Research Article GU J Sci 39(2): 1020-1032 (2026) DOI: 10.35378/gujs.1817298

Gazi University

Journal of Science

http://dergipark.org.tr/gujs

Using Parametric Bootstrapping for Estimation of the Incidence of Inefficiency: Improving Features of Non-Parametric Bootstrap Estimator

Denız OZONUR[1,*], Mehmet Guray UNSAL[2]

[1]Gazi University, Faculty of Science, Department of Statistics, Ankara, Türkiye
[2]Başkent University, Faculty of Business and Administrative Sciences, Department of Technology and Knowledge Management, Ankara, Türkiye

Highlights

- This study provides bootstrap based simulations for estimating the incidence of inefficiency (IOI).
- ML-based parametric bootstrapping yields better results even with small sample sizes.
- The Bayesian estimator outperforms as the number of input and output variables increases.

Article Info

Received: 4 Nov 2025
Accepted: 30 Mar 2026

Keywords

Bootstrapping
Estimation
Incidence of inefficiency
Simulation

Abstract

This study proposes parametric bootstrap estimation methods to improve features of the non-parametric bootstrap estimator of Incidence of Inefficiency proposed in the literature. This study is the first to propose parametric bootstrap estimation methods to improve features of the IOI estimators. In the simulation study, the Maximum Likelihood-based parametric bootstrap method yields the best results in small sample sizes and a limited number of input and output variable situations. However, in cases where the number of input and output variables increases, which reduces the discrimination power of classical Data Envelopment Analysis models, the Bayesian estimator with latent variable adjustment tends to yield better results than the proposed estimators for IOI in the literature. Additionally, it is experimentally demonstrated that the parametric bootstrap based on the Bayesian method applied to a specific posterior distribution converges to the same features as the estimators obtained with the Bayesian estimator based on that posterior distribution.

1. INTRODUCTION

The Incidence of Inefficiency (IOI) is a term introduced to the literature by [1] within the scope of Data Envelopment Analysis (DEA). It is calculated as the ratio of inefficient decision-making units (DMUs) to all DMUs in the analysis. Various studies in the literature focus on estimating the population IOI based on the IOI value obtained from the sample [1–5]. These studies can be categorized into two main approaches: Bayesian estimation and bootstrap estimation.

Firstly, Friesner et al. [1] introduce the IOI and demonstrate that Bayesian estimation methods could be used to estimate the IOI. Ünsal et al. [2] examine posterior distributions associated with the IOI in DEA, focusing on scenarios where sampling is conducted with replacement or when the population size is infinity. Ünsal et al. [3] work on the Curse of Dimensionality (COD) subject based on the IOI. They choose an appropriate prior structure using a simulation approach to investigate the empirical distribution of the IOI. They conclude that using left skewed priors gives better estimates in terms of mean absolute error measure. Ünsal [4] obtains alternative Bayesian Estimators for the IOI by considering right skewed prior in the cases without a rule of thumb (for the cases when the discrimination power problem is valid). Ünsal et al. [5] modify the Bayesian Estimators proposed by [1,2]. They also compare the performance of Bayesian estimators and nonparametric bootstrapping in the estimation of the IOI term. In Ünsal et al. [5]'s study, Bayesian estimators generally perform better than Bootstrap estimators, except in cases of small sample sizes and limited input-output scenarios. At this point, an interesting research question arises: Whether the

*Corresponding author, e-mail: denizozonur@gazi.edu.tr

predictive properties of non-parametric bootstrap estimators can be improved, and whether better or comparable results to Bayesian estimators can be achieved using parametric bootstrapping.

