A Novel Parametric Predictive Bootstrap Method

Abdulrahman M. A. Aldawsari[a,*], Frank P.A. Coolen[b], Tahani Coolen-Maturi[b]

[a]*Department of Mathematics, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia*
[b]*Department of Mathematical Sciences, Durham University, Durham, UK*

Abstract

Bootstrap methods are widely used statistical techniques known for their simplicity and good properties. This paper introduces a novel bootstrap method called the parametric predictive bootstrap (PP-B), which relies on parametric models and is designed for predictive inference. The PP-B method is evaluated in various scenarios typically used with other bootstrap methods to assess its performance in estimation and prediction inference. Comparisons of PP-B with other bootstrap methods are made in terms of the coverage probabilities of confidence and prediction intervals. Simulation results indicate that PP-B excels in predictive inference due to its explicitly predictive nature.

Keywords: Bootstrap, Confidence intervals, Prediction intervals, Prediction regions, Nonparametric predictive inference bootstrap

1. Introduction

Measuring the uncertainty of a sample estimate is an important aspect of statistical inference. Bootstrap methods are sampling techniques used to quantify the uncertainty of sample estimates [10]. They have been applied to a wide range of statistical problems due to their simplicity of implementation and the ability to provide good approximate results for sample estimates. A

*Corresponding author

Email addresses: abd.aldawsari@psau.edu.sa (Abdulrahman M. A. Aldawsari), frank.coolen@durham.ac.uk (Frank P.A. Coolen), tahani.maturi@durham.ac.uk (Tahani Coolen-Maturi)

researcher may use the bootstrap method to avoid performing complicated mathematical derivations or, in some instances, to offer a solution where no analytical answer is possible [25]. The bootstrap method has contributed to resolving problems such as the estimation of the standard error for statistical estimators. The standard error can be used to evaluate the accuracy of an estimator, but for the majority of statistical estimators, there are no mathematical formulas to estimate the standard error. The bootstrap exploits the power of computers to assess the statistical accuracy of complicated procedures. Additionally, the bootstrap method is capable of efficiently determining the confidence interval for a parameter of interest. The use of the bootstrap method has been extended to many problems, including hypothesis testing, because of its simplicity of implementation and good performance.

The first presentation of the bootstrap method was in a Stanford University technical report by Bradley Efron in 1977, followed by his famous paper in the Annals of Statistics in 1979 [10, 20]. Many efforts have been made to popularise the bootstrap method in the statistical community, such as Diaconis and Efron [19], Efron [21], and Efron and Gong [23]. There are many modifications to Efron's bootstrap, such as double bootstrap, smooth bootstrap, and Bayesian bootstrap that have been presented in the literature; see e.g. [6, 18, 39]. Bootstrap methods have been introduced for different types of data, e.g. real data [27], right-censored data [1], and ordinal data [8]. Chernick [9] described bootstrap methods along with examples and applications such as hypothesis testing, confidence intervals, regression, and time series. As a result, the importance of the bootstrap approach has been widely recognised.

This paper presents a new bootstrap method, the parametric predictive bootstrap, which we denote as PP-B. It relies entirely on parametric models and aims to predict future observations. The proposed bootstrap methods will be evaluated in a range of scenarios that have been used with other bootstrap methods. This will enable us to investigate the performance of PP-B in estimation and prediction inference. In Section 2, we introduce several bootstrap methods from the literature for later comparison with our PP-B method. In Section 3, the concept of the parametric predictive bootstrap is introduced, clarifying how it differs from other bootstrap methods described in Section 2. In Section 4, the performance of PP-B for estimation is compared with different bootstrap methods using percentile confidence intervals. In Section 5, we consider the percentile prediction interval to predict the future sample statistic in order to investigate the performance of PP-B in prediction inference. Section 6 provides further investigation to evaluate the performance of PP-B as a prediction approach using prediction regions for the bootstrap prediction interval. In the last section, we present some concluding remarks of this paper.

2. Bootstrap methods

In this section, we describe three different bootstrap methods: Efron's bootstrap (EB), parametric bootstrap (PB), and nonparametric predictive inference bootstrap (NPI-B). These bootstrap methods will be compared with the parametric predictive bootstrap (PP-B) introduced in this paper. The classical Efron bootstrap is a nonparametric sampling technique that does not make any assumptions about how observations are distributed. In contrast, the parametric bootstrap requires assumptions regarding the distribution of the data. The nonparametric predictive inference bootstrap is formulated for predictive inference and does not use an assumed parametric model. The PP-B is similar to NPI-B in terms of focusing on prediction, but it requires assumptions about data distribution.

2.1. Efron's bootstrap

The bootstrap method has become an essential technique for researchers because of its good properties and general applicability to a variety of statistical situations. The standard version of the bootstrap method is introduced by Efron [25], which is a resampling technique from the original data set. This bootstrap method employs the empirical distribution to quantify the uncertainty of sample estimates. The basic idea of Efron's bootstrap (EB) is to resample with replacement from the original observations repeatedly, where each observation has an equal probability of being selected during the resampling process [31]. It has been widely used in applied statistics as it relies on few mathematical assumptions and can be easily implemented using statistical software. It is important to note that EB makes no assumptions regarding the distribution of observations [26, 36].

Suppose that there is a random sample $x_1, x_2, \ldots, x_n$ from an unknown distribution F, and we want to estimate the parameter of interest $\theta(F)$, e.g. the mean or variance by the statistic T. The bootstrap method can be used to construct the sampling distribution of any statistic. A bootstrap sample is denoted by $X^* = (x_1^*, x_2^*, \ldots, x_n^*)$, which consists of members of the original data set $X = (x_1, x_2, \ldots, x_n)$. It is obtained by randomly sampling n times with replacement from the original sample. The size of a bootstrap sample can be chosen differently from the original sample size. The basic bootstrap method generates an empirical estimate of the sampling distribution of the statistic (bootstrap distribution). The procedure involves drawing a large number of samples from the observations and determining the statistics for each sample. The statistic's sampling distribution can be estimated by the relative frequency distribution of these statistics. The bootstrap distribution typically mirrors the shape of the actual sampling distribution resulting from the sampling process.

A point worth noting here is that some of the observations will be repeated once or more in a bootstrap sample, which makes them different from the original sample. Also, specific observations may not appear at all in a particular bootstrap sample. Consequently, there will be a variation of the values for the parameter of interest. We should draw large numbers of bootstrap samples to approximate the variation of a sampling distribution. The EB method is described in many references with examples and applications, e.g. Berrar [7], Davison and Hinkley [18], and Efron [22]. The idea of bootstrap has been applied to a variety of statistical inferences. For example, Rosenkranz [37] estimated the bias of treatment effect estimators using the bootstrap method.

2.2. Parametric bootstrap

The parametric bootstrap (PB) method assumes that the data come from a known distribution with unknown parameters. In this method, samples are drawn from the assumed distribution with the estimated parameters instead of resampling with replacement from the original data. The idea of the PB method is to estimate the parameters of the assumed distribution using available data and to generate a number of PB samples from the assumed distribution with the estimated parameters [26, 33]. The PB method requires knowledge of the data distribution and can contain observations that were not included in the original sample, but this method may produce misleading results if the assumed model is wrong. Conversely, the EB method does not assume a distribution for the data; all observations are included in the original sample, and tied observations occur. The PB method can be used in situations where some knowledge about the form of the underlying population is available.

2.3. Nonparametric predictive inference bootstrap

The nonparametric predictive inference (NPI) method has been developed over the past two decades for a wide range of applications and problems in statistics, along with various data types. NPI is a statistical technique based on Hill's assumption $A_{(n)}$ that makes inferences on a future observation based on past data observations [11, 12]. Hill [28, 29, 30] introduced the assumption $A_{(n)}$ for the prediction of one future observation X_{n+1} with no prior knowledge about the underlying distribution. Suppose that $x_1, \ldots, x_n$ are the observed data corresponding to real-valued and exchangeable random quantities $X_1, \ldots, X_n$. Let $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ be the ordered observations and define $x_{(0)} = -\infty$ and $x_{(n+1)} = +\infty$ for ease of notation. For one future observation X_{n+1}, the assumption $A_{(n)}$ is:

$$P(X_{n+1} \in I_i) = \frac{1}{n+1} \tag{1}$$

where $I_i = (x_{(i-1)}, x_{(i)})$ and $i = 1, \ldots, n+1$. The assumption $A_{(n)}$ states that the future observation X_{n+1} is equally likely to fall within any open interval $(x_{(i-1)}, x_{(i)})$. These intervals were created by the previous n observations between consecutive order statistics of the given sample. The assumption $A_{(n)}$ itself is not sufficient to derive precise probabilities for any event of interest, but it can be used to derive bounds (lower and upper) of probabilities, which are called imprecise probabilities. The NPI approach is introduced by Coolen and Augustin [4, 5], which uses lower and upper probabilities for events of interest considering future observations based on Hill's assumption. The lower probability is the maximum lower bound for the precise probability for the event and is denoted by $\underline{P}(\cdot)$. The upper probability is the minimum upper bound for the event and is denoted by $\overline{P}(\cdot)$. The NPI lower and upper probabilities become precise probability if they are equal $\underline{P}(\cdot) = \overline{P}(\cdot)$, $0 \leq \underline{P}(\cdot) \leq \overline{P}(\cdot) \leq 1$. The NPI lower and upper probabilities for the event $X_{n+1} \in B$, where $B \subset \mathbb{R}$ are:

$$\underline{P}(X_{n+1} \in B) = \frac{1}{n+1}|\{i : I_i \subseteq B\}| \tag{2}$$

$$\overline{P}(X_{n+1} \in B) = \frac{1}{n+1}|\{i : I_i \cap B \neq \emptyset\}| \tag{3}$$

The lower probability (2) is the total probability mass assigned to intervals I_i that are completely contained within B, and the upper probability (3) is taking into account all probability masses assigned to intervals that can be in B.

