(d; y) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}y'\Sigma^{-1}y\,, \tag{5.2}$$

where $\Sigma = \mathsf{toeplitz}(\gamma)$, and $\gamma = ACF(y)$.

The empirical distribution of $\hat{d}$ is obtained by using the R realized values for each statistic as the basis for a kernel density estimate of the associated distribution. We use the Gaussian kernel, with bandwidth equal to 75% of the Wand and Jones (1995) over-smoothed bandwidth; i.e.,

$$h = 0.75\,\sqrt[5]{\frac{243}{35R}}\;s(X)$$

where $s(X)$ is the empirical standard deviation of the R-element series X. This will be referred to as the "Monte-Carlo" or "simulated" distribution of $\hat{d}$, and is our estimate of the true distribution of the MLE, to which the bootstrap and Edgeworth distributions are compared.

For each realisation of the data vector y (and hence the MLE $\hat{d}$) we compute the corresponding bootstrap distribution of $\hat{d}$ based on $B = 500$ resamples, using both the model-based and sieve procedures as outlined above. The "average" bootstrap distribution is computed by taking a sample of 10,000 of the 500,000 bootstrap samples and obtaining the corresponding kernel density estimate. This (when plotted on top of the simulated distribution of $\hat{d}$) allows a visual comparison of the overall location of the bootstrap distributions with that of $\hat{d}$; however, it must be noted that the location of the individual bootstrap distributions varies with $\hat{d}$, and so the average *appears* to have a much higher variance than that of $\hat{d}$ itself. For this reason the bootstrap distributions must be "re-centered" by subtracting the corresponding value of $\hat{d}$; this removes the $\hat{d}$-related variability, and so gives an impression of the *average* dispersion of the bootstrap distributions for comparison with the distribution of $\hat{d}$. Neither average should be taken to be "the" bootstrap distribution of $\hat{d}$.

For each model (value of d) and sample size (value of T) we therefore have

- the empirical distribution of $\hat{d}$,
- the averaged bootstrap distribution (sieve and model-based)
- the averaged-recentered bootstrap distribution (sieve and model-based)

on which we plot:

- The distribution based on the first-order Edgeworth expansion of Lieberman and Phillips (2004, equation 13);
- The Lieberman & Phillips first-order Edgeworth expansion with "asymptotic" coefficients; and
- The Normal approximation.

Note that these Edgeworth expansions are for the distribution of the normalized MLE of the fractional integration parameter d in the zero mean fractional noise model; i.e., for $\hat{\delta} = \sqrt{T}(\hat{d} - d)$. However, as the asymptotic distribution of $\hat{\delta}$ is $\mathrm{N}(0, 6/\pi^2)$, we further standardize by $\sqrt{6}/\pi$, and plot the standard Normal ($\mathrm{N}(0,1)$) as an overlay. Accordingly, all distributions are plotted in terms of

$$\zeta_T = \pi\sqrt{\frac{T}{6}}(\hat{d} - d) \approx 1.28\hat{\delta}$$

with asymptotic distribution $\mathrm{N}(0,1)$.

To assess the effect of model misspecification on both the bootstrap procedures and the Edgeworth expansion, we repeat the experiment according to the following schemes:

	DGP		Model assumes	
Scheme	μ	σ	mean	variance
S1	0	1	0	1
S2	0.5	1	0	1
S3	1	1	0	1
S4	0	2	0	1
S5	0	1	unknown	1
S6	0	1	0	unknown
S7	0	1	unknown	unknown

The data generating process (DGP) is now taken to be $y(t) - \mu = \varepsilon(t)/(1-L)^d$, with $\varepsilon(t) \sim \mathrm{IN}(0, \sigma^2)$, and μ and σ as listed in the table. The "assumed" model is that actually estimated; i.e., for the first four schemes we assume $\mu = 0$, $\sigma = 1$, and so the MLE $\hat{d}$ is obtained by maximizing (5.2).

Scheme S1 is correctly specified. Schemes 5 – 7 are overparameterized – they estimate nuisance parameters that do not in fact appear in the DGP. Schemes 2 – 3 are underparameterized – the estimation phase in this case omits parameters that should in fact appear.

Schemes 1 – 4 all assume fractional noise with known mean and variance, and so the MLE $\hat{d}$ is obtained by maximizing (5.2). Schemes 5 – 7 (more realistically) assume one or other or both these parameters is unknown, and hence we maximize the log-profile likelihood discussed in the next section.

5.1 Note on the Log-likelihood

The log-likelihood of d in model (5.1) is (allowing for possibly non-zero mean and/or non-unit variance)

$$\ell(d, \sigma, \mu; y) = -\frac{T}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}z'\Sigma^{-1}z\,, \tag{5.3}$$

where $z = (y-\mu)/\sigma$ and $\Sigma = \mathsf{toeplitz}(\gamma)$; $\gamma = ACF(y)$.

If the variance is unknown we instead use the "profile" likelihood, in which σ^2 is replaced by its MLE, yielding

$$\ell_p(d, \mu; y) = c - \frac{1}{2}\ln|\Sigma| - \frac{T}{2}\ln\hat{\sigma}^2;\ \hat{\sigma}^2 = \frac{e'\Sigma^{-1}e}{T},\ e = y-\mu. \tag{5.4}$$

An unknown mean may similarly be replaced by its MLE, resulting in (once 5.4 is maximised) a corresponding MLE for d. This may be easily extended to regressors other than, or additional to, a simple constant; in which case $\hat{\sigma}^2$ is the mean-squared error from the regression of $\Sigma^{-1/2}y$ on $\Sigma^{-1/2}X$.

However, for the "constant-only" case a convenient alternative is to simply "de-mean" y prior to estimation; i.e., we replace μ in (5.4) by the sample mean $\bar{y}$. Since $\bar{y}$ is a consistent estimator of μ, the resulting estimator of d, often referred to as the "plug-in" estimator, or PMLE, is also consistent, asymptotically normal, and asymptotically efficient (Dahlhaus, 1989). In any case, given a means of computing the ACF for given d , both $|\Sigma|$ and $\hat{\sigma}^2$ are most efficiently computed via the Durbin-Levinson algorithm.

