Supplement of Adv. Stat. Clim. Meteorol. Oceanogr., 3, 33–53, 2017
https://doi.org/10.5194/ascmo-3-33-2017-supplement

Supplement of

Estimating trends in the global mean temperature record

Andrew Poppick et al.

Correspondence to: Andrew Poppick (apoppick@carleton.edu)

Model	Autoregressive parameters				Moving average parameter	Innovation standard		
	ϕ_1	ϕ_2	ϕ_3	ϕ_4	θ_1	deviation (°C)	AICc	BIC
ARMA(4,1)	-0.29 (0.19)	0.36 (0.12)	0.05 (0.09)	0.24 (0.09)	0.80 (0.19)	0.09	-258	-242
AR(1)	0.52 (0.07)	–	–	–	–	0.09	-257	-252

Table S1. Coefficient estimates (with standard errors in parentheses), innovation standard deviations, and AICc and BIC associated with ARMA(4,1) and AR(1) fits to the residuals from model (4). These two models minimize AICc and BIC, respectively. We use the ARMA(4,1) model because this model appears to better represent the low-frequency variation in the residuals (see Figure 3), which is crucial for inferences about trends.

S1 Coefficient estimates for noise model

Table S1 gives information about the ARMA(4,1) and AR(1) models fit to the residuals of model (4) used in Section 4. The ARMA(4,1) model give more conservative inferences about the systematic trend, and is the model we adopt in Section 4 (see Figure 3).

S2 Performance of misspecified methods under AR(2) and ARMA(4,1) simulations

In Section 5.2, we compared the five methods for generating nominal p-values in two settings where the pre-specified AR(1) model was incorrect, where the AR(1) model either over- or under-represented low-frequency variability. Here we present two more comparisons. For these, Table S2 gives the model parameters and Figure S1 shows the two true spectra and the corresponding best AR(1) approximations. The first comparison is to an AR(2) noise model under which the AR(1) approximation over-represents variability at the lowest frequencies but under-represents variability at intermediate frequencies (Figure S1, left). The second comparison is to the ARMA(4,1) model used in our main analysis, under which the AR(1) approximation under-represents variability at the lowest frequencies (Figure S1, right).

First, we show results for the AR(2) true noise model model (Figure S2). In this setting, all methods perform better than in the setting of Figure 9, but the relative performance of the different methods is largely the same as there, except that selection by AICc now performs even more poorly than the block bootstrap at the smallest sample sizes.

Second, we show results for the ARMA(4,1) model used in our main analysis (Figure S3). In this setting, the results are similar in nature to but less extreme than those from the fractionally differenced AR(1) model shown in Figure 10. All methods perform poorly but the parametric methods tend to perform better, especially when the time series length is short.

The behavior from both of these simulations again serves to emphasize that when making inferences about smooth trends, it is most crucial to represent low-frequency variability well.

S3 Model coefficients for simulations in Sections 5 and S2

In Sections 5 and S2, we generate synthetic, mean zero time series with different correlation structures. The models that we simulate from are summarized in Table S2. The general form for an autoregressive fractionally integrated moving average model (ARFIMA) of order (p,d,q) (of which all the simulated models are special cases) is

$$\left(1-\sum_{k=1}^{p}\phi_k B^k\right)(1-B)^d Y_t = (1+\sum_{k=1}^{q}\theta_k B^k)\epsilon(t),$$

where Y_t is the time series at time t, the ϕ's are the AR parameters, the θ's are the MA parameters, d is the (fractional) differencing parameter, B is the backshift operator (i.e., $B^k Y_t = Y_{t-k}$), and $\epsilon(t)$ are uncorrelated innovations with constant

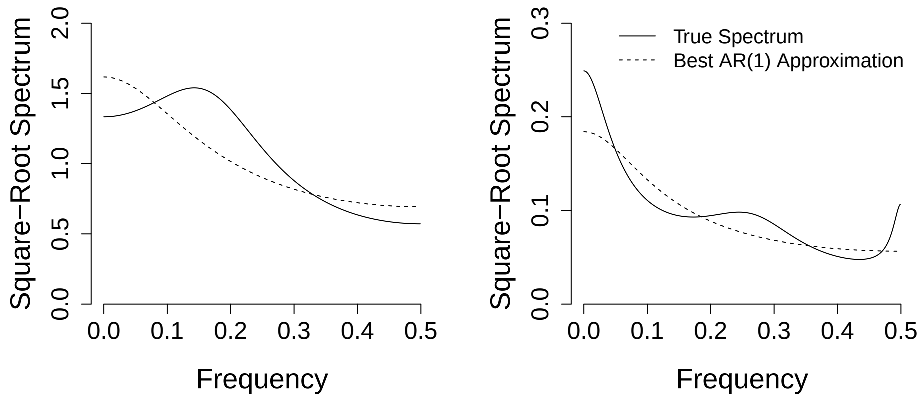

Figure S1. Analogous to Figure 8 but for the AR(2) (left) and ARMA(4,1) (right) models considered in Section S2. See Table S2 for noise model parameters.