Sequential application of the assumptions $A_{(n)}, \ldots, A_{(n+m-1)}$ can be used to generalise NPI for $(m \geq 1)$ future real-valued observations based on n real data observations. These assumptions imply that all $\binom{n+m}{n}$ possible different orderings of the m future observations among the n data observations are equally likely to appear, with no further assumptions made on where future observations will be within any of these intervals I_i [15]. The NPI approach is considered for statistical inference, e.g. acceptance sampling [13], precedence testing for two groups [17], and the accuracy of diagnostic tests [16].

Coolen and Binhimd [14] introduced a predictive bootstrap method based on NPI, called nonparametric predictive inference bootstrap (NPI-B). The NPI-B method involves creating $n+1$ intervals between the n ordered observations of the original data, then selecting one of these intervals randomly. The first observation is drawn uniformly from the selected interval and then this observation is added to the original data, resulting in $n+1$ observations. This leads to the creation of a partition consisting of $n+2$ intervals, from which the second observation is sampled. The process continues until m observations are drawn, where m is predefined. These m observations constitute one NPI-B sample (which, of course, does not include the n original observations). In NPI-B, all possible orderings of the new observations among the past observations are equally likely to occur. NPI-B's sampling method, which involves drawing each observation from the intervals in the partition created by combining the n original observations with all previously drawn observations belonging to the same bootstrap sample, leads to greater variation in bootstrap samples than Efron's and parametric bootstrap samples. It is worth mentioning that one observation is sampled uniformly from each chosen interval when applying NPI-B. However, it cannot be sampled uniformly from an open-ended interval; e.g., data defined on the whole real line lead to the first and last intervals in the form of $(-\infty, x_{(1)})$ and $(x_{(n)}, +\infty)$. Coolen and Binhimd [14] suggest using the tail of a Normal distribution for real-valued data, and the tail of an Exponential distribution for non-negative real-valued data. It is important to note that the conditional tail distribution is only used to sample an observation from open-ended intervals; otherwise, the observation is sampled uniformly from finite intervals. The NPI-B algorithm for real-valued data on finite and infinite intervals is as follows:

1. Create $n+1$ intervals between the n ordered observations $(x_{(0)}, x_{(1)}), (x_{(1)}, x_{(2)}), \ldots, (x_{(n-1)}, x_{(n)}), (x_{(n)}, x_{(n+1)})$, where $x_{(0)}$ and $x_{(n+1)}$ are the end points of the possible data range.
2. Select one of the $n+1$ intervals randomly, each with equal probability, and sample one future observation uniformly from this selected interval.
   (a) We sample the future value uniformly for any finite interval.
   (b) For the case with data on the whole real line $(-\infty, +\infty)$: If the chosen interval is $(-\infty, x_{(1)})$ or $(x_{(n)}, +\infty)$, we sample the future value from the tail of Normal distribution with mean $\mu = \frac{x_{(1)}+x_{(n)}}{2}$ and standard deviation $\sigma = \frac{x_{(n)}-\mu}{\Phi^{-1}(\frac{n}{n+1})}$, where Φ^{-1} indicates the inverse function of a standard normal cumulative distribution function.
   (c) For the case with data on the $(0, +\infty)$: If the chosen interval is $(x_{(n)}, +\infty)$, we sample the future value from the tail of Exponential distribution with rate $\lambda = \frac{\ln(n+1)}{x_{(n)}}$.
3. Add this sampled observation x_1^* to the data; increase n to $n+1$.
4. Repeat Steps 1-3, now with $n+1$ data, to obtain a further future value. This is continued to sample m future observations from the intervals in the partition created by combining the n original observations with all previously drawn observations that belong to the bootstrap sample. These m drawn observations $(x_1^*, x_2^*, \ldots, x_m^*)$ form one NPI-B sample of size m.
5. Repeat Steps 2-4 to obtain B of NPI-B samples of size m.

3. The general idea of parametric predictive bootstrap

In this section, we present the main idea of PP-B for real-valued data, followed by a brief comparison with other bootstrap methods described in Section 2. In the PP-B method, a single observation is sampled from an assumed distribution with estimated parameters based on an original data set of size n. Then, this observation is added to the data and the process is repeated, now with $n+1$ observations. We re-estimate the distribution parameters with the new observation added to the data in order to sample the second observation. This process continues to sample m further values in the same way, each observation adding to the data and re-estimating the parameters before sampling the next one. The PP-B sample consists of these

m sampled observations, excluding the n original data observations. The PP-B algorithm for one-dimensional real-valued data is as follows:

1. We have a random sample consisting of n observations $x_1, x_2, \ldots, x_n$ from a known distribution $F(x;\theta)$, with parameter θ.
2. The parameter θ of the assumed distribution is estimated by $\hat{\theta}$ from the available data, using maximum likelihood estimation (MLE) or any other estimation method.
3. Sample one future observation x_1^* randomly from the fitted distribution $F(x;\hat{\theta})$.
4. Add x_1^* to the data results in the data set $(x_1, x_2, \ldots, x_n, x_1^*)$; increase n to $n+1$.
5. Repeat Steps 2-4, now with $n+1$ data, to obtain a further future value. This is continued to sample m observations in total, with each one added to the data and the parameter re-estimated before sampling the next observation. These sampled observations $x_1^*, x_2^*, \ldots, x_m^*$ are a PP-B sample of size m.
6. Repeat Steps 2-5 to obtain B of PP-B samples of size m.

As a consequence of the method of sampling observations in PP-B, with sampled observation added to the data set and the parameter estimated before sampling the next one, the bootstrap samples show more variation than the EB and PB samples. The method for sampling observations in NPI-B, with each observation drawn from the intervals created by combining the n original observations with all previously drawn observations belonging to the same bootstrap sample, also causes more variation in the bootstrap samples than the EB and PB samples. In the EB and PB methods, all observations are sampled based on the original data only. EB depends on resampling with replacement from the original data, where each value of the original data set has the same probability of being chosen by random selection during the resampling process [24]. The PB method assumes the data to come from a known distribution with unknown parameters. The parameters of the assumed distribution are estimated from the available data, then observations are sampled from the assumed distribution with the estimated parameters in order to obtain PB sample [26]. The bootstrap samples in PP-B, NPI-B, and PB are not restricted to already observed values, whereas all observations in EB samples are in the original sample.

We give a brief comparison of variations in bootstrap samples for each bootstrap method and leave a more detailed comparison for the following sections. We compute the variance for a statistic of interest T using the bootstrap technique to measure the spread of these statistic values based on the bootstrap samples. The bootstrap estimate of variance $\hat{\sigma}^2_{boot}$ can be computed by generating B bootstrap samples, then calculate the statistic of interest T for each bootstrap sample to obtain $T^*_1, \ldots, T^*_B$. The value of $\hat{\sigma}^2_{boot}$ is given by:

$$\hat{\sigma}^2_{boot} = \frac{\sum_{j=1}^{B}[T^*_j - (\sum_{j=1}^{B} T^*_j/B)]^2}{B-1} \tag{4}$$

To conduct the simulation study, we begin by generating one original sample of size n from each of the two distributions: N(0,1) and Exp(0.5). Then, we apply different bootstrap methods $B = 1000$ times for each generated sample. The mean and variance of each bootstrap sample are computed, and then we estimate the variance based on different bootstrap methods. This procedure is repeated for various original sample sizes $n = 5, 25, 100, 200, 500$ drawn from each of the two distributions: N(0,1) and Exp(0.5). For consistency, the same original data sets for each distribution and sample size are used across all bootstrap methods. It is important to note that, for each method, the bootstrap samples are generated with the same size as the corresponding original sample.