6 Results

Beginning with the correctly specified, zero mean, unit variance fractional noise model (schemes B1, M1), we find that for $d = 0.1$ the Edgeworth distributions are reasonable approximations for the true distribution of $\hat{d}$, although their dispersion is slightly low, and they 'miss' the tails quite badly (especially the right-hand tail, even for $T = 160$). The bootstrap distributions are not quite correctly centered, the sieve having a small negative bias (on average), and the model-based a small positive bias.

As d increases the bootstrap distributions shift rightwards. This means that for $d = 0.2$ the bias in the sieve bootstrap is negligible, while that in the model-based bootstrap is worse, while for $d = 0.3$ we see a slight rightward bias in the sieve bootstrap, and even more bias in the model-based. The Exact Edgeworth approximation is still quite good, except in the tails for the smaller sample sizes ($T = 40, 80$).

However, for $d = 0.4$ the asymptotic Edgeworth and Normal approximations are now very poor. The exact Edgeworth does much better, and is still quite good for $T = 160$. There is also a peculiar (negative probability) dip in the right-hand tail, although this is outside the actual parameter space. The averaged bootstrap distributions are now highly negatively skewed; indicating a "pile up" of distributions against the right-hand ($d = 0.5$) boundary. This appears to be the result of the 'flatness' of the likelihood function for d when

the true d is close to 0.5. This effect does not diminish as T increases (in fact it actually gets worse, particularly for the sieve bootstrap). Re-centering with respect to $\hat{d}$substantially reduces this effect, although some rightwards "bias" still remains.

To summarize: for a correctly specified fractional noise model the exact Edgeworth expansion gives a good approximation for $d < 0.25$ and reasonable for $d > 0.25$, for all sample sizes considered here. The asymptotic-Edgeworth is reasonable for $d = 0.1$, even for $T = 40$; but very poor for $d > 0.25$. The averaged sieve bootstrap seems to be "located" correctly for $d \leq 0.3$, but for $d = 0.4$ the distribution tends to "pile up" against the right-hand ($d = 0.5$) boundary.

6.1 Over-specified models

We continue to assume a data-generating process consisting of zero mean, unit variance fractional noise, but now suppose that the practitioner has chosen to estimate the mean, or the variance, or both.

Zero mean, unit variance fractional noise model: estimated mean

$d = 0.1$: Estimation of the mean has resulted in a leftward (negative) bias in $\hat{d}$ of approximately half an asymptotic standard deviation unit (ASDU). The sieve bootstrap distributions are correctly located wrt the distribution of $\hat{d}$, and re-centering shows them to have around the correct variance on average. Re-centering the average model-based bootstrap, on the other hand, seems to indicate that its variance is lower than it should be. The Edgeworth distributions have not of course "followed" the true distribution, and so are now incorrectly located, as well as having a variance that appears too low.

$d = 0.2, 0.3$: As for $d = 0.1$, except that the sieve bootstrap distribution variance is now also lower than it should be. The negative bias in $\hat{d}$ is now around 1 ADSU for $d = 0.3$, and the distribution of $\hat{d}$ is substantially negatively skewed.

$d = 0.4$: The Edgeworth expansions are now extremely inaccurate, as the true distribution has shifted further to the left (negative bias now around 1.5 ADSU), and become more skewed and perhaps more variable as well. The bootstrap distributions have again "followed" the actual distribution, with (as usual) the sieve bootstrap doing slightly better than the model-based in this regard. On re-centering, however, we find that neither bootstrap distribution has sufficient variability (this would cause lower than nominal coverage probabilities). Again, the sieve seems to do better than the model-based. Oddly enough, the "pile-up" on the right-hand boundary now longer occurs.

Zero mean, unit variance fractional noise model: estimated variance

The mean is now correctly taken to be zero, but the variance is estimated. Results are indistinguishable from those for the case where the variance is taken (correctly) to be one; this was because estimating the variance resulted in a set of realizations for $\hat{d}$ that were very close to those for $\sigma = 1$.

Zero mean, unit variance fractional noise model: both mean and variance esti-

mated

$d = 0.1$: Estimating the variance as well as the mean (the most likely situation on practice) has increased the leftward (negative) bias in $\hat{d}$ to approximately one ASDU for $T = 40$, reducing to 0.5 ADSU for $T = 160$, making the difference in location between the actual distribution of $\hat{d}$ and the Edgeworth expansion of the supposed distribution more marked. The averaged bootstrap distributions tend to be skewed-left, but are more-or-less correctly located with respect to the distribution of $\hat{d}$. The skewness is most marked for for small T, with the sieve bootstrap worst affected. Re-centering the sieve bootstrap distributions shows them to have variance close to that of $\hat{d}$ on average. Re-centering the average model-based bootstrap, on the other hand, produces a curve that closely approximates the Edgeworth expansion, and so tends to similarly under-estimate the variance.

$d = 0.2$: As for $d = 0.1$, except that for $T < 160$ the sieve bootstrap distribution variance is now also lower than it should be. The negative bias in $\hat{d}$ is now around 1 ADSU, but, in contrast to the "estimate the mean only" case, its distribution is not particularly skewed.

$d = 0.4$: The Edgeworth expansions are now extremely inaccurate, as the true distribution has shifted further to the left (negative bias now around 1.5 ADSU), and become more skewed and perhaps more variable as well. The bootstrap distributions have again "followed" the actual distribution, with (as usual) the sieve bootstrap doing slightly better than the model-based in this regard. On re-centering, however, we find that neither bootstrap distribution has sufficient variability (this would cause lower than nominal coverage probabilities). Again, the sieve seems to do better than the model-based. Oddly enough, the "pile-up" on the right-hand boundary now longer occurs.

