arXiv: arXiv:0000.0000

REJOINDER
Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions

Li Pan and Dimitris N. Politis*

Li Pan
Department of Mathematics
University of California–San Diego
La Jolla, CA 92093-0112, USA
e-mail: lipan@ucsd.edu

Dimitris N. Politis
Department of Mathematics
University of California–San Diego
La Jolla, CA 92093-0112, USA
e-mail: dpolitis@ucsd.edu

We would like to sincerely thank all discussants for their insightful comments and encouraging remarks. Notably, in addition to offering helpful suggestions on our paper, all four discussion pieces provide innovative extensions to the methodology of bootstrap prediction intervals.

To elaborate, Profs. Nordman and Lahiri provide a full coverage of the construction of bootstrap prediction intervals in the case of an unstable linear autoregression. In the case of a (stable) linear AR model, Prof. Kim gives details on the implementation of bias-corrected estimators, while Profs. Goncalves and Perron give a complete treatment of prediction intervals in linear ARX models where prediction is also based on additional factors.

Last but certainly not least, Prof. Kreiss inquires about the possibility of obtaining *Model-Free* prediction intervals, and suggests a way to go about it. Coincidentally, Model-Free prediction intervals is the subject of our follow-up paper, Pan and Politis [6], in which we work under the weak assumption that the data series is Markov of order p. In this case, we show how it is indeed possible to construct bootstrap Model-Free prediction intervals using one of three methods: (a) generating bootstap Markov processes based on an estimated transition density as in Rajarshi [10]; (b) using forward or backward local bootstrap as in Paparoditis and Politis [7] [8]; or (c) invoking the *Model-Free prediction principle* of Politis [9] and constructing an appropriate transformation towards "i.i.d.–ness".

Coming back to the current paper, we have formulated the main part of the rejoinder on the basis of three general themes that are brought forth in the discussion pieces.

A. On incorporating model selection variability in the prediction intervals

Profs. Kim and Kreiss independently bring up the question of incorporating model selection variability in the bootstrap prediction intervals. Indeed, this is an important and timely concern; see e.g. the recent paper by Efron [2] on the linear regression analog. Both Profs. Kim and Kreiss suggest using a consistent model selector; in such a case, the asymptotic justification of the bootstrap procedure is automatic since the model identification error disappears in the limit.

*Research partialy supported by NSF grants DMS 13-08319 and DMS 12-23137.

In the case of a linear AR(p) model, it is well known that—under some regularity conditions—the BIC criterion leads to consistent estimation of p; see e.g. Choi [1]. Nevertheless, it is common knowledge that the AIC criterion is much more popular with practitioners in spite—or maybe because—of its tendency to over-estimate model order.

Luckily, there is some recent evidence that linear AR(p) model selection via AIC minimization can be bootstrapped; see Fenga and Politis [3] and the references therein. As an example, consider the AR(2) model

$$X_t = 1.55X_{t-1} - 0.6X_{t-2} + \epsilon_t \text{ with } \epsilon_t \sim \text{ i.i.d. } (0, 1). \tag{R.1}$$

Table 1 shows the relative frequency of orders p chosen by AIC in the real and bootstrap world respectively, p_{AIC} and p^*_{AIC}; in both cases, the maximum order under consideration was 5. The entries of Table 1 show that the statistical behavior of p_{AIC} can be mimicked reasonably well by that of p^*_{AIC}. Table 1 shows results obtained via resampling fitted residuals with innovations ϵ_t that were Gaussian or Laplace; resampling predictive residuals gave qualitatively similar results.

order chosen	1	2	3	4	5
$n = 50$, Normal					
p_{AIC}	0.060	0.762	0.146	0.020	0.012
p^*_{AIC}	0.103	0.639	0.174	0.052	0.032
$n = 50$, Laplace					
p_{AIC}	0.072	0.756	0.130	0.026	0.016
p^*_{AIC}	0.123	0.627	0.167	0.052	0.031
$n = 100$, Normal					
p_{AIC}	0.004	0.746	0.174	0.044	0.032
p^*_{AIC}	0.007	0.636	0.228	0.071	0.058
$n = 100$, Laplace					
p_{AIC}	0.000	0.742	0.178	0.042	0.038
p^*_{AIC}	0.004	0.630	0.232	0.073	0.061

Table 1: Statistics for p chosen by AIC in the real and bootstrap world, p_{AIC} and p^*_{AIC}, respectively; AR(2) model (R.1) with Normal or Laplace innovations.