Table 1 shows the estimate of variance using different bootstrap methods for the mean and variance. The results were approximated to four decimal digits, but we used additional digits with some values to make the results more informative and to avoid the inclusion of zeros "0.0000". PP-B and NPI-B have the largest estimated variance values for the mean and variance among these bootstrap methods, as expected due to the method of sampling observations in both methods. The results for N(0,1) show that the NPI-B method has the largest variance in all cases except for the mean when $n = 500$, in which the PP-B method has a larger variance. Also, the NPI-B method provides the largest variance in most cases of the mean and all cases of the variance for Exp(0.5), followed by the PP-B method. The NPI-B method has a larger variance in most cases compared to the PP-B method due to the assumption of a parametric model in the PP-B method.

We observe that the PP-B and NPI-B methods give the largest variance for the statistics compared to the other methods. However, this is actually a good point because their variances are the closest to the variance of the underlying distribution. For example, with the normal distribution N(0,1), which has a variance of 1, the variance of the statistics using PP-B and NPI-

(a) N(0,1)

method	statistics	$n = 5$	$n = 25$	$n = 100$	$n = 200$	$n = 500$
PP-B	mean	0.3320	0.0606	0.0146	0.0083	0.0041
	variance	0.5491	0.1201	0.0249	0.0151	0.0084
NPI-B	mean	0.6067	0.0892	0.0165	0.0085	0.0040
	variance	2.9020	0.3624	0.0444	0.0172	0.0112
PB	mean	0.1981	0.0316	0.0075	0.0042	0.0020
	variance	0.0009	0.0001	0.00002	0.00001	0.000004
EB	mean	0.1515	0.0359	0.0074	0.0043	0.0021
	variance	0.1743	0.0735	0.0133	0.0071	0.0047

(b) Exp(0.5)

method	statistics	$n = 5$	$n = 25$	$n = 100$	$n = 200$	$n = 500$
PP-B	mean	0.4517	0.1764	0.0789	0.0386	0.0149
	variance	4.3163	2.0995	1.9951	0.8993	0.3314
NPI-B	mean	0.6725	0.2146	0.0839	0.0321	0.0140
	variance	11.7162	9.6144	4.5507	1.1390	0.5424
PB	mean	0.2385	0.0887	0.0396	0.0194	0.0075
	variance	2.6386	1.5059	1.3511	0.5997	0.2195
EB	mean	0.1211	0.0717	0.0352	0.0143	0.0070
	variance	0.1063	0.3657	0.6540	0.2132	0.1408

Table 1: *The bootstrap estimate of variance for the mean and variance when the original sample was from N(0,1) and Exp(0.5).*

B is the largest, but they are also the closest to 1, except that the statistic variance obtained using the NPI-B method is overestimated when $n = 5$. Similarly, for the exponential distribution with $\lambda = 0.5$, which has a variance of 4, the variance of the statistics derived from PP-B and NPI-B are again the largest and the closest to 4, but for the statistic variance the NPI-B method is overestimated when $n = 5, 25, 100$.

4. Confidence intervals

In this section, we consider the comparison of PP-B among bootstrap methods using confidence intervals to investigate its performance in estimation inference. First, we give a general review of confidence intervals and describe how a bootstrap technique can be used to construct confidence intervals. A $100(1-2\alpha)\%$ confidence interval for the parameter θ is an interval constructed from a random sample, such that if we were to repeat the experiment a large number of times, the interval would contain the true value of θ in $100(1-2\alpha)\%$ of the cases. It is important to note that the interval will depend on the value of the estimate $\hat{\theta}$ and the sampling distribution of the estimator $\hat{F}$. The sample size, confidence level, and the variability in the sample are all factors that influence the width of the interval. The larger samples produce narrower confidence intervals when all other factors are equal, while a higher confidence level or greater variability in the sample produces wider confidence intervals when all other factors are equal. In one-sample case, we have a random sample observations $x_1, x_2, ..., x_n$ from distribution F. Assume the parameter of interest θ, e.g. the mean or variance, which can be estimated by the statistic T. We need to determine the sampling distribution of the estimator $\hat{\theta}$. The bootstrap method can be used to estimate the sampling distribution of the statistic.

Efron and Tibshirani [25] introduced different ways to construct confidence intervals based on bootstrap technique. In this paper, we will use the percentile confidence interval which depend on the percentiles of the bootstrap distribution of a statistic. The $100(1-2\alpha)\%$ percentile interval is giving by:

$$\begin{aligned} \hat{\theta}_B^{*(\alpha)} &< \theta < \hat{\theta}_B^{*(1-\alpha)} \\ T_B^{*(\alpha)} &< \theta < T_B^{*(1-\alpha)} \end{aligned} \tag{5}$$

where T_j^*, $j = 1, \ldots, B$ is the computed statistic of interest T for each bootstrap sample. So, $T_j^{*(\alpha)}$ is the $100 \cdot \alpha$th percentile of the T_j^* values, that means the $B \cdot \alpha$th of the ordered list of the B replications of T^* and it is likewise for $T_j^{*(1-\alpha)}$ indicate the $100 \cdot (1-\alpha)$th percentile of the T_j^* values. For example, if $B = 1000$ and $\alpha = 0.05$, then lower endpoint of the percentile interval $T_j^{*(\alpha)}$ is the 50th ordered value of replications and upper endpoint of the percentile interval $T_j^{*(1-\alpha)}$ is the 950th ordered value of replications. If $B \cdot \alpha$ is not an integer, we assuming $\alpha \leq 0.5$ and let $k = \lfloor (B+1)\alpha \rfloor$ is the largest integer $\leq (B+1)\alpha$, then we define α and $1-\alpha$ by the kth largest and $(B+1-k)$th largest value of T_j^*, respectively.

The simulation study is conducted to find the coverage proportion and average width of confidence intervals for the mean and variance. In this study, we use Beta(8,2) with a different original sample size $n = 50, 100, 200, 400$ and confidence level 95% and 90%. We generate an original sample of size n from Beta(8,2) and then apply different bootstrap methods $B = 1000$ times.

(a) mean

Bootstrap	measures	Confidence level							
		95%				90%			
		$n = 50$	$n = 100$	$n = 200$	$n = 400$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
PP-B	CP	0.9850	0.9910	0.9930	0.9910	0.9670	0.9750	0.9820	0.9700
	AL	0.0933	0.0665	0.0471	0.0333	0.0780	0.0558	0.0396	0.0280
NPI-B	CP	0.9970	0.9970	0.9950	0.9930	0.9860	0.9850	0.9840	0.9690
	AL	0.1089	0.0724	0.0492	0.0341	0.0901	0.0605	0.0412	0.0286
PB	CP	0.9430	0.9410	0.9540	0.9440	0.8800	0.8980	0.9080	0.9050
	AL	0.0664	0.0472	0.0334	0.0236	0.0558	0.0397	0.0281	0.0198
EB	CP	0.9350	0.9430	0.9580	0.9460	0.8800	0.9030	0.9090	0.9040
	AL	0.0656	0.0469	0.0332	0.0235	0.0552	0.0395	0.0279	0.0198

(b) variance

Bootstrap	measures	Confidence level							
		95%				90%			
		$n = 50$	$n = 100$	$n = 200$	$n = 400$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
PP-B	CP	0.9820	0.9900	0.9930	0.9940	0.9590	0.9770	0.9800	0.9740
	AL	0.0175	0.0126	0.0090	0.0063	0.0144	0.0105	0.0075	0.0053
NPI-B	CP	0.9980	0.9960	0.9960	0.9950	0.9920	0.9860	0.9870	0.9870
	AL	0.0390	0.0227	0.0134	0.0082	0.0308	0.0181	0.0108	0.0067
PB	CP	0.9450	0.9410	0.9520	0.9520	0.9000	0.9020	0.9020	0.9000
	AL	0.0128	0.0091	0.0064	0.0045	0.0108	0.0076	0.0054	0.0038
EB	CP	0.8820	0.9320	0.9380	0.9450	0.8310	0.8750	0.8880	0.8870
	AL	0.0117	0.0087	0.0063	0.0045	0.0099	0.0074	0.0053	0.0037

Table 2: *Coverage of* $(1-2\alpha)\%$ *confidence interval using percentile method for mean and variance parameters when the original sample from* $Beta(8,2)$.

It is important to note that the bootstrap samples for each method are the same size as the original samples. The statistics are computed for each bootstrap sample to construct percentile intervals using Equation (5). Then, we discover which percentile confidence intervals include the true statistics of the Beta(8,2) distribution. This procedure is repeated $N = 1000$ times in order to find the coverage proportions of different bootstrap methods. The performance assessment of each bootstrap method is based on two criteria: coverage proportion and the average width of the intervals. It is desirable to have a proportion of coverage that is close to these advertised confidence levels with a smaller average width of intervals.