6.2 Under-specified models

As an extreme, and possibly unrealistic worst case example, consider the estimation of a fractional noise model where the mean is wrongly assumed to be zero. That is, the DGP is fractional noise with with non-zero mean and unit variance (in fact), but estimation procedure assumes the mean is zero.

For schemes B4, M4 we have a DGP with $\mu = 1$, $\sigma = 1$. The bias in $\hat{d}$ is now severe for any value of d: for $d = 0.1$ we find a severe positive bias of around 3–3.5 ASDU's. The sieve bootstrap more-or-less correctly centered, while the model-based bootstrap is badly biased to the right. Both bootstrap distributions are heavily left-skewed, especially for the smaller sample sizes. The Edgeworth distributions are nowhere near the true distribution, being of course still centered on zero. Re-centering the bootstrap distributions causes them to shift towards the Edgeworth; the sieve is now pretty well centered on zero while the model-based is still (relatively) right-shifted. The actual variance of $\hat{d}$ is, however, much lower than for any of the approximating distributions, with the model-based bootstrap coming closest in this regard.

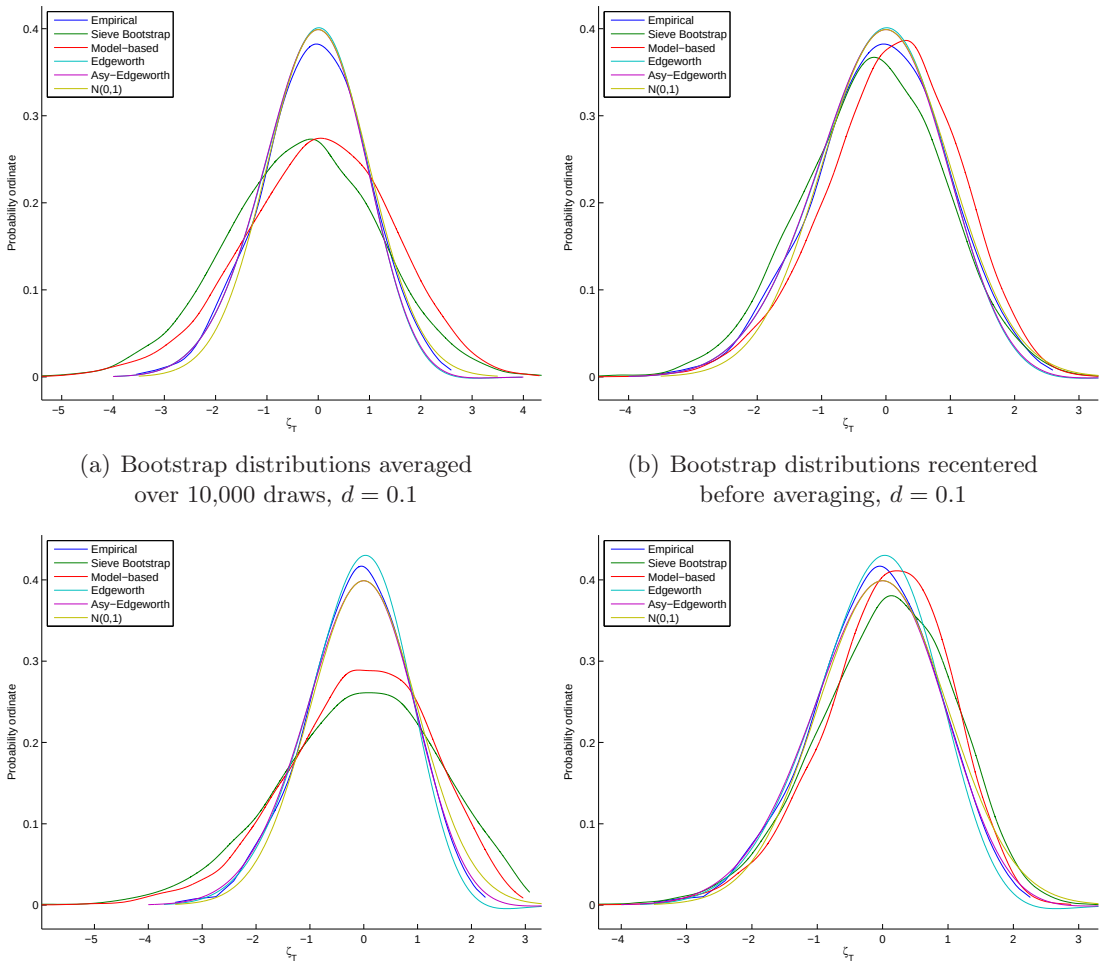
Oddly enough, because of the 'truncation' effect the bias actually decreases as d gets closer to 0.5. For $d = 0.2$ the positive bias in $\hat{d}$ varies between 1.2–2.2 ASDU's depending on

sample size (smaller for smaller T). The sieve bootstrap now also exhibits a small rightward bias as T increases. For $d = 0.3$ the bias in $\hat{d}$ has decreased to about 1.2 ASDU, and the rightward bias in the sieve bootstrap is now comparable to the model-based bootstrap for $T = 160$; for smaller T the greater degree of left-skewness in the sieve distribution counteracts this.

For $d = 0.4$ the bias in $\hat{d}$ is around 0.5 ASDU. The averaged bootstrap distributions are negatively skewed, and showing "pile up" on the $d = 0.5$ boundary. The sieve bootstrap is worst affected, at least for $T = 160$. Re-centering removes this effect, but shows that the variance of the bootstrap distributions is slightly low.