To see how an estimated model order plays out in prediction intervals we conducted a small simulation based on the above AR(2) model (R.1); the simulation follows the pattern outlined in the main paper. Table 2 shows the performance of prediction intervals when the order p is treated as known as is the *modus operandi* in the main paper. By contrast, Table 3 shows what would happen if a practitioner uses AIC to pick the order in the real world but uses the same order in the bootstrap world; this quite realistic setting results in slight under-coverages in all four methods considered here. The under-coverages are corrected if the practitioner also uses AIC minimization to pick the order in the bootstrap world as shown in Table 4. Note that Tables 2—4 were constructed using Gaussian innovations; using Laplace innovations gave qualitatively similar results.

normal errors	nominal coverage 95%				nominal coverage 90%		
$n = 50$	CVR	LEN	st.err		CVR	LEN	st.err
Ff	0.931	3.933	0.521		0.883	3.328	0.417
Fp	0.946	4.171	0.543		0.902	3.527	0.434
FSf	0.946	4.185	0.547		0.901	3.521	0.441
FSp	0.945	4.159	0.537		0.899	3.496	0.431
$n = 100$	CVR	LEN	st.err		CVR	LEN	st.err
Ff	0.939	3.908	0.358		0.889	3.295	0.280
Fp	0.945	4.026	0.366		0.899	3.399	0.283
FSf	0.946	3.020	0.362		0.898	3.384	0.286
FSp	0.945	4.014	0.360		0.897	3.381	0.278

Table 2: Simulation Results for AR(2) model (R.1); the order p is treated as known.

normal errors	nominal coverage 95%				nominal coverage 90%		
$n = 50$	CVR	LEN	st.err		CVR	LEN	st.err
Ff–AIC	0.929	3.952	0.565		0.877	3.336	0.465
Fp–AIC	0.944	4.216	0.601		0.899	3.566	0.485
FSf–AIC	0.944	4.216	0.605		0.896	3.532	0.486
FSp–AIC	0.942	4.199	0.614		0.895	3.530	0.480
$n = 100$	CVR	LEN	st.err		CVR	LEN	st.err
Ff–AIC	0.937	3.899	0.370		0.886	3.295	0.299
Fp–AIC	0.945	4.046	0.385		0.898	3.415	0.297
FSf–AIC	0.945	4.023	0.377		0.896	3.390	0.303
FSp–AIC	0.945	4.032	0.376		0.896	3.388	0.289

Table 3: Simulation Results for AR(2) model (R.1); the order p is estimated by AIC in the real world, and the same value is used in the bootstrap world.

normal errors	nominal coverage 95%				nominal coverage 90%		
$n = 50$	CVR	LEN	st.err		CVR	LEN	st.err
Ff–AIC	0.933	4.042	0.589		0.882	3.406	0.481
Fp–AIC	0.947	4.308	0.624		0.904	3.633	0.501
FSf–AIC	0.948	4.320	0.633		0.902	3.613	0.508
FSp–AIC	0.947	4.297	0.640		0.900	3.602	0.503
$n = 100$	CVR	LEN	st.err		CVR	LEN	st.err
Ff–AIC	0.938	3.917	0.366		0.887	3.315	0.296
Fp–AIC	0.946	4.069	0.382		0.899	3.433	0.295
FSf–AIC	0.946	4.058	0.376		0.898	3.420	0.301
FSp–AIC	0.946	4.067	0.378		0.898	3.419	0.291

Table 4: Simulation Results for AR(2) model (R.1); the order p is estimated by AIC in the real world, and re-estimated by AIC minimization in each of the bootstrap samples.

B. On h-step ahead prediction intervals

Profs. Goncalves and Perron rightfully bring attention to some issues related to the construction of h-step ahead prediction intervals with $h > 1$. To discuss this further, consider the simple homoscedastic case with no additional regressors, e.g. the causal model (1.1) of the main paper which is re-iterated below:

$$X_t = m(X_{t-1}, ..., X_{t-p}) + \epsilon_t. \tag{R.2}$$

In Section 4 of the main paper the following recommendation was given: *if h-step ahead prediction*

intervals are desired, then work directly with the model

$$X_t = \mu(X_{t-h}, ..., X_{t-h-p+1}) + \zeta_t \tag{R.3}$$

instead of model (R.2); the switch to notation $\mu(\cdot)$ and ζ_t in the above serves to emphasize that these are different quantities than the $m(\cdot)$ and ϵ_t appearing in (R.2).

Note that if (R.2) holds true with $\epsilon_t \sim$ i.i.d. $(0, \sigma^2)$, then the ζ_t will be h–dependent and not i.i.d. By contrast, it may be true—although quite rare in practice—that (R.3) holds true with $\zeta_t \sim$ i.i.d. in which case the ϵ_t appearing in (R.2) will not be i.i.d. Hence, Profs. Goncalves and Perron correctly point out that although eq. (R.3) is tailor-made for prediction, it may not be such a good vehicle for generating the whole bootstrap series.