Table 2 presents the coverage proportions and average interval widths for the mean and variance based on the four bootstrap procedures. The notation CP and AW refer to the coverage proportion and average interval widths, respectively. The NPI-B method produces the largest average width of confidence intervals in all cases for these two statistics, followed by the PP-B method. As a result, over-coverage occurs in all cases of the NPI-B and PP-B methods. The sampling methods in PP-B and NPI-B, which add a sampled observation to the data set before sampling the next one, leads to more variation in the bootstrap samples, as discussed in Section 3. The greater variability in the sample produces wider intervals, so as a result PP-B and NPI-B lead to wider confidence intervals than other bootstrap methods. The NPI-B method produces a wider average width of intervals than the PP-B method when the sample size and confidence level of both methods are equal. We conclude that the NPI-B method has more variation than the PP-B method, as we had expected, due to the assumption of a parametric model in the PP-B method. The NPI-B method does not use an assumed parametric model, leading to greater variability compared to the PP-B method.

The method that has a coverage proportions closer to nominal coverage probability is the preferred one. PB and EB have coverage that is closer to the presumed coverage probabilities with narrower intervals on average than NPI-B and PP-B. For the variance, the PB method achieved the best coverage proportion of all cases and it has the nominal coverage probability 0.90% when $n = 50, 400$. The EB method shows almost 7% under-coverage result below their 95% and 90% nominal confidence level when $n = 50$, but it is improved when the sample size gets large. The PP-B and NPI-B methods have an over-coverage tendency due to wider intervals arising from greater variability. The PP-B method does not perform well in confidence intervals, as it is not developed for estimating population characteristics, but for predictive inference. It is explicitly aimed at predictive inference, with variability in different bootstrap samples reflecting uncertainty in prediction in line with the NPI-B method.

5. Prediction intervals

We begin this section by providing an overview of prediction intervals and explaining how to construct prediction intervals using the bootstrap technique. Lu and Chang [32] used the bootstrap method to construct a prediction interval for one or more future values from a Birnbaum-Saunders distribution. They constructed the prediction interval using the bootstrap percentile method with bootstrap calibration. The Birnbaum-Saunders distribution is used in reliability applications to model failure times. They assumed that a random sample $x_1, \ldots, x_n$, is taken from Birnbaum-Saunders distribution function F with parameters α and β. The density of F is defined by

$$f(x;\alpha,\beta) = \frac{1}{2\alpha\beta\sqrt{2\pi}}\left[\left(\frac{x}{\beta}\right)^{-\frac{1}{2}} + \left(\frac{x}{\beta}\right)^{-\frac{3}{2}}\right] exp\left[-\frac{1}{2\alpha^2}\left(\frac{x}{\beta} - 2 + \frac{\beta}{x}\right)\right],$$

where $x > 0$ and the two parameters of Birnbaum–Saunders distribution $\alpha, \beta > 0$.

A bootstrap sample of size n, $x_1^*, \ldots, x_n^*$ is n random values drawn with replacement from $x_1, \ldots, x_n$, each with a probability of $1/n$, to construct the estimated distribution F^*. In this case, the bootstrap sample is considered as a sample of the unknown distribution. Then, generate $y_1^*, \ldots, y_m^*$ from the estimated distribution F^*, where m is the number of future observations. Thereafter, the mean of $y_1^*, \ldots, y_m^*$ is obtained and denoted by $\bar{y}_m^*$. Repeat the previous procedure B times to obtain B values of $\bar{y}_m^*$, denoted by $\bar{y}_m^*(1), \ldots, \bar{y}_m^*(B)$. Then, construct $100(1-2\alpha)\%$ prediction interval for the mean of future observations $\bar{x}_m$ as:

$$\left(\bar{y}_{m,B}^{(\alpha)}, \bar{y}_{m,B}^{(1-\alpha)}\right) \tag{6}$$

where the lower endpoint $\bar{y}_{m,B}^{(\alpha)}$ is the $B \times \alpha$th value in the ordered list of the B replications of $\bar{y}_m^*$ and the upper endpoint $\bar{y}_{m,B}^{(1-\alpha)}$ is the $B \times (1-\alpha)$th value in this ordered list. If $B \times \alpha$ is not an integer, the same procedure of bootstrap-t interval is used as discussed in previous section (use the largest integer).

Lu and Chang [32] investigate the performance of the bootstrap prediction intervals for a single future observation and for the mean of five future observations through simulations. They draw a sample of size $n+m$ from the Birnbaum-Saunders distribution $x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m}$, where $x_1, \ldots, x_n$ represents the past sample and $x_{n+1}, \ldots, x_{n+m}$ represents the future sample. Then, find the observed mean of m future observations $x_{n+1}, \ldots, x_{n+m}$, $\bar{x}_m$. The prediction interval for the mean of future observations is constructed by drawing the bootstrap sample $x_1^*, \ldots, x_n^*$, then generating from them $y_1^*, \ldots, y_m^*$ and finding $\bar{y}_m^*$. Repeat this $B = 1000$ times to have the list of B values $\bar{y}_m^*(1), \ldots, \bar{y}_m^*(B)$ in order to construct the prediction interval for $\bar{x}_m$ as described earlier. In the case of prediction for a single future observation, we draw the bootstrap sample $x_1^*, \ldots, x_n^*$, then generate y_1^* from them and repeat this $B = 1000$ times to have the list of B values $y_1^*(1), \ldots, y_1^*(B)$ and construct the prediction interval for x_{n+1}. The 90% and 95% prediction intervals for a single future observation x_{n+1} and the mean of m future observations $\bar{x}_m$ are computed in the Lu and Chang study [32]. They conducted a Monte-Carlo simulation to determine the coverage probability by counting how many intervals contain x_{n+1} and $\bar{x}_m$. A percentile prediction interval used by Lu and Chang [32] is defined as the LC method.

Mojirsheibani and Tibshirani [35] introduced different ways to construct prediction intervals such as bootstrap-t, percentile and BCa prediction intervals. The percentile prediction interval is used in this paper because the BCa prediction interval cannot be constructed for a single future observation and the bootstrap-t prediction interval is not transformation respecting. They assumed that a random sample $X = (x_1, \ldots, x_n)$ represent past sample and $Y = (y_1, \ldots, y_m)$ represent a future sample, where X and Y are independent and identically distributed with a common distribution F and $\hat{\theta} = T$ is the estimator of scalar parameter θ. Let F_n and F_m are the CDF's of $\hat{\theta}_n = T_n$ and $\hat{\theta}_m = T_m$, which are the estimators of a scalar parameter θ from the past sample and future sample respectively. Let $\hat{F}_n$ and $\hat{F}_m$ are the CDF's of $\hat{\theta}_n^* = T_n^*$ and $\hat{\theta}_m^* = T_m^*$, the bootstrap version of $\hat{\theta}_n = T_n$ and $\hat{\theta}_m = T_m$. The bootstrap samples X^* and Y^* are drawn with replacement from the past sample X. The $100(1 - 2\alpha)\%$ percentile prediction interval for $\hat{\theta}_m = T_m$ is:

$$(\hat{\theta}_{lo}, \hat{\theta}_{up}) = \left(\hat{F}_m^{-1}\left[\Phi(z^{(\alpha)}(1 + m/n)^{1/2})\right], \hat{F}_m^{-1}\left[\Phi(z^{(1-\alpha)}(1 + m/n)^{1/2})\right]\right) \quad (7)$$

where $\hat{F}_m$ is the bootstrap distribution of $\hat{\theta}_m^* = T_m^*$, and $z^{(\alpha)} = \Phi^{-1}(\alpha)$.

Mojirsheibani [34] studies the effects of bootstrap iteration (calibration) as a method to improve the coverage accuracy. They generated X^* and Y^* from the past sample X and then resample Y^{**} from X^*. All previous studies of constructing prediction intervals were based on Efron's bootstrap method. We refer to the percentile prediction interval recommended by Mojirsheibani and Tibshirani [35] as the MT method. In this paper, we focus on the percentile prediction interval using MT and LC methods, but without iterated bootstrap.