7 Conclusion

Appendix A: Figures



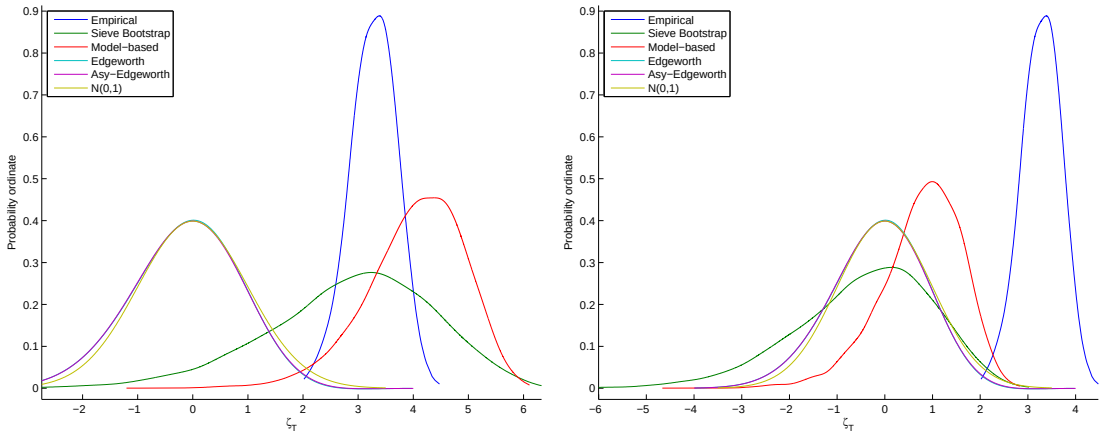
(a) Bootstrap distributions averaged over 10,000 draws, $d = 0.1$

(b) Bootstrap distributions recentered before averaging, $d = 0.1$

(c) Bootstrap distributions averaged over 10,000 draws, $d = 0.3$

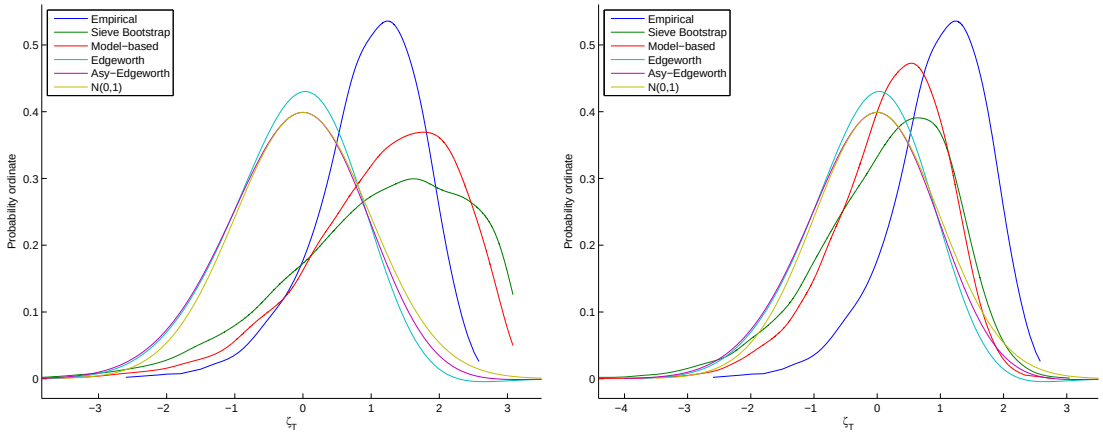
(d) Bootstrap distributions recentered before averaging, $d = 0.3$

Figure 1: Probability densities for $T = 160$: (i) Empirical (ii) Sieve bootstrap (iii) Model-based bootstrap (iv) Edgeworth (v) Asymptotic Edgeworth and (vi) Standard Normal. Model is correctly specified – fractional noise with zero mean and unit variance.



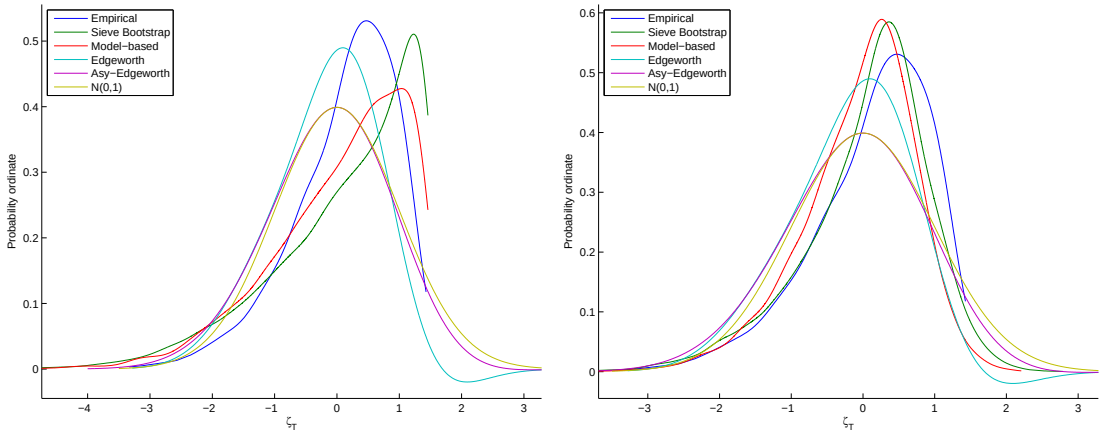
(a) Bootstrap distributions averaged over 10,000 draws, $d = 0.1$