Let us now assume the more standard assumption (R.2) with $\epsilon_t \sim$ i.i.d. $(0, \sigma^2)$. In this setting, if one were to use the 'direct' approach offered by (R.3), the dependence of the ζ_t's must be taken into account. As suggested by Profs. Goncalves and Perron, employing a time series bootstrap on the residuals, i.e., the estimated ζ_t's, gives a viable way out. Indeed, a resampling scheme that can incorporate the h–dependence of the ζ_t's may be particularly appealing. The dependent wild bootstrap is such a possibility, and so is the Linear Process Bootstrap of McMurry and Politis [4].

An easier approach might be to estimate *both* functions, i.e., $m(\cdot)$ in (R.2) and $\mu(\cdot)$ in (R.3). Then, use the fitted model (R.2) for the purpose of resampling in order to generate $X_1^*, \ldots, X_n^*, \ldots, X_{n+h}^*$, and save the fitted model (R.3) for the purpose of constructing the optimal predictor. Note, however, that this presupposes identifying the form of the optimal h–step ahead predictor function $\mu(\cdot)$ having assumed model (R.2) with $\epsilon_t \sim$ i.i.d. $(0, \sigma^2)$.

Consider the following three cases; only the last one seems to be not straightforward.

1. If $m(\cdot)$ is linear, then $\mu(\cdot)$ is immediately obtained by iterating the one-step-ahead predictor $m(\cdot)$.
2. If $m(\cdot)$ is nonparametrically specified, then $\mu(\cdot)$ can also be estimated nonparametrically from the scatterplot of X_t vs. $(X_{t-h}, ..., X_{t-h-p+1})$ that is associated with model (R.3).
3. If $m(\cdot)$ is nonlinear but it is parametrically specified, then the form of the h–step ahead conditional expectation $\mu(\cdot)$ is not obvious except in special cases. However, a hybrid approach is always possible, e.g., to estimate $\mu(\cdot)$ nonparametrically but perform a model-based resampling using the (parametric) nonlinear model (R.2).

C. Conditional vs. unconditional validity

Profs. Nordman and Lahiri inquire on the importance of the conditioning step of the Forward Bootstrap algorithms that ensures that the last p values of the bootstrap series are identical to those in the original series; recall that Step B in Section 2.1 of the main paper is meant to render conditional validity to the bootstrap prediction intervals. Obviously, conditional validity implies unconditional validity but the converse is not necessarily true.

For conciseness, the simulations in the main paper focused on unconditional coverages. We now revert to a simple AR(1) model in order to investigate the conditional coverage of some intervals of interest, and complement the simulations offered by Profs. Nordman and Lahiri. In order to investigate conditional coverages, we now fix the last value of each of the 'true' datasets to some chosen value; the 'true' datasets in the simulation are then generated using backward bootstrap with X_n fixed to a particular value, say x_n.

Table 5 compares Masarotto's (1990) method—denoted by M—with our four forward methods. As discussed in Remark 3.2 of the main paper, the only difference between Masarotto's method to our FSf method is that Masarotto does not fix the last p values of the bootstrap series to match the ones from the original series, i.e., Step B of Section 2.1 is omitted. In spite of this, Masarotto's

intervals appear to have quite accurate conditional coverages as the entries of Table 5 suggest; similar findings using Laplace errors can be found in Pan and Politis (2014a).

To further shed light on this phenomenon which was astutely pointed out by Profs. Nordman and Lahiri, recall that the distribution of the bootstrap predictive root depends on $X_n = x_n$ because

$$X^*_{n+1} - \hat{X}^*_{n+1} = (\hat{\phi} - \hat{\phi}^*)x_n + \epsilon^*_{n+1}. \tag{R.4}$$

Since $\hat{\phi} - \hat{\phi}^* = O_p(1/\sqrt{n})$, it is apparent that the term $(\hat{\phi} - \hat{\phi}^*)x_n$ is small compared to the error term ϵ^*_{n+1}; this is why using the wrong x_n—as Masarotto's method does—can still yield accurate coverages. The situation is similar for studentized bootstrap roots since the first term of the numerator contains a term including x_n. Nevertheless, there seems no reason to forego using the correct x_n in the bootstrap predictive root (R.4).