Even though they are different subjects, estimation of the efficiency scores and estimation of the IOI are sometimes conflated. Kneip et al. [6] derive the asymptotic distribution of DEA estimators, and they use non-parametric bootstrap to estimate the efficiency scores. It is commonly known that the estimated frontier does not coincide with the true frontier, therefore, a latent variable correction is attempted to be integrated into the estimation process in Bayesian estimation of IOI [1]. A substantial and expanding body of research has explored the application of Bayesian methods for modeling and evaluating latent variable issues across various theoretical and empirical settings [7-9]. Without disregarding the fact that DEA is a non parametric method to determine efficiency and inefficiency, this study aims to model the distributional structure of the proportion of inefficient DMUs. In this context, the study builds upon a left-skewed empirical distribution structure, in line with the approach adopted by [3]. A parametric bootstrap procedure is then implemented based on this probability distribution.

From a methodological perspective, non-parametric and parametric bootstrap approaches differ fundamentally in their theoretical foundations. The non-parametric bootstrap does not impose any distributional assumptions on the data-generating process; instead, it relies on repeated resampling from the empirical distribution of the observed data. In contemporary statistical research, non-parametric bootstrap procedures play a central role in several advanced fields. In contemporary statistical research, non-parametric bootstrap procedures play a central role in several advanced fields, including robust statistics, financial econometrics, biomedical data analysis, and environmental modeling. For instance, Heng and Lange [10] present a non-parametric bootstrap procedure to estimate the proportion of outliers within a robust regression framework. They repeatedly resample the observed data and recompute the estimated contamination rate in each bootstrap sample to approximate its sampling distribution and construct confidence intervals without relying on parametric distributional assumptions. Wu et al. [11] apply a nonparametric two-sample bootstrap to estimate the uncertainty of ROC curve measures, particularly the standard error of the area under the curve (AUC), for large datasets. They validate the bootstrap results by comparing them to the analytical Mann-Whitney method, showing close agreement and demonstrating that the bootstrap provides a reliable probability distribution for uncertainty estimation.

In contrast, the parametric bootstrap relies on a fully specified probabilistic model for the data-generating process, rather than resampling directly from the empirical distribution. After estimating the model parameters, new samples are generated from the fitted parametric distribution. By incorporating structural information about the assumed probability model, this approach can yield more efficient and precise inference when the model is correctly specified. Parametric bootstrap methods are widely applied in reliability analysis, survival analysis, and other model-based statistical frameworks where specific stochastic assumptions are theoretically justified. For example, in industry, Zhuang et al. [12] develop a multivariate reparameterized inverse Gaussian (rIG) model to capture dependence among multiple performance characteristics that deteriorate simultaneously under a shared failure mechanism. They estimate the model parameters using an expectation–maximization algorithm and employ a parametric bootstrap procedure to generate confidence intervals by simulating new degradation paths from the fitted rIG model, thereby quantifying uncertainty under the assumed stochastic structure.

In the context of IOI estimation, adopting a parametric bootstrap framework allows the modeling of the distributional characteristics of the proportion of inefficient DMUs while preserving the flexibility of resampling-based inference. Therefore, investigating whether parametric bootstrap estimators can achieve predictive performance comparable to or better than Bayesian estimators constitutes a relevant methodological contribution to the literature.

In this study, parametric versions of the bootstrap estimator are proposed as alternative methods for estimating the IOI. In the previous study by Ünsal et al. [5], the comparison includes Bayesian estimators and a non-parametric bootstrap estimator of the IOI. The present study extends this comparison by incorporating a parametric bootstrap estimator of the IOI, in addition to the estimators considered in Ünsal et al. [5], thereby providing originality to this research. The parametric bootstrap estimators based on both

the ML and the Bayesian methods are compared with previously developed Bayesian estimators and the non-parametric bootstrap estimator in terms of predictive features such as bias and Mean Squared Error (MSE). In this respect, it represents the most comprehensive simulation study in the literature on this subject. The study is constructed as follows: Section 2 investigates Bayesian and non-parametric bootstrapping approaches for estimation of the IOI and parametric bootstrapping methods used in this study are also presented. Section 3 gives the results of simulation studies to compare all these estimators in terms of features. Section 4 includes a real-world data application for a multi-dimensional case. Section 5 concludes the study.