(b) Bootstrap distributions recentered before averaging, $d = 0.1$



(c) Bootstrap distributions averaged over 10,000 draws, $d = 0.3$

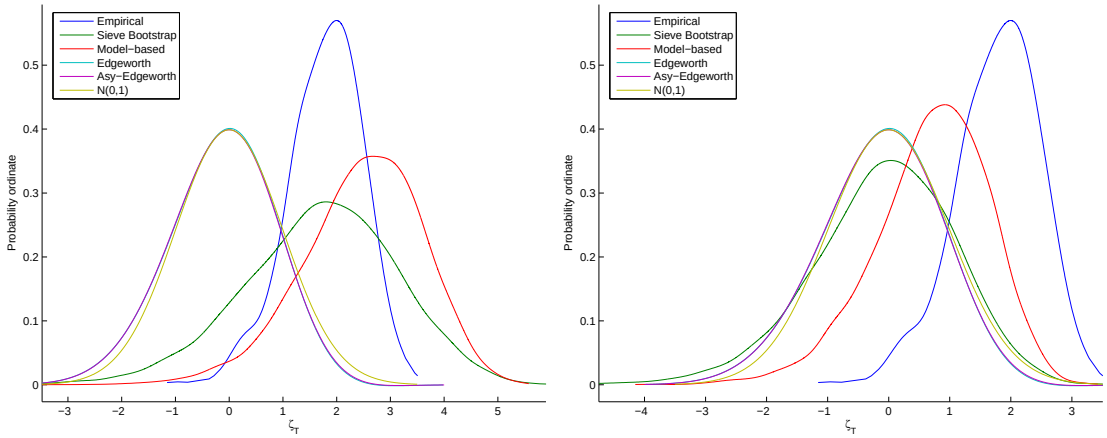
(d) Bootstrap distributions recentered before averaging, $d = 0.3$



(e) Bootstrap distributions averaged over 10,000 draws, $d = 0.4$

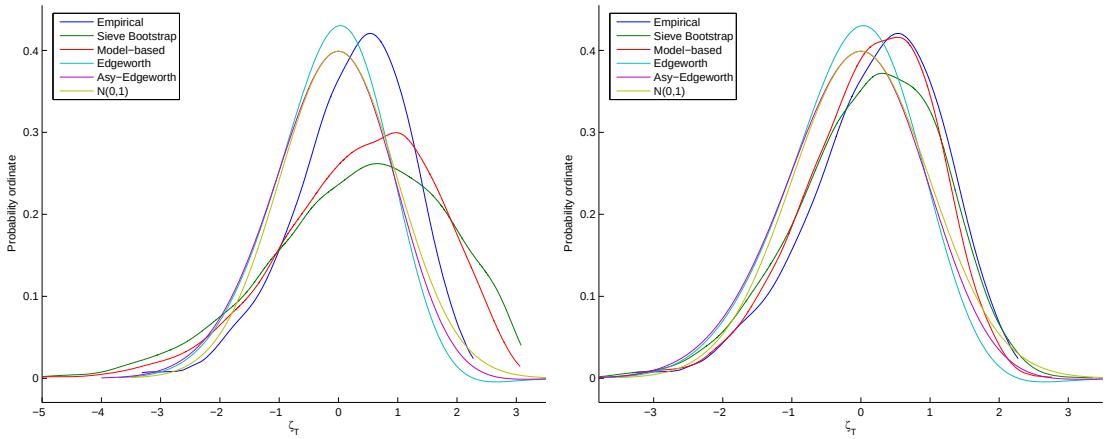
(f) Bootstrap distributions recentered before averaging, $d = 0.4$

Figure 2: Probability densities for $T = 160$: (i) Empirical (ii) Sieve bootstrap (iii) Model-based bootstrap (iv) Edgeworth (v) Asymptotic Edgeworth, and (vi) Standard Normal. Model is misspecified – fractional noise with zero mean and unit variance; true mean is 1.0.



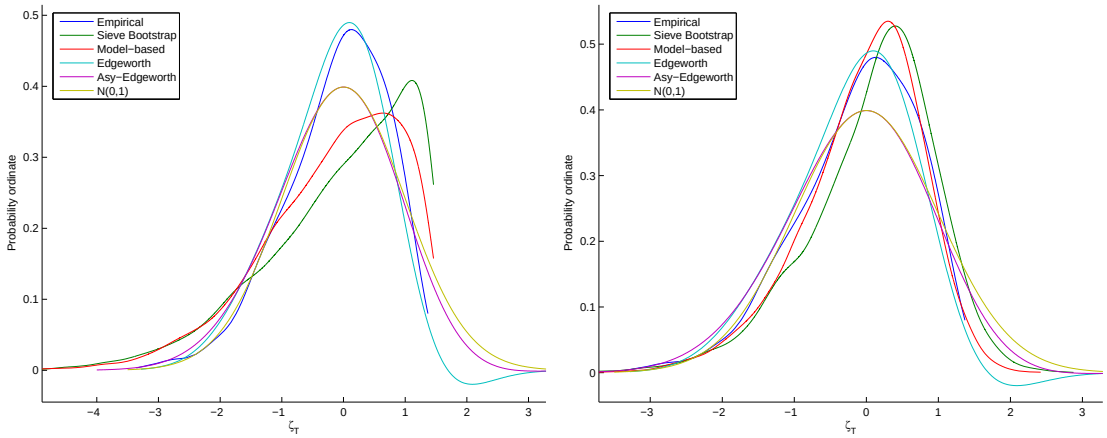
(a) Bootstrap distributions averaged over 10,000 draws, $d = 0.1$

(b) Bootstrap distributions recentered before averaging, $d = 0.1$



(c) Bootstrap distributions averaged over 10,000 draws, $d = 0.3$

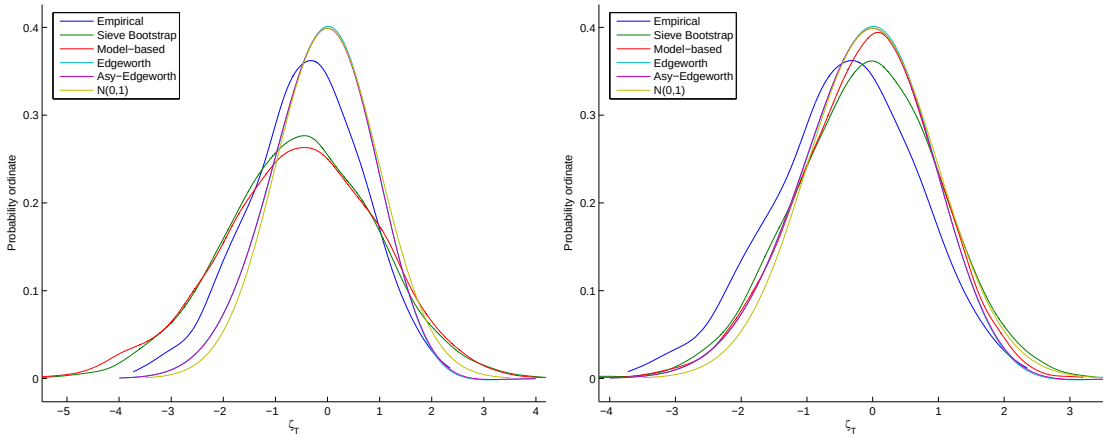
(d) Bootstrap distributions recentered before averaging, $d = 0.3$