A comparison of PP-B with other bootstrap methods is carried out using prediction intervals in order to investigate their performance in prediction inference. The percentile prediction intervals are constructed based on the LC and MT methods but without bootstrap iteration. Here we will draw the past and future samples separately as done by Mojirsheibani and Tibshirani [35], where both samples are independent and identically distributed. The percentile prediction interval based on the bootstrap method can be generlized to a large class of statistics and is not restricted to sample means. We extend studies to investigate the performance of the percentile prediction with LC and MT methods for future sample variance. The following processes are used to find the coverage proportion and average interval widths for the percentile prediction interval of a statistic:

1. Draw an original sample of size n from specific distribution $X = (x_1, \ldots, x_n)$ to be the past sample and then draw another original sample of size m from the same distribution $Y = (y_1, \ldots, y_m)$ to be the future sample, where the two samples are independent and identically distributed.
2. Compute the statistic of the future sample T_m using $Y = (y_1, \ldots, y_m)$.
3. Draw B bootstrap samples of size m from $x_1, \ldots, x_n$ and find the statistic for each bootstrap sample T^*_{mj}, where $j = 1, \ldots, B$.
4. Construct an $100(1-2\alpha)\%$ prediction interval of T_m by LC method and MT method:
   (a) Lu and Chang (LC) method: lower bound is the $\alpha \cdot B$th value in the ordered list of T^*_{mj} and the upper bound is the $(1-\alpha) \cdot B$th value in this list (use the largest integer if these values are not integer).
   (b) Mojirsheibani and Tibshirani (MT) method:
   Lower bound: $F_m^{-1}\left[\Phi(z_\alpha(1+\frac{m}{n})^{1/2})\right] = F_m^{-1}[\alpha_1]$ is the $\alpha_1 \cdot B$th value in the ordered list of T^*_{mj}.
   Upper bound: $F_m^{-1}\left[\Phi(z_{1-\alpha}(1+\frac{m}{n})^{1/2})\right] = F_m^{-1}[\alpha_2]$ is the $\alpha_2 \cdot B$th value in the ordered list of T^*_{mj}.
   If $\alpha_1 \cdot B$ or $\alpha_2 \cdot B$ are not integer, use the largest integer.
5. Determine if this interval contains the statistic of the future sample T_m in Step 2 and compute the width of the prediction interval for both methods.
6. Steps 1-5 are repeated N times to find the coverage proportion (number of times out of N that interval captures its corresponding future sample mean) and the average interval widths.

Bootstrap	measures	Confidence level							
		95%				90%			
		$n = 50$	$n = 100$	$n = 200$	$n = 400$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
PP-B	CP_{LC}	0.9510	0.9390	0.9600	0.9480	0.8950	0.8910	0.9080	0.9040
	CP_{MT}	0.9900	0.9890	0.9940	0.9940	0.9710	0.9740	0.9830	0.9750
	AW_{LC}	0.1493	0.1063	0.0755	0.0534	0.1252	0.0892	0.0634	0.0449
	AW_{MT}	0.2091	0.1484	0.1054	0.0742	0.1770	0.1256	0.0894	0.0632
NPI-B	CP_{LC}	0.9600	0.9470	0.9660	0.9450	0.9220	0.8990	0.9180	0.9020
	CP_{MT}	0.9930	0.9920	0.9940	0.9950	0.9800	0.9810	0.9840	0.9800
	AW_{LC}	0.1574	0.1090	0.0763	0.0537	0.1317	0.0915	0.0641	0.0451
	AW_{MT}	0.2211	0.1524	0.1071	0.0750	0.1868	0.1292	0.0905	0.0635
PB	CP_{LC}	0.8390	0.8220	0.8360	0.8300	0.7760	0.7440	0.7450	0.7510
	CP_{MT}	0.9460	0.9380	0.9540	0.9420	0.9030	0.8870	0.9120	0.9050
	AW_{LC}	0.1066	0.0753	0.0535	0.0378	0.0896	0.0634	0.0450	0.0318
	AW_{MT}	0.1478	0.1048	0.0744	0.0525	0.1258	0.0890	0.0632	0.0447
EB	CP_{LC}	0.8380	0.8150	0.8380	0.8380	0.7630	0.7410	0.7420	0.7450
	CP_{MT}	0.9490	0.9350	0.9510	0.9420	0.8950	0.8850	0.9110	0.9030
	AW_{LC}	0.1054	0.0749	0.0533	0.0377	0.0885	0.0630	0.0449	0.0317
	AW_{MT}	0.1461	0.1041	0.0742	0.0525	0.1245	0.0884	0.0631	0.0446

Table 3: *Coverage of $100(1-2\alpha)\%$ prediction interval for the mean of m future observations from Beta(3,1), when $m = n$.*

We conduct a simulation study as shown in the steps above to investigate the coverage performance and average width of intervals for each bootstrap method. The number of simulations is set equal to $N = 1000$ and the bootstrap methods are applied to each past sample $B = 1000$ times. The percentile prediction intervals are constructed for the mean of $m = n$ future observations with various original sample sizes $n = 50, 100, 200, 400$ from Beta(3,1) and Gamma(6,3) at confidence levels 90% and 95%. Table 3 and Table 4 show the coverage proportions and average width of intervals for LC and MT methods using different bootstrap methods. The notation CP_{LC} and AW_{LC} refer to the coverage proportion and average interval widths for the LC prediction interval, respectively. In the MT prediction interval, CP_{MT} and AW_{MT} represent coverage proportion and average interval widths, respectively.

First, we compare the performance of different bootstrap methods with the LC prediction interval. In all future sample sizes and confidence levels, the PP-B and NPI-B methods provide coverage that is close to the coverage probability. Conversely, the coverage of PB and EB is considerably below the nominal coverage probability for all cases irrespective of sample size and confidence level. In comparison to their nominal coverage probabilities, the observed coverage is at least 12% lower than the nominal coverage probability of 0.95, and at least 14.5% below the nominal coverage probability of 0.90 for the Gamma(6,3) distribution. Similarly, for the Beta(3,1) distribution, the coverage proportions are at least 11% lower than the corresponding nominal coverage probabilities. Bootstrap methods with a predictive nature, such as PB-B and NPI-B, perform well and provide good coverage for LC prediction intervals. PP-B has the advantage of achieving good coverage with a narrower interval, where the average interval widths for PP-B are smaller than those for NPI-B in all cases.

We also compare different bootstrap methods in terms of their performance based on MT prediction intervals. The PP-B and NPI-B methods have over-coverage in all cases as a result of the large average interval width in both methods. The PB and EB methods provide coverage that is closer to the presumed coverage probabilities. Although the MT method improves coverage for PB and EB, it is still possible to obtain coverage closer to the coverage probability using PP-B and NPI-B with the LC method. For example, the coverage of PP-B and NPI-B with the LC method is closer to the nominal coverage probabilities when $m = 100$ compared to PB and EB with the MT method for the Beta(3,1) distribution. Similarly, for the Gamma(6,3) distribution, PP-B and NPI-B with the LC method achieve coverage that is closer to the nominal probabilities at $m = 400$ than PB and EB with the MT method.

Mojsheibani [34] investigated prediction intervals using a future sample of size m with a different size from the past sample n. The simulation studies are conducted using various bootstrap methods, incorporating different statistics and sample sizes m, to evaluate their performance in terms of coverage proportion and the average width of intervals. In our study, we construct percentile prediction intervals for the variance of $m = n/2$ future observations. In this study, we use different original samples sizes $n = 50, 100, 200, 400$ with confidence level 95% and 90% from Beta(3,1) and Gamma(6,3). The results of coverage proportions and interval average widths of the future sample variance using LC and MT methods for Beta(3,1) and Gamma(6,3) are presented in Tables 5 and 6, respectively. The notation CP_{LC} and AW_{LC} refer to the coverage proportion and average interval widths for the LC prediction interval, respectively. In the MT prediction interval, CP_{MT} and AW_{MT} represent coverage proportion and average interval widths, respectively.

The performance of different bootstrap procedures is first compared with the LC prediction interval. PP-B and NPI-B have good coverage in all cases of future sample sizes $m = n/2$ at confidence levels 95% and 90%. The

Bootstrap	measures	Confidence level							
		95%				90%			
		$n = 50$	$n = 100$	$n = 200$	$n = 400$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
PP-B	CP_{LC}	0.9270	0.9330	0.9330	0.9420	0.8770	0.8790	0.8810	0.8890
	CP_{MT}	0.9830	0.9860	0.9870	0.9920	0.9630	0.9660	0.9720	0.9740
	AW_{LC}	0.6289	0.4469	0.3172	0.2243	0.5245	0.3748	0.2663	0.1884
	AW_{MT}	0.8947	0.6314	0.4438	0.3135	0.7506	0.5321	0.3757	0.2657
NPI-B	CP_{LC}	0.9580	0.9440	0.9380	0.9510	0.9050	0.9050	0.8990	0.8950
	CP_{MT}	0.9930	0.9940	0.9890	0.9930	0.9820	0.9730	0.9760	0.9790
	AW_{LC}	0.7416	0.4927	0.3344	0.2312	0.6087	0.4089	0.2797	0.1940
	AW_{MT}	1.1587	0.7270	0.4786	0.3257	0.9137	0.5933	0.3994	0.2743
PB	CP_{LC}	0.8030	0.8160	0.8240	0.8280	0.7270	0.7460	0.7520	0.7460
	CP_{MT}	0.9240	0.9250	0.9310	0.9390	0.8840	0.8820	0.8790	0.8830
	AW_{LC}	0.4473	0.3176	0.2243	0.1589	0.3759	0.2668	0.1886	0.1337
	AW_{MT}	0.6219	0.4422	0.3118	0.2211	0.5288	0.3754	0.2653	0.1877
EB	CP_{LC}	0.8100	0.8100	0.8300	0.8290	0.7270	0.7450	0.7470	0.7550
	CP_{MT}	0.9180	0.9180	0.9290	0.9380	0.8730	0.8720	0.8870	0.8860
	AW_{LC}	0.4431	0.3150	0.2237	0.1587	0.3724	0.2648	0.1882	0.1334
	AW_{MT}	0.6152	0.4387	0.3116	0.2210	0.5225	0.3726	0.2647	0.1875

Table 4: *Coverage of* $100(1-2\alpha)\%$ *prediction interval for the mean of* m *future observations from Gamma(6,3), when* $m = n$.

superiority of PP-B is that it achieves good coverage with shorter intervals, where the average interval widths of PP-B are smaller than NPI-B in all cases. Additionally, its coverage proportions are closer to the coverage probabilities than those of NPI-B in most cases. In contrast, PB and EB show worse under-coverage results for all cases of Beta(3,1) and Gamma(6,2). Their coverage proportions with Beta(3,1) are at least 7% lower than 0.95 and 0.90 nominal coverage probabilities. Also, they provide coverage proportions that are at least 5.2% below their nominal coverage probabilities with Gamma(6,3).