2. MATERIAL METHOD

2.1. Bayesian and Bootstrap Estimators of the IOI

Friesner et al. [1] introduce the concept of the IOI and estimate it using the Bayesian estimation method. They adopt a non-informative prior distribution, specifically a uniform distribution, for the Bayesian estimators of the IOI. Additionally, they incorporate sample information employing the Binomial distribution for sampling with replacement and the Hypergeometric distribution for sampling without replacement. Ünsal et al. [2] derive posterior distributions for the IOI scores in DEA under two scenarios: when the population approaches infinity and when the population is finite with sampling conducted with replacement. Ünsal et al. [3] investigate the empirical distribution of the IOI, and according to the results of this study, they use left skewed priors as an informative prior distribution structure in obtaining the Bayesian estimator for the IOI. As for the information derived from the sample, they adhered to the concept proposed by [1]. The studies on Bayesian estimators for the IOI are not limited to these. Furthermore, Ünsal [4] develops alternative Bayesian estimators for the IOI term using a right-skewed prior in situations where no rule of thumb applies, particularly in cases where the discrimination power issue is relevant. In rule of thumb cases, the IOI tends to exhibit a right-skewed distribution structure, as already identified in the findings of [3]. In Ünsal [4]'s research, this information is utilized to develop Bayesian estimators for rule of thumb scenarios. As mentioned, the IOI estimation is not limited to the use of the Bayesian estimation method.

Ünsal et al. [5] estimate the IOI using bootstrapping. They compare two methods (one based on Bayesian methods and one based on non-parametric bootstrapping methods) for the estimation of the population IOI. In this most recent study, a comparison is made between the Bayesian estimator obtained combining information derived from the sample, represented by the Binomial distribution, with an informative prior (a left-skewed distribution) under the assumption of the latent version (see [1]) of the IOI, and the estimator derived using the bootstrapping approach. The binomial probability distribution can be used to express the information from the sample as below:

$$P(x|\pi,n) = C(n,x)\,\pi^{x}(1-\pi)^{n-x}. \tag{1}$$

Here, π represents the proportion of inefficient firms (or units) in the population, in other words, parameter value of the IOI term, n is the number of DMUs in the sample, x is the number of DMUs classified as inefficient in the sample. In the study of [5], the prior structure of π is also given below:

$$P(\pi) = \frac{2N\pi}{(N-x)(N+x-1)}. \tag{2}$$

Here, N is the population size (the number of DMUs in the population), and the true range of the IOI in the population is from x/N to $(N-1)/N$. $C(n,x)$ is combination of n items taken x at a time (n choose x). The posterior structure of this Bayesian estimator is described in [3, 5] can be given as follows:

$$P(\pi|x,n) = \frac{P(\pi)P(x|\pi,n)}{\sum P(\pi)P(x|\pi,n)} = \frac{\frac{2N\pi}{(N-x)(N+x-1)}\,C(n,x)\pi^{x}(1-\pi)^{n-x}}{\sum_{\pi}\frac{2N\pi}{(N-x)(N+x-1)}\,C(n,x)\,\pi^{x}(1-\pi)^{n-x}}. \tag{3}$$

It is also possible to derive the expected value of this posterior as Bayesian estimator (minumum posterior quadratic risk-MPQR) and its latent version as mentioned in [5] (for theoretical procedures for obtaining the posterior distribution, the estimator and definition of notations and indices, see Appendix in [5]).

Except for certain special cases in simulations (such as small sample sizes and scenarios with a limited number of input and output variables), non-parametric version of bootstrap estimators has failed to outperform or even match the performance of Bayesian estimators in terms of absolute error measure according to the findings in [5]. However, the effectiveness of the parametric bootstrap approach and its adequacy in improving the properties of estimators are well-known in the literature [13,14].