normal errors	nominal coverage 95%				nominal coverage 90%		
$X_n = 3$	CVR	LEN	st.err		CVR	LEN	st.err
M	0.947	4.108	0.368		0.899	3.450	0.290
FSf	0.951	4.216	0.355		0.905	3.540	0.272
FSp	0.951	4.222	0.337		0.905	3.535	0.264
Ff	0.946	4.125	0.342		0.898	3.466	0.269
Fp	0.951	4.217	0.341		0.904	3.537	0.265
$X_n = 2$	CVR	LEN	st.err		CVR	LEN	st.err
M	0.945	4.002	0.362		0.899	3.384	0.283
FSf	0.947	4.045	0.357		0.902	3.417	0.274
FSp	0.947	4.049	0.349		0.902	3.413	0.263
Ff	0.943	3.959	0.350		0.895	3.350	0.270
Fp	0.947	4.047	0.358		0.902	3.415	0.270
$X_n = 1$	CVR	LEN	st.err		CVR	LEN	st.err
M	0.944	3.960	0.364		0.897	3.340	0.282
FSf	0.944	3.957	0.369		0.897	3.336	0.279
FSp	0.945	3.968	0.366		0.897	3.335	0.269
Ff	0.939	3.877	0.370		0.891	3.273	0.275
Fp	0.944	3.966	0.380		0.898	3.340	0.269
$X_n = 0$	CVR	LEN	st.err		CVR	LEN	st.err
M	0.945	3.956	0.366		0.897	3.329	0.283
FSf	0.944	3.937	0.371		0.895	3.313	0.281
FSp	0.944	3.949	0.374		0.895	3.312	0.272
Ff	0.939	3.861	0.379		0.889	3.252	0.281
Fp	0.943	3.944	0.389		0.896	3.318	0.273

Table 5: Simulation Results of AR(1) model $X_t = 0.5X_{t-1} + \epsilon_t$ with normal innovations when $n = 100$.

To elaborate, Masarotto replaces the term $(\hat{\phi} - \hat{\phi}^*)x_n$ in (R.4) with $(\hat{\phi} - \hat{\phi}^*)X^*_n$ where X^*_n is random (with mean zero). If x_n is near zero and X^*_n happens to be near its mean, then the terms match well. However, there is an issue of unnecessary variability here that is manifested with slightly higher standard errors of the lengths of the Masarotto intervals and with inflated CVRs—but the CVR inflation is due to a fluke, not a *bona fide* capturing of the predictor variability. Now if x_n is large (in absolute value), there is an issue of bias in the centering of the Masarotto intervals but this is again masked by the unnecessary/excess variability of the term $(\hat{\phi} - \hat{\phi}^*)X^*_n$.

All in all, adjusting the last p values of the bootstrap series to match the original ones is highly advisable in a causal, linear AR(p) model. Furthermore, it may achieve particular importance under a nonlinear and/or nonparametric model in which the above arguments break down. Adjusting the last p values certainly becomes crucial in autoregressions with heteroscedastic errors as in eq. (1.2)

of the main paper where the scale of the error also depends on these last p values. For example, without this adjustment prediction intervals for the next observation coming from an ARCH model would have the same width regardless of the recent past; this amounts to ignoring the ARCH effect on the volatility of the process which is obviously suboptimal .

In closing, we would like to re-iterate our thanks to the six discussants for their stimulating and thought-provoking contributions.

References

[1] Choi, B.-S. (1992). *ARMA Model Identification*, Springer, New York.

[2] Efron, B. (2014). Model selection, estimation, and bootstrap smoothing (with discussion), *J. Amer. Statist. Assoc.,* 2014.

[3] Fenga, L. and Politis, D.N. (2011). Bootstrap based ARMA order selection, *J. Statist. Comput. Simul.*, vol 81, no.7, pp. 799-814.

[4] McMurry, T. and Politis, D.N. (2010). Banded and tapered estimates of autocovariance matrices and the linear process bootstrap, *J. Time Ser. Anal.*, vol. 31, pp. 471-482. [Corrigendum, *J. Time Ser. Anal.*, vol. 33, 2012.]

[5] Pan, L. and Politis, D.N. (2014a). Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions, Working Paper, Department of Economics, UCSD; retrievable from: http://www.escholarship.org/uc/item/67h5s74t.

[6] Pan, L. and Politis, D.N. (2014b). Bootstrap prediction intervals for Markov processes, Working Paper, Department of Economics, UCSD.

[7] Paparoditis, E. and Politis, D.N. (1998). The backward local bootstrap for conditional predictive inference in nonlinear time series, in *Proceedings of the 4th Hellenic-European Conference on Computer Mathematics and its Applications (HERCMA'98), E. A. Lipitakis (Ed.)*, Lea Publishing, Athens, Greece, pp. 467-470.

[8] Paparoditis, E. and Politis, D.N. (2002). The local bootstrap for Markov processes, *Journal of Statistical Planning and Inference*, vol. 108, no. 1, pp. 301–328.

[9] Politis, D.N. (2013). Model-free Model-fitting and Predictive Distributions (with discussion), *Test*, vol. 22, no. 2, pp. 183-250.

[10] Rajarshi, M.B. (1990). Bootstrap in Markov sequences based on estimates of transition density, *Annals of the Institute of Statistical Mathematics*, vol. 42, no. 2, pp. 253-268.