(e) Bootstrap distributions averaged over 10,000 draws, $d = 0.4$

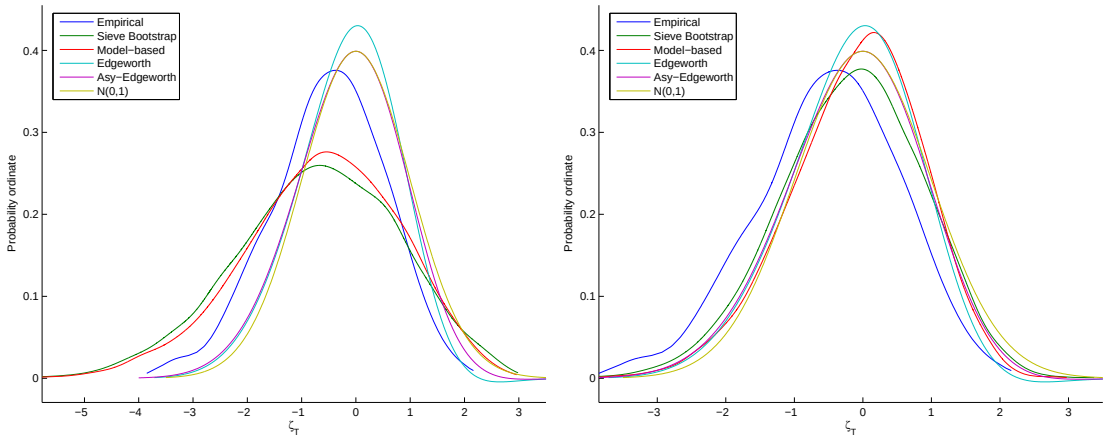
(f) Bootstrap distributions recentered before averaging, $d = 0.4$

Figure 3: Probability densities for $T = 160$: (i) Empirical (ii) Sieve bootstrap (iii) Model-based bootstrap (iv) Edgeworth (v) Asymptotic Edgeworth, and (vi) Standard Normal. Model is misspecified – fractional noise with zero mean and unit variance; true mean is 0.5.



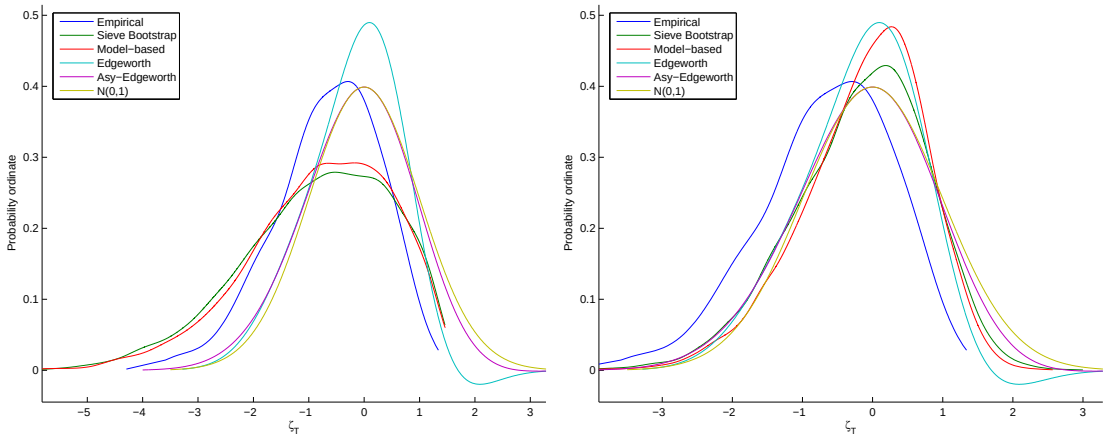
(a) Bootstrap distributions averaged over 10,000 draws, $d = 0.1$