In this study, the parametric bootstrap approaches are used for the IOI estimation to enhance the performance of the bootstrap estimator in terms of bias and MSE. Through detailed simulation studies, all the approaches previously proposed in the literature are examined, and their performance is simultaneously compared with the parametric bootstrap estimators discussed in this paper. As mentioned above, unlike the previous studies in the literature, this comparison is made not only based on bias but also on MSE values in order to see the properties of estimators from a broader perspective.

2.2. Parametric Bootstrapping for Estimation of the IOI

Parametric bootstrap is a resampling technique used to estimate the uncertainty of parameter estimates generating new datasets based on a fitted parametric model. It relies on the assumption that the data follows a known probability distribution. Parametric bootstrap is a computational method used when classical estimation methods often rely on large-sample approximations. This method can be more accurate in small-sample scenarios, where analytical results may not hold well. Efron and Tibshirani [15] introduce various bootstrap techniques, including parametric bootstrap, and provide an accessible explanation of their implementation. The parametric bootstrap estimation method is widely used in the literature [16–20].

This study aims to use the parametric bootstrap estimation method for the first time in estimating the population IOI. Before applying parametric bootstrap estimation methods, population sizes (N), parameter values (π), sample sizes (n), and the numbers of inputs and outputs (m, s) are determined. In the simulation, the B value (the number of generated data for each replication) is determined as 1000, which is a commonly-used value in parametric bootstrapping in the literature [13,15,21].

The parametric bootstrap procedure is conducted by following the steps for Bayesian estimation of IOI values:

i) Obtain the initial estimate of the IOI value from the observed sample (n).
ii) Draw a set of parameter IOI values $\pi^{(b)}$ ($b=1,2,...,B$) from the CDF of the posterior distribution using the estimated IOI values.
iii) Generate a new bootstrap dataset $x^{(b)}$ from Binomial distribution using the $\pi^{(b)}$ IOI values.
iv) Fit the Bayesian model to the bootstrap dataset $x^{(b)}$ and obtain the Bayesian estimates of IOI values as parametric bootstrap Bayesian estimates.
v) Repeat the steps (ii)-(iv) for B number of times.

To obtain parametric bootstrap ML of IOI values, the following steps are applied:

i) Obtain the initial estimate of the IOI value from the observed sample (n).
ii) Generate a new bootstrap dataset $x_*^{(b)}$ ($b=1,2,...,B$) from Binomial distribution using the estimated IOI values.
iii) For each bootstrap dataset $x_*^{(b)}$, calculate the ML of IOI values as parametric bootstrap ML.
iv) Repeat the steps (ii)-(iii) for B number of times.

In summary, although both procedures follow the same resampling framework, they differ fundamentally in the way the parameter is handled during the bootstrap stage. The Bayesian parametric bootstrap

repeatedly samples parameter values from the posterior distribution and then generates data conditional on these draws. In contrast, the ML-based parametric bootstrap conditions on a single fixed point estimate obtained from the observed data and generates all bootstrap samples using this value. Thus, the distinction between the two approaches arises from whether parameter uncertainty is explicitly incorporated into the resampling process.

3. SIMULATION STUDY

A simulation analysis is implemented to compare the relative performance of parametric, non-parametric bootstrapping, and Bayesian methods in estimating the population IOI. As different from the past studies related to the estimation of the IOI, the performances are compared based on estimator values and MSE values. The production process used to generate the simulation data is designed following [22], similar to [3,5]. The simulation results include both bootstrap estimators based on parametric and non-parametric approaches and also Bayesian estimators based on informative-non informative priors and latent-non latent versions, separately mentioned in [3,5]. From this aspect, it is the most comprehensive study in the literature to see the IOI estimation performance of the estimators. The simulation is conducted twice, once where the production process allows for a single ray technology and once where it allows for multiple rays production technology. The analysis is replicated several times, each time adjusting for different population sizes (N), parameter values (π), sample sizes (n), and numbers of inputs and outputs (m, s). Each of these simulation parameters is consistent with guidelines in the literature related to the rule of thumb discussed in [23]. In other words, the selected values of N, π, m, s and n in the simulation study ensure adequate discrimination power of the DEA models, consistent with the commonly used rule of thumb ($n \geq 3(m+s)+1$). In this manuscript, the simulations for each combination are conducted with R=1000 replications, which should be sufficient according to [24] and [25].