The performance of different bootstrap methods is also compared based on MT prediction intervals. The wide average width of intervals in both PP-B and NPI-B leads to over-coverage for all cases. The MT method improves the coverage proportions of PB and EB, but PB is at least 4.4% below their nominal coverage probabilities with Beta(3,1) when $n = 50, 100, 200$ as shown in Table 5. Also, the EB method gives a result of 5.4% under-coverage below the nominal level of 95% and 6.6% lower than the nominal level of 90% with Gamma(6,3) when $n = 50$ as shown in Table 6. We observe that the MT method improves the coverage probability of PB and EB, however we can obtain a coverage proportion that is close to the nominal coverage probability using PP-B and NPI-B with the LC method. For example, the coverage of PP-B and NPI-B with Beta(3,1) based on LC method is closer to the nominal

Bootstrap	measures	Confidence level							
		95%				90%			
		$n = 50$	$n = 100$	$n = 200$	$n = 400$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
PP-B	CP_{LC}	0.9510	0.9460	0.9490	0.9410	0.8990	0.8800	0.9020	0.8760
	CP_{MT}	0.9800	0.9770	0.9810	0.9760	0.9560	0.9500	0.9550	0.9490
	AW_{LC}	0.0506	0.0361	0.0258	0.0183	0.0425	0.0303	0.0217	0.0154
	AW_{MT}	0.0618	0.0440	0.0315	0.0223	0.0520	0.0371	0.0265	0.0188
NPI-B	CP_{LC}	0.9660	0.9610	0.9570	0.9530	0.9320	0.9230	0.9070	0.8950
	CP_{MT}	0.9960	0.9880	0.9890	0.9880	0.9690	0.9660	0.9620	0.9570
	AW_{LC}	0.0629	0.0416	0.0279	0.0191	0.0522	0.0346	0.0233	0.0160
	AW_{MT}	0.0782	0.0514	0.0343	0.0233	0.0648	0.0428	0.0287	0.0196
PB	CP_{LC}	0.8110	0.8430	0.8310	0.8590	0.7250	0.7650	0.7600	0.7860
	CP_{MT}	0.8920	0.9060	0.9050	0.9240	0.8230	0.8520	0.8440	0.8730
	AW_{LC}	0.0417	0.0299	0.0211	0.0150	0.0353	0.0252	0.0178	0.0126
	AW_{MT}	0.0504	0.0363	0.0257	0.0183	0.0428	0.0307	0.0217	0.0154
EB	CP_{LC}	0.8700	0.8620	0.8800	0.8560	0.8050	0.7980	0.8050	0.7850
	CP_{MT}	0.9230	0.9170	0.9390	0.9350	0.8780	0.8750	0.8940	0.8710
	AW_{LC}	0.0405	0.0293	0.0210	0.0149	0.0344	0.0248	0.0177	0.0125
	AW_{MT}	0.0485	0.0355	0.0255	0.0181	0.0415	0.0301	0.0216	0.0153

Table 5: *Coverage of* $100(1-2\alpha)\%$ *prediction interval for the variance of* m *future observations from Beta(3,1), when* $m = n/2$.

coverage probabilities when $n = 50$ than PB and EB with MT method. The MT method enhances the coverage proportions of PB and EB by expanding the prediction interval width, but it provides under-coverage results in some cases. It is obvious that the PP-B method performs best for LC prediction intervals, as it is developed for predictive inference in line with the NPI-B method.

6. Prediction regions

In this section, we consider Banks' comparison method for prediction intervals to explore the performance of different bootstrap methods in predictive inference. Here, we intend to investigate the global measure of coverage accuracy for prediction intervals, which are called prediction regions. The main requirement for predection regions is that the nominal coverage probability closely resembles the actual coverage probability. This has motivated several simulations with accuracy at particular customary confidence levels, such as 0.99, 0.95 and 0.90. Banks [6] investigated the global measure of coverage accuracy to compare different bootstrap methods. The 20 prediction

Bootstrap	measures	Confidence level							
		95%				90%			
		$n=50$	$n=100$	$n=200$	$n=400$	$n=50$	$n=100$	$n=200$	$n=400$
PP-B	CP_{LC}	0.9300	0.9480	0.9450	0.9420	0.8670	0.8910	0.8980	0.8830
	CP_{MT}	0.9750	0.9840	0.9840	0.9770	0.9390	0.9530	0.9530	0.9490
	AW_{LC}	1.0694	0.7646	0.5456	0.3875	0.8631	0.6272	0.4520	0.3228
	AW_{MT}	1.3956	0.9715	0.6804	0.4787	1.1071	0.7890	0.5621	0.3989
NPI-B	CP_{LC}	0.9630	0.9750	0.9650	0.9660	0.9200	0.9410	0.9310	0.9250
	CP_{MT}	0.9900	0.9930	0.9960	0.9900	0.9680	0.9780	0.9720	0.9700
	AW_{LC}	2.2296	1.3233	0.8047	0.5070	1.5779	0.9800	0.6180	0.4025
	AW_{MT}	3.6831	2.0642	1.1798	0.6992	2.3717	1.3936	0.8430	0.5272
PB	CP_{LC}	0.8810	0.8980	0.8920	0.8800	0.8180	0.8320	0.8320	0.8110
	CP_{MT}	0.9510	0.9560	0.9480	0.9440	0.8920	0.9090	0.9030	0.8890
	AW_{LC}	0.9109	0.6378	0.4500	0.3176	0.7517	0.5302	0.3755	0.2664
	AW_{MT}	1.1492	0.7909	0.5526	0.3890	0.9399	0.6564	0.4630	0.3264
EB	CP_{LC}	0.8310	0.8640	0.8730	0.8730	0.7550	0.8080	0.8090	0.8010
	CP_{MT}	0.8960	0.9190	0.9200	0.9340	0.8340	0.8760	0.8790	0.8860
	AW_{LC}	0.7936	0.5934	0.4317	0.3102	0.6773	0.5020	0.3631	0.2612
	AW_{MT}	0.9565	0.7193	0.5244	0.3778	0.8137	0.6092	0.4431	0.3186

Table 6: *Coverage of* $100(1-2\alpha)\%$ *prediction interval for the variance of* m *future observations from Gamma(6,3), when* $m = n/2$.

regions with a nominal coverage probability of 0.05 can be obtained by

$$PRL_{(i)} = \left(q_{(\frac{\alpha_{i+1}}{2})}, q_{(\frac{\alpha_i}{2})}\right) \quad (8)$$

$$PRR_{(i)} = \left(q_{(1-\frac{\alpha_i}{2})}, q_{(1-\frac{\alpha_{i+1}}{2})}\right) \quad (9)$$

where $i = 1, 2, \ldots, 10$, $\alpha_{i+1} = \alpha_i - 0.10$, $\alpha_1 = 1$ and $q_{(z)}$ is the z^{th} quantile of statistical values, so $PRL_{(i)}$ and $PRR_{(i)}$ are the prediction regions representing the left tail and right tail of the global measure of coverage accuracy, respectively. A total of 10 prediction regions are created, each with a nominal coverage probability of 0.10 can be obtained using Equations (8) and (9) as follows:

$$PR_{(i)} = PRL_{(i)} \cup PRR_{(i)} \quad (10)$$

Both divisions of prediction regions are used to show the best bootstrap method that have the closest true coverage probability to the nominal coverage probability for a specific parameter of interest. Banks [6] used a chi-squared test of goodness of fit to assess the discrepancy in coverage proportion with different parameters, distributions and sample sizes, to compare his bootstrap method to other bootstrap techniques, e.g. Efron's method [20], Rubin's Bayesian bootstrap [38] and smoothed Rubin's bootstrap [6]. He

$PR_{(i)}$	PP-B	NPI-B	PB	EB
1	0.096	0.108	0.073	0.068
2	0.106	0.097	0.071	0.070
3	0.090	0.119	0.072	0.077
4	0.094	0.091	0.076	0.077
5	0.104	0.087	0.078	0.072
6	0.102	0.112	0.072	0.065
7	0.110	0.113	0.090	0.098
8	0.103	0.092	0.100	0.106
9	0.090	0.103	0.144	0.130
10	0.105	0.078	0.224	0.237

Table 7: *The coverage proportions for the mean in the 10 prediction regions, where* $m = n = 50$.

considered the best bootstrap method to be the one having the lowest chi-squared (χ^2) values. We here intend to use the prediction regions technique for comparison of PP-B with other methods of bootstrap, described in Section 2. The following processes are used to study the coverage proportions in the 10 and 20 prediction regions for the future sample statistic based on the bootstrap method:

1. Draw a sample $X = (x_1, \dots, x_n)$ of n observations from a specific distribution to be the past sample and then draw a sample $Y = (y_1, \dots, y_m)$ of m observations from the same distribution to be the future sample. The samples are assumed to be independent samples.
2. Compute the statistic of the Y sample, T_m.
3. Draw B bootstrap samples of size m from the X sample and compute the statistic T_m^* for each bootstrap sample to obtain a list of $T_m^*(j)$ for $j = 1, \dots, B$.
4. Create the 10 and 20 prediction regions for T_m by Equations (8), (9) and (10).
5. Determine if these prediction regions include the statistic T_m.
6. Steps 1-5 are performed in total N times in order to find the coverage proportions.