(b) Bootstrap distributions recentered before averaging, $d = 0.1$



(c) Bootstrap distributions averaged over 10,000 draws, $d = 0.3$

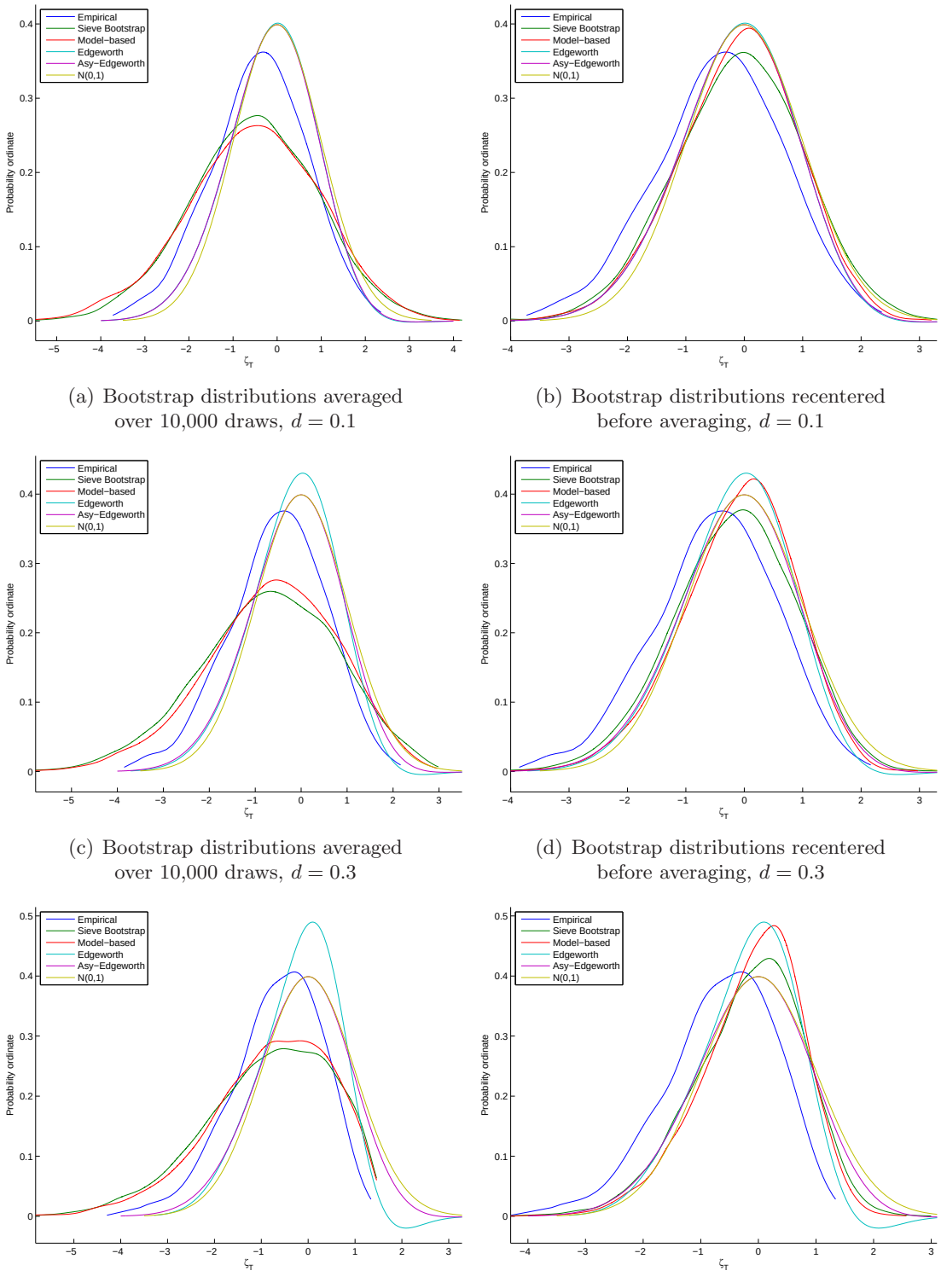
(d) Bootstrap distributions recentered before averaging, $d = 0.3$



(e) Bootstrap distributions averaged over 10,000 draws, $d = 0.4$

(f) Bootstrap distributions recentered before averaging, $d = 0.4$

Figure 4: Probability densities for $T = 160$: (i) Empirical (ii) Sieve bootstrap (iii) Model-based bootstrap (iv) Edgeworth (v) Asymptotic Edgeworth, and (vi) Standard Normal. DGP is fractional noise with zero mean and unit variance; model assumes the mean is unknown.



(a) Bootstrap distributions averaged over 10,000 draws, $d = 0.1$

(b) Bootstrap distributions recentered before averaging, $d = 0.1$

(c) Bootstrap distributions averaged over 10,000 draws, $d = 0.3$

(d) Bootstrap distributions recentered before averaging, $d = 0.3$

(e) Bootstrap distributions averaged over 10,000 draws, $d = 0.4$

(f) Bootstrap distributions recentered before averaging, $d = 0.4$

Figure 5: Probability densities for $T = 160$: (i) Empirical (ii) Sieve bootstrap (iii) Model-based bootstrap (iv) Edgeworth (v) Asymptotic Edgeworth, and (vi) Standard Normal. DGP is fractional noise with zero mean and unit variance; model assumes both mean and variance unknown.

Appendix B: Tables

TABLE 1

Average coverage probabilities: Sieve Bootstrap: $T = 160$

Model (DGP)				Nominal confidence level (%)				
	σ	μ	d	50	80	90	95	99
B1	1	0	0.1	48.4	78.1	88	93.3	98.7
			0.2	47.5	75.2	85.3	92.2	98
			0.3	47.3	76.8	86.8	92.3	98.5
			0.4	48.6	75.2	87.1	93.1	97.5
Average				47.95	76.325	86.8	92.725	98.175
B2	1	0	0.1	48.2	78.1	88.3	93.4	98.9
			0.2	47.6	75.3	85.1	92.6	98.2
			0.3	46.5	77	86.8	92.3	98.5
			0.4	49	75.4	87.4	92.9	97.7
Average				47.825	76.45	86.9	92.8	98.325
B3	1	0	0.1	45.6	74.7	85	91.7	97.9
			0.2	44.1	69.9	80.4	87.4	95.6
			0.3	42.9	70.9	80.6	86.3	95.1
			0.4	41	68.1	78.6	85.3	93.9
Average				43.4	70.9	81.15	87.675	95.625
B4	1	1	0.1	2.2	13.6	26.1	40.8	72.8
			0.2	7.6	28	45.3	61.8	85.6
			0.3	26.2	54.3	70.8	81.5	94.2
			0.4	39.5	70.1	81.6	89.5	97.9
Average				18.875	41.5	55.95	68.4	87.625
B5	2	0	0.1	49	78.3	88.7	93.9	99
			0.2	47.8	75.5	86.5	93.4	98.5
			0.3	46.8	76.7	86.9	93.3	98.7
			0.4	45.5	72.2	84.6	92.4	97.6
Average				47.275	75.675	86.675	93.25	98.45
B6	1	5	0.1	1.7	10.6	21.3	33.7	62.7
			0.2	4.5	15.6	29.8	43.1	73.3
			0.3	8.6	27	42.4	59.5	82.6
			0.4	19.3	44.4	60.7	71.6	90.6
Average				8.525	24.4	38.55	51.975	77.3
B7	1	0.5	0.1	13.3	37.5	57.5	72.4	92
			0.2	30.4	61.2	76	87.1	97.6
			0.3	41.8	69.8	82.4	90.9	98.2
			0.4	44.7	73.4	84.6	92.3	98.3
Average				32.55	60.475	75.125	85.675	96.525
B8	1	0	0.1	45.6	74.7	85	91.7	97.9
			0.2	44.1	69.9	80.4	87.4	95.6
			0.3	42.9	70.9	80.6	86.3	95.1
			0.4	41	68.1	78.6	85.3	93.9
Average				43.4	70.9	81.15	87.675	95.625