The simulation analysis for a single ray technology is summarized in Tables 1 and 2. Table 1 gives estimator values, and Table 2 indicates the results of MSE values for single ray technology.

Table 1. *Estimator values for the single ray technology*

					Non-Informative Prior		Informative Prior		Bootstrapping		
N	π	m	s	n	Non-Latent	Latent	Non-Latent	Latent	Non-Parametric	Parametric Bayesian	Parametric ML
200	0.9950	1	2	20	0.899981	0.947979	0.904323	0.950138	0.934011	0.912013	0.939994
				40	0.948670	0.969677	0.949857	0.970330	0.967789	0.948059	0.971297
				60	0.965493	0.977770	0.966033	0.978088	0.977901	0.962188	0.981241
				80	0.974330	0.981846	0.974631	0.982040	0.983425	0.969639	0.986498
200	0.9750	2	1	20	0.774364	0.930963	0.784174	0.933896	0.784418	0.898053	0.801760
				40	0.868643	0.954266	0.871698	0.955303	0.872785	0.934364	0.887121
				60	0.903548	0.962112	0.905079	0.962701	0.902873	0.947897	0.917026
				80	0.924512	0.967229	0.925422	0.967617	0.921168	0.956186	0.935109
250	0.9640	2	2	20	0.702000	0.921294	0.714957	0.924674	0.701485	0.890223	0.722183
				40	0.820976	0.944891	0.825140	0.946159	0.820665	0.926195	0.837155
				60	0.869323	0.952568	0.871397	0.953315	0.865655	0.939329	0.881690
				80	0.896415	0.957360	0.897663	0.957871	0.889765	0.947088	0.906401

Table 2. *MSE values for the single ray technology*

					Non-Informative Prior		Informative Prior		Bootstrapping		
N	π	*m*	*s*	*n*	Non-Latent	Latent	Non-Latent	Latent	Non-Parametric	Parametric Bayesian	Parametric ML
200	0.9950	1	2	20	0.009406	0.002372	0.008568	0.002159	0.004332	0.006996	0.003489
				40	0.002226	0.000711	0.002113	0.000675	0.000915	0.002259	0.000650
				60	0.000903	0.000333	0.000870	0.000321	0.000391	0.001107	0.000224
				80	0.000439	0.000195	0.000427	0.000189	0.000187	0.000661	0.000086
200	0.9750	2	1	20	0.042546	0.002718	0.038510	0.002399	0.039645	0.006445	0.032785
				40	0.011969	0.000779	0.011299	0.000720	0.011579	0.001929	0.008450
				60	0.005448	0.000382	0.005220	0.000360	0.005914	0.000917	0.003726
				80	0.002770	0.000201	0.002673	0.000192	0.003392	0.000475	0.001824
250	0.9640	2	2	20	0.074332	0.002937	0.067227	0.002562	0.076303	0.006204	0.065384
				40	0.022372	0.000958	0.021110	0.000883	0.023277	0.001908	0.018211
				60	0.009780	0.000513	0.009365	0.000484	0.010986	0.000934	0.007646
				80	0.005031	0.000292	0.004853	0.000279	0.006357	0.000504	0.003803