A simulation study is conducted as previously described, employing various bootstrap methods to estimate the coverage proportions for the 10 and 20 prediction regions. Simulations are performed $N = 1000$ times for the mean

method	PP-B		NPI-B		PB		EB	
i	$PRL_{(i)}$	$PRR_{(i)}$	$PRL_{(i)}$	$PRR_{(i)}$	$PRL_{(i)}$	$PRR_{(i)}$	$PRL_{(i)}$	$PRR_{(i)}$
1	0.051	0.045	0.057	0.051	0.042	0.031	0.039	0.029
2	0.056	0.050	0.047	0.050	0.040	0.031	0.038	0.032
3	0.043	0.047	0.058	0.061	0.034	0.038	0.038	0.039
4	0.049	0.045	0.042	0.049	0.039	0.037	0.038	0.039
5	0.055	0.049	0.045	0.042	0.042	0.036	0.041	0.031
6	0.051	0.051	0.053	0.059	0.034	0.038	0.030	0.035
7	0.054	0.056	0.055	0.058	0.046	0.044	0.050	0.048
8	0.057	0.046	0.033	0.059	0.050	0.05	0.051	0.055
9	0.044	0.046	0.044	0.059	0.066	0.078	0.060	0.070
10	0.052	0.053	0.028	0.050	0.113	0.111	0.119	0.118

Table 8: *The coverage proportions for the mean in the 20 prediction regions, where* $m = n = 50$.

of $m = n$ future observations with a sample size of 50 from Beta(3,1). The bootstrap methods are applied to each past sample $B = 1000$. The coverage proportions for the mean in the 10 and 20 prediction regions are outlined in Tables 7 and 8, respectively. The PP-B and NPI-B methods illustrate their superiority in achieving coverage proportions in each of the 10 and 20 prediction regions close to 0.10 and 0.05, respectively. In contrast, the PB and EB methods lead to coverage proportions far from the nominal level of 0.10 in most of the 10 prediction regions, and far from 0.05 in most of the 20 prediction regions. We use the chi-square test to assess the discrepancy between the nominal coverage probabilities and coverage proportions in order to show the best bootstrap method. The resulting χ^2 values are presented in the first row of Table 9. The PP-B and NPI-B methods achieve good coverage accuracy, which is reflected by the low chi-squared value in both divisions of the prediction regions. They make the discrepancies between coverage proportions and nominal coverage probabilities lower than the other bootstrap methods. Simulations were repeated several times and consistent results were obtained, as illustrated in Table 9. It is obvious that the PP-B method shows its superiority to the other bootstrap methods in achieving the smallest chi-squared values. It distributes the coverage proportions more accurately in most of the prediction region divisions than the other bootstrap methods and this is apparent from having the lowest chi-squared values.

A variety of sample sizes are considered to determine whether the size of the sample affects the performance of different bootstrap methods. Table 10

Repetition	10 PR				20 PR			
	PP-B	NPI-B	PB	EB	PP-B	NPI-B	PB	EB
1	4.42	15.54	216.10	247.00	7.12	30.96	220.36	250.84
2	3.90	9.14	225.62	252.12	13.44	25.16	233.08	259.00
3	5.46	9.34	236.24	253.40	10.88	27.96	241.12	261.44
4	5.48	10.72	234.06	246.24	11.52	27.64	236.24	251.92
5	4.70	11.70	246.48	262.76	11.28	32.64	258.04	265.60
6	4.72	10.42	233.60	241.06	9.80	32.16	238.00	246.12
7	5.34	15.40	227.66	254.52	12.36	32.64	230.64	261.84
8	6.96	10.24	227.04	248.82	18.24	25.56	233.08	253.96
9	6.88	15.76	239.58	238.62	12.48	36.40	243.04	240.52
10	3.18	10.76	227.48	261.34	9.04	33.52	229.96	265.16

Table 9: *The chi-squared values obtained from coverage proportions for the mean, where* $m = n = 50$.

n	measures	10 PR				20 PR			
		PP-B	NPI-B	PB	EB	PP-B	NPI-B	PB	EB
30	χ^2	7.36	16.30	235.78	266.18	25.32	45.60	261.28	287.96
	p-value	0.600	0.061	0.000	0.000	0.150	0.001	0.000	0.000
100	χ^2	7.50	7.56	303.52	314.24	15.32	18.76	306.64	321.48
	p-value	0.585	0.579	0.000	0.000	0.702	0.472	0.000	0.000
200	χ^2	9.66	22.28	281.50	289.32	16.56	31.80	287.28	296.24
	p-value	0.379	0.008	0.000	0.000	0.620	0.033	0.000	0.000
300	χ^2	5.44	6.32	244.30	250.02	13.60	14.48	250.20	252.92
	p-value	0.794	0.708	0.000	0.000	0.806	0.755	0.000	0.000

Table 10: *The chi-squared values for the mean and their* p*-values with different sample sizes* $m = n$.

presents the chi-squared values obtained from the coverage proportions for the mean using different bootstrap methods at different sample sizes $m = n$. The chi-squared values for all bootstrap methods show no clear pattern as the sample size increases. In both prediction regions, chi-squared values are consistent across all bootstrap methods regardless of sample size. The PP-B method performs better in both prediction region divisions at different sample sizes than any other bootstrap method, followed by NPI-B. For both the PB and the EB methods, the chi-squared values are large because of the great discrepancies between the nominal coverage probabilities and coverage proportions.

We also evaluate different bootstrap methods based on the global accuracy of prediction intervals for the variance. The chi-squared goodness of fit test is used as the basis for comparing the performance of the bootstrap method. We consider several sample sizes to investigate whether or not sample size affects the performance of different bootstrap techniques. In Table 11, we present the chi-squared values obtained from the coverage proportions for the variance based on different bootstrap methods with different sample sizes $m = n$. The results of this table are computed in the same manner as before, to demonstrate the performance of these bootstrap techniques. The results of χ^2 values in both prediction region divisions indicate that PP-B and NPI-B are both performing better than any other bootstrap methods. The reason for this is that both methods are able to distribute coverage proportions more accurately across 10 and 20 prediction regions. In contrast, the χ^2 values of PB and EB are high due to the great discrepancies between the nominal coverage probabilities and coverage proportions. The PP-B method has the lowest chi-squared value among these bootstrap methods, which indicates its superiority in achieving coverage proportions close to nominal levels in most of the prediction region divisions.