Table 2

Average coverage probabilities: Model-based Bootstrap: $T = 160$

Model (DGP)				Nominal confidence level (%)				
	σ	μ	d	50	80	90	95	99
M1	1	0	0.1	48.2	79	90	94.6	99.1
			0.2	48.1	78.9	90	94.6	99.1
			0.3	48.6	80.1	89.9	94.7	98.8
			0.4	53	83	90.6	94.1	98
M1 Total				49.475	80.25	90.125	94.5	98.75
M2	1	0	0.1	47.9	79	89.9	94.6	99.1
			0.2	47.7	79	89.9	94.5	99.1
			0.3	48.4	80.2	89.9	94.8	98.8
			0.4	51.5	83.8	91	94.1	98.3
M2 Total				48.875	80.5	90.175	94.5	98.825
M3	1	0	0.1	44.5	73.5	82.4	88.4	95.7
			0.2	42.9	71.7	81.8	87.1	94.9
			0.3	42.3	71.2	79.8	85.9	94.1
			0.4	39.8	65.8	77	82.1	91.1
M3 Total				42.375	70.55	80.25	85.875	93.95
M4	1	1	0.1	0	0	0	0	0.2
			0.2	0.1	1.1	4.1	11.7	60.4
			0.3	14.6	42	65.9	83.7	98.8
			0.4	44.4	79.9	92.8	95.8	98.6
M4 Total				14.775	30.75	40.7	47.8	64.5
M5	2	0	0.1	47.2	78	89.3	94.7	99.1
			0.2	46.9	77.4	89.1	95.2	99.1
			0.3	45.9	77	88.6	94.7	99.1
			0.4	44.4	73.8	87.3	95.2	99.3
M5 Total				46.1	76.55	88.575	94.95	99.15
M7	1	0.5	0.1	1.1	4.2	11	23.9	66.8
			0.2	17.8	43.3	63.7	79.7	97.2
			0.3	39.7	72.5	85.9	94.2	99.1
			0.4	47.6	82.3	91.5	95.1	97.6
M7 Total				26.55	50.575	63.025	73.225	90.175
M8	1	0	0.1	44.5	73.5	82.4	88.4	95.7
			0.2	42.9	71.7	81.8	87.1	94.9
			0.3	42.3	71.2	79.8	85.9	94.1
			0.4	39.8	65.8	77	82.1	91.1
M8 Total				42.375	70.55	80.25	85.875	93.95

References

Andrews, D. W., Lieberman, O. and Marmer, V. (2006). Higher-order improvements of the parametric bootstrap for long-memory Gaussian processes. *Journal of Econometrics*, **133** 673–702.

Andrews, D. W. K. and Lieberman, O. (2005). Valid edgeworth expansions for the Whittle maximum likelihood estimator for stationary long-memory gaussian time series. *Econometric Theory*, **21** 710–734.

Beran, J. (1994). *Statistics for Long Memory Processes*. Chapman and Hall, New York.

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society*, **B 57** 654–672.

Box, G. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco.

Brockwell, P. L. and Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd ed. Springer Series in Statistics, Springer-Verlag, New York.

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

Burg, J. (1968). A new analysis technique for time series data. Paper presented at the Advanced Study Institute on Signal Processing, N.A.T.O., Enschede, Netherlands.

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.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics*, **17** 1749–1766.

Fox, R. and Taqqu, M. S. (1986). Large sample properties of parameter estimates for strongly dependent stationary gaussian time series. *Annals of Statistics*, **14** 517–532.

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.

Grose, S. D. and Poskitt, D. S. (2006). The finite-sample properties of autoregressive approximations of fractionally-integrated and non-invertible processes. Working paper 15/06, Department of Econometrics and Business Statistics, Melbourne, Australia.

Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer series in Statistics, Springer-Verlag, New York.

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

Hosking, J. R. M. (1980). Fractional differencing. *Biometrika*, **68** 165–176.

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

Lieberman, O. and Phillips, P. C. B. (2004). Expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter. *Econometric Theory*, **20** 464–484.

Lieberman, O., Rousseau, J. and Zucker, D. M. (2000). Small-sample likelihood-based inference in the ARFIMA model. *Econometric Theory*, **16** 231–248.

Lieberman, O., Rousseau, J. and Zucker, D. M. (2003). Valid asymptotic expansions for the maximum likelihood estimator of the parameter of a stationary, gaussian, strongly dependent process. *The Annals of Statistics*, **31** 586–612.

Poskitt, D. S. (2006a). Autoregressive approximation in nonstandard situations: The fractionally integrated and non-invertible cases. *Annals of Institute of Statistical Mathematics* forthcoming.

Poskitt, D. S. (2006b). Properties of the sieve bootstrap for fractionally integrated and non-invertible processes. Working paper 12/06, Melbourne, Australia.

Robinson, P. M. (1995). Log periodogram regression of time series with long memory. *Annals of Statistics*, **23** 1048–1072.

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 estmation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, **53** 165–188.

Wand, M. and Jones, M. (1995). *Kernel Smoothing*. Chapman and Hall.