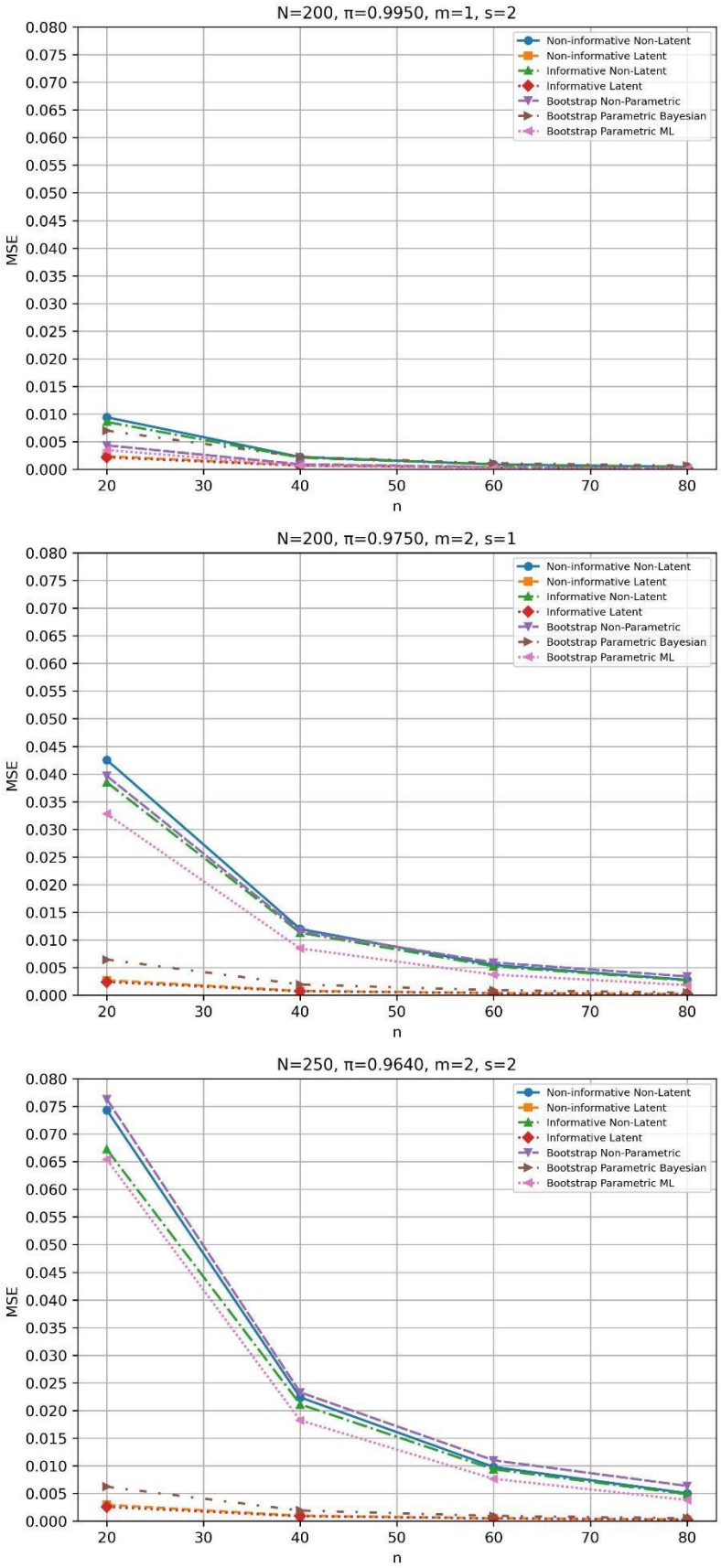


Figure 1. *MSE comparison for the single ray technology under different (N, π, m, s, n) configurations*