A simulation study is conducted as previously described, employing various bootstrap methods to estimate the coverage proportions for the 10 and 20 prediction regions. In this simulation, data are generated from N(0,1), using various combinations of sample sizes where $m = n/2$. Simulations are performed $N = 1000$ times for the variance of future observations. The bootstrap methods are applied to each past sample with $B = 1000$ times. The coverage proportions for the variance in the 10 and 20 prediction regions are outlined in Table 12. We observe that the PP-B method consistently outperforms other bootstrap methods across various sample sizes in both the 10 and 20 prediction regions. This superior performance is evidenced by

n	measures	10 PR				20 PR			
		PP-B	NPI-B	PB	EB	PP-B	NPI-B	PB	EB
50	χ^2	16.20	28.82	220.08	327.56	53.20	88.24	232.84	367.44
	p-value	0.063	0.001	0.000	0.000	0.000	0.000	0.000	0.000
100	χ^2	11.56	19.44	257.24	325.08	45.80	53.20	288.64	383.48
	p-value	0.239	0.022	0.000	0.000	0.001	0.000	0.000	0.000
200	χ^2	13.64	15.14	177.42	193.54	23.84	46.04	182.96	205.48
	p-value	0.136	0.087	0.000	0.000	0.005	0.000	0.000	0.000
300	χ^2	7.12	13.02	242.42	268.70	22.32	31.80	267.08	292.52
	p-value	0.625	0.162	0.000	0.000	0.269	0.033	0.000	0.000

Table 11: *The chi-squared values for the variance and their p-values with different sample sizes $m = n$.*

n	measures	10 PR				20 PR			
		PP-B	NPI-B	PB	EB	PP-B	NPI-B	PB	EB
100	χ^2	7.58	14.98	102.96	148.56	37.52	52.52	113.68	172.24
	p-value	0.577	0.091	0.000	0.000	0.007	0.000	0.000	0.000
200	χ^2	5.76	9.60	88.72	129.38	22.96	31.60	101.36	154.52
	p-value	0.764	0.384	0.000	0.000	0.239	0.035	0.000	0.000
300	χ^2	8.14	10.7	111.88	118.08	19.24	42.04	116.40	128.44
	p-value	0.520	0.297	0.000	0.000	0.442	0.002	0.000	0.000
400	χ^2	5.60	9.60	73.82	83.06	18.52	44.76	81.84	95.16
	p-value	0.779	0.384	0.000	0.000	0.488	0.001	0.000	0.000

Table 12: *The chi-squared values for the variance and their p-values with different sample sizes $m = n/2$.*

lower chi-squared values, indicating smaller discrepancies between the nominal coverage probabilities and the observed coverage proportions. The NPI-B method also shows better performance compared to PB and EB methods. In contrast, the PB and EB methods have higher chi-squared values, reflecting greater discrepancies between the nominal coverage probabilities and coverage proportions.

7. Conclusions and future works

This paper has introduced a new version of bootstrap, which we call the parametric predictive bootstrap (PP-B). The proposed bootstrap method is explicitly aimed at predictive inference. The confidence intervals and prediction intervals are being used to assess the strength of estimation and prediction inference of PP-B. We evaluate its performance via simulations, which show that it works well as a method for predictive inference. These results motivate us to investigate further properties and performance of PP-B for other inferences.

In future works, it would be of interest to evaluate this new bootstrap method in a range of scenarios that have been used with other bootstrap methods. Another work which is currently being done is evaluating the performance of PP-B with reproducibility of tests [2, 3]. A hypothesis test is one of the most important tools in the practical application of statistics. Statistical hypothesis tests can have different results each time they are repeated. The reproducibility probability of tests (RP) has gained increasing attention due to its importance in evaluating the variability and the stability of test results. Test reproducibility is more naturally viewed as a prediction problem than as an estimation problem. The explicitly predictive nature of PP-B provides an appropriate formulation for inferring RP, as the nature of RP is explicitly predictive as well.

Acknowledgements

This research took place during Abdulrahman Aldawsari's PhD at the Department of Mathematical Sciences, Durham University. Abdulrahman is grateful to Prince Sattam bin Abdulaziz University in Saudi Arabia for financial support (project number PSAU/2024/R/1446).

References

[1] Asamh S M Al Luhayb, F P A Coolen, and Tahani Coolen-Maturi. Generalizing Banks' smoothed bootstrap method for right-censored data. In *Proceedings of the 29th European Safety and Reliability Conference (ESREL 2019)*, pages 894–901, Hannover, Germany, 2019.

[2] Abdulrahman Aldawsari. *Parametric Predictive Bootstrap and test reproducibility*. PhD thesis, Durham University, 2023.

[3] Abdulrahman Aldawsari, Tahani Coolen-Maturi, and Frank Coolen. Parametric predictive bootstrap method for the reproducibility of hypothesis tests. *Journal of Statistical Theory and Practice*, 19(2):1–27, 2025.

[4] Thomas Augustin and Frank P A Coolen. Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, 124(2):251–272, 2004.

[5] Thomas Augustin, Frank PA Coolen, Gert De Cooman, and Matthias CM Troffaes. *Introduction to imprecise probabilities*. John Wiley & Sons, 2014.

[6] David L Banks. Histospline smoothing the Bayesian bootstrap. *Biometrika*, 75(4):673–684, 1988.

[7] Daniel Berrar. Introduction to the non-parametric bootstrap. In *Encyclopedia of Bioinformatics and Computational Biology*, pages 766–773. Academic Press Oxford, 2019.

[8] Wai Chan, Yiu-Fai Yung, Peter M Bentler, and Man-Lai Tang. Tests of independence for ordinal data using bootstrap. *Educational and Psychological Measurement*, 58(2):221–240, 1998.

[9] Michael R Chernick. *Bootstrap Methods: A Guide for Practitioners and Researchers*. John Wiley & Sons, 2011.

[10] Michael R Chernick and Robert A LaBudde. *An Introduction to Bootstrap Methods with Applications to R*. John Wiley & Sons, 2014.

[11] Frank P A Coolen. On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information*, 15(1-2):21–47, 2006.

[12] Frank P A Coolen. Nonparametric predictive inference. In Miodrag Lovric, editor, *International Encyclopedia of Statistical Sciences*, pages 968–970. Springer, Berlin, 2011.

[13] Frank P A Coolen and Mohamed A Elsaeiti. Nonparametric predictive methods for acceptance sampling. *Journal of Statistical Theory and Practice*, 3(4):907–921, 2009.

[14] Frank P A Coolen and Sulafah Bin Himd. Nonparametric predictive inference bootstrap with application to reproducibility of the two-sample Kolmogorov–Smirnov test. *Journal of Statistical Theory and Practice*, 14(2):1–13, 2020.

[15] Frank P A Coolen and Filipe J Marques. Nonparametric predictive inference for test reproducibility by sampling future data orderings. *Journal of Statistical Theory and Practice*, 14(4):1–22, 2020.

[16] Tahani Coolen-Maturi, Pauline Coolen-Schrijner, and Frank PA Coolen. Nonparametric predictive inference for binary diagnostic tests. *Journal of Statistical Theory and Practice*, 6:665–680, 2012.

[17] Pauline Coolen-Schrijner, Tahani A Maturi, and Frank P A Coolen. Nonparametric predictive precedence testing for two groups. *Journal of Statistical Theory and Practice*, 3(1):273–287, 2009.

[18] Anthony Christopher Davison and David Victor Hinkley. *Bootstrap Methods and their Application.* Cambridge University Press, 1997.

[19] Persi Diaconis and Bradley Efron. Computer-intensive methods in statistics. *Scientific American*, 248(5):116–131, 1983.

[20] Bradley Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7:1–26, 1979.

[21] Bradley Efron. Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, 9(2):139–158, 1981.

[22] Bradley Efron. More efficient bootstrap computations. *Journal of the American Statistical Association*, 85(409):79–89, 1990.

[23] Bradley Efron and Gail Gong. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36–48, 1983.

[24] Bradley Efron and Robert Tibshirani. The bootstrap method for assessing statistical accuracy. *Behaviormetrika*, 12(17):1–35, 1985.

[25] Bradley Efron and Robert J Tibshirani. *An Introduction to the Bootstrap.* Boca Raton, Florida : Chapman & Hall/CRCl, 1994.

[26] Peter Hall. Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics*, pages 927–953, 1988.

[27] Tim Hesterberg. Bootstrap. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(6):497–526, 2011.

[28] Bruce M Hill. Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association*, 63(322):677–691, 1968.

[29] Bruce M Hill. De Finetti's theorem, induction, and A(n) or bayesian nonparametric predictive inference (with discussion). *Bayesian Statistics*, 3:211–241, 1988.

[30] Bruce M Hill. Parametric models for A_n: splitting processes and mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(2):423–433, 1993.

[31] Kris N Kirby and Daniel Gerlanc. Bootes: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods*, 45(4):905–927, 2013.

[32] Ming-Che Lu and Dong Shang Chang. Bootstrap prediction intervals for the birnbaum-saunders distribution. *Microelectronics Reliability*, 37(8):1213–1216, 1997.

[33] Michael A Martin. On bootstrap iteration for coverage correction in confidence intervals. *Journal of the American Statistical Association*, 85(412):1105–1118, 1990.

[34] Majid Mojirsheibani. Iterated bootstrap prediction intervals. *Statistica Sinica*, pages 489–504, 1998.

[35] Majid Mojirsheibani and Robert Tibshirani. Some results on bootstrap prediction intervals. *Canadian Journal of Statistics*, 24(4):549–568, 1996.

[36] Maria L Rizzo. *Statistical Computing with R*. Boca Raton ; London : CRC Press, 2008.

[37] Gerd K Rosenkranz. Bootstrap corrections of treatment effect estimates following selection. *Computational Statistics & Data Analysis*, 69:220–227, 2014.

[38] Donald B Rubin. The Bayesian bootstrap. *The Annals of Statistics*, 9:130–134, 1981.

[39] BW Silverman and GA Young. The bootstrap: to smooth or not to smooth? *Biometrika*, 74(3):469–479, 1987.