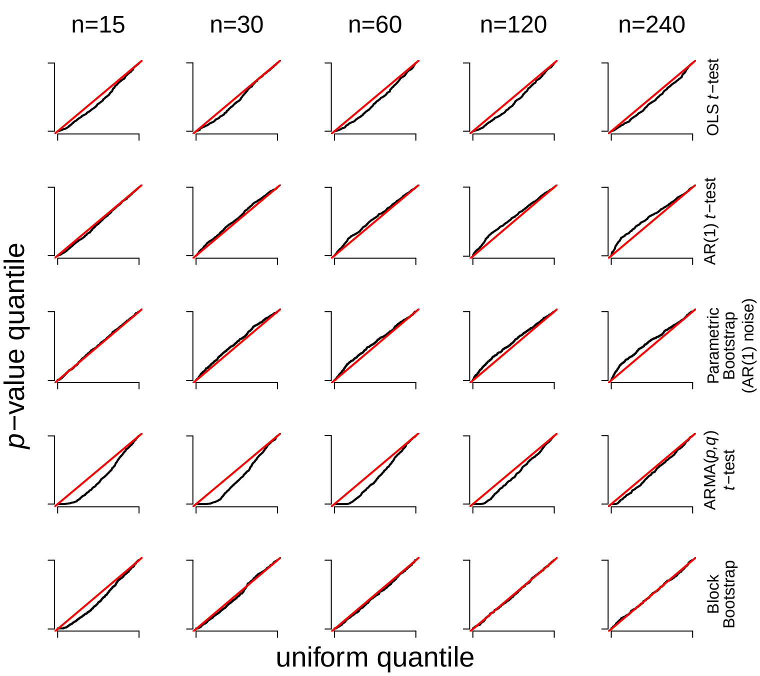

Figure S2. Same as Figure 7 but with simulations from an AR(2) model. The relative results are similar to the setting of Figure 9 but all methods perform better here.

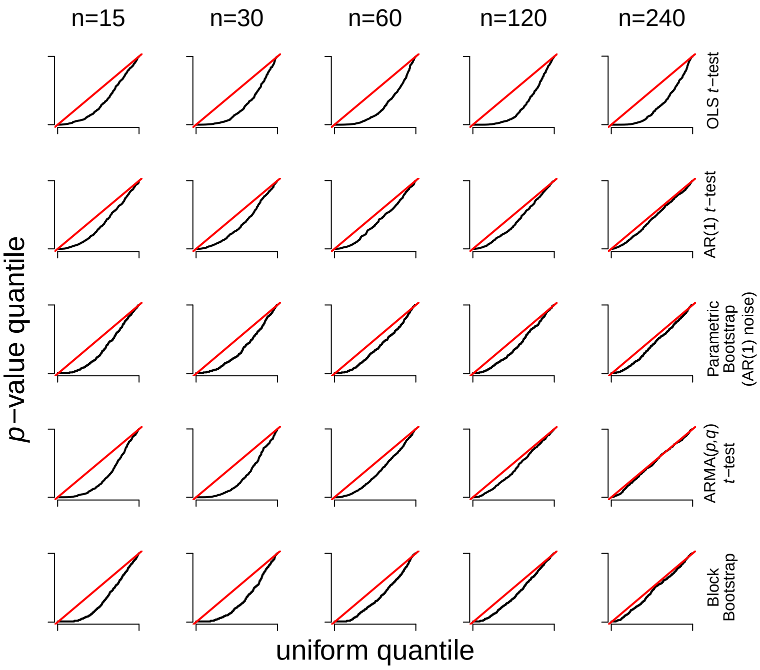

Figure S3. Same as Figure 7 but with simulations from a the ARMA(4,1) model used as the noise model in our main analysis. The relative results are similar to but less extreme than those in the setting of Figure 10.

Figure	Name	AR parameter(s)	MA parameter	Differencing parameter
7	AR(1)	0.5	0	0
9	ARMA(1,1)	0.5	0.25	0
10	fractional AR(1)	0.5	0	0.25
S2	AR(2)	(0.5,-0.25)	0	0
S3	ARMA(4,1)	(-0.29, 0.36, 0.05, 0.24)	0.80	0

Table S2. Parameters in models from which we simulate in Sections 5 and S2.

variance. (The convention is that the acronym is shortened to account for parameters that are set to zero, so for example an ARFIMA(1,0,0) model is called an AR(1) model.)

S4 Performance of maximum likelihood vs. REML

In Section 5, parametric inference was done using maximum likelihood estimators. As shown there, the MLE for covariance parameters can give anticonservative estimates of standard errors for trend parameters in small sample sizes. This problem can be substantially ameliorated using restricted maximum likelihood (REML) instead. Figure S4 repeats the tests in Section 5.1 (where both the true and assumed models are AR(1)) and compares the performance of maximum likelihood and REML. For larger sample sizes, the two procedures are comparable, but in small sample sizes REML is much better calibrated (although still slightly anticonservative in the smallest sample sizes).

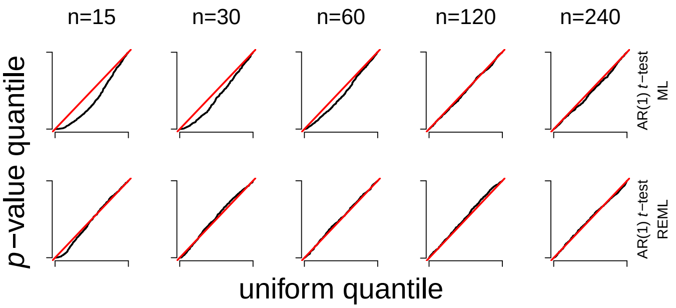

Figure S4. Comparison of maximum likelihood and REML in the same context as Figure 7. The first row is the same as the second row of Figure 7. In the second row here, the estimation is instead done using REML. The REML standard errors give better calibrated inferences in small sample sizes.