Figure 1 presents a comparison of the MSE values of various estimators under different parameter configurations (*N, π, m, s, n*) for single-ray technology. When comparing noninformative and informative approaches, the latent versions perform better than the non-latent ones. In all scenarios, latent models produce lower MSE values. Furthermore, the MSE values of the latent versions are quite close to each other across different combinations of mmm and sss, indicating that the latent approach is more stable with respect to parameter changes. Across different combinations of (*N, π, m, s, n*), parametric bootstrapping based on ML improves estimator performance in terms of MSE compared to nonparametric bootstrapping. When the bootstrap methods are evaluated internally, the results vary depending on the values of mmm and sss. The Bootstrap Parametric ML method yields slightly lower MSE values in the case *m=1, s=2*. In contrast, the Parametric Bootstrap Bayesian method produces lower MSE values when *m=2*. In summary, the latent version of the informative Bayesian estimator performs best in terms of MSE for the cases *m=2, s=1* and *m=2, s=2*. For the case *m=1, s=2*, parametric bootstrapping based on ML provides the best MSE performance.

The simulation results for the multiple-ray technology are presented in Tables 3 and 4, which report the parameter estimates and the corresponding MSE values, respectively.

Table 3. *Estimator values for the multiple rays technology*

					Non-Informative Prior		Informative Prior		Bootstrapping		
N	*π*	*m*	*s*	*n*	Non-Latent	Latent	Non-Latent	Latent	Non-Parametric	Parametric Bayesian	Parametric ML
300	0.9667	1	2	20	0.649864	0.925226	0.665087	0.928437	0.640654	0.893577	0.664785
				40	0.793452	0.946772	0.798256	0.947998	0.790705	0.927988	0.808215
				60	0.852694	0.954667	0.855032	0.955381	0.848354	0.941238	0.864531
				80	0.886341	0.959551	0.887711	0.960036	0.880385	0.949204	0.896055
300	0.9200	2	1	20	0.464273	0.874412	0.487565	0.879859	0.430172	0.851676	0.460622
				40	0.625643	0.894345	0.634349	0.896801	0.601711	0.880613	0.631971
				60	0.709645	0.904784	0.714254	0.906296	0.683793	0.894833	0.716622
				80	0.761110	0.911313	0.763988	0.912381	0.731761	0.903429	0.767637
600	0.8933	2	2	20	0.303955	0.856648	0.334217	0.862877	0.261133	0.837084	0.284405
				40	0.455071	0.877586	0.467744	0.880433	0.428851	0.865495	0.452982
				60	0.546161	0.885289	0.553365	0.887110	0.520925	0.876560	0.547774
				80	0.610488	0.889427	0.615181	0.890759	0.583401	0.882652	0.613257

Table 4. *MSE values for the multiple rays technology*

					Non-Informative Prior		Informative Prior		Bootstrapping		
N	*π*	*m*	*s*	*n*	Non-Latent	Latent	Non-Latent	Latent	Non-Parametric	Parametric Bayesian	Parametric ML
300	0.9667	1	2	20	0.107031	0.002885	0.097050	0.002528	0.114820	0.006140	0.099234
				40	0.032336	0.000943	0.030588	0.000870	0.034100	0.001937	0.027686
				60	0.014086	0.000511	0.013524	0.000482	0.015581	0.000964	0.011607
				80	0.007002	0.000305	0.006771	0.000292	0.008355	0.000530	0.005564
300	0.9200	2	1	20	0.205347	0.004617	0.184557	0.003972	0.238510	0.005692	0.210025
				40	0.083814	0.001738	0.078803	0.001599	0.098945	0.002091	0.080560
				60	0.042007	0.000948	0.040123	0.000905	0.053879	0.001038	0.039388

				80	0.023401	0.000650	0.022523	0.000639	0.033808	0.000633	0.021567
600	0.8933	2	2	20	0.354774	0.005291	0.319388	0.004537	0.408073	0.005881	0.379739
				40	0.197252	0.002237	0.186066	0.002064	0.221704	0.002412	0.199616
				60	0.124645	0.001359	0.119566	0.001293	0.143556	0.001422	0.123825
				80	0.082912	0.001008	0.080210	0.000976	0.099713	0.001017	0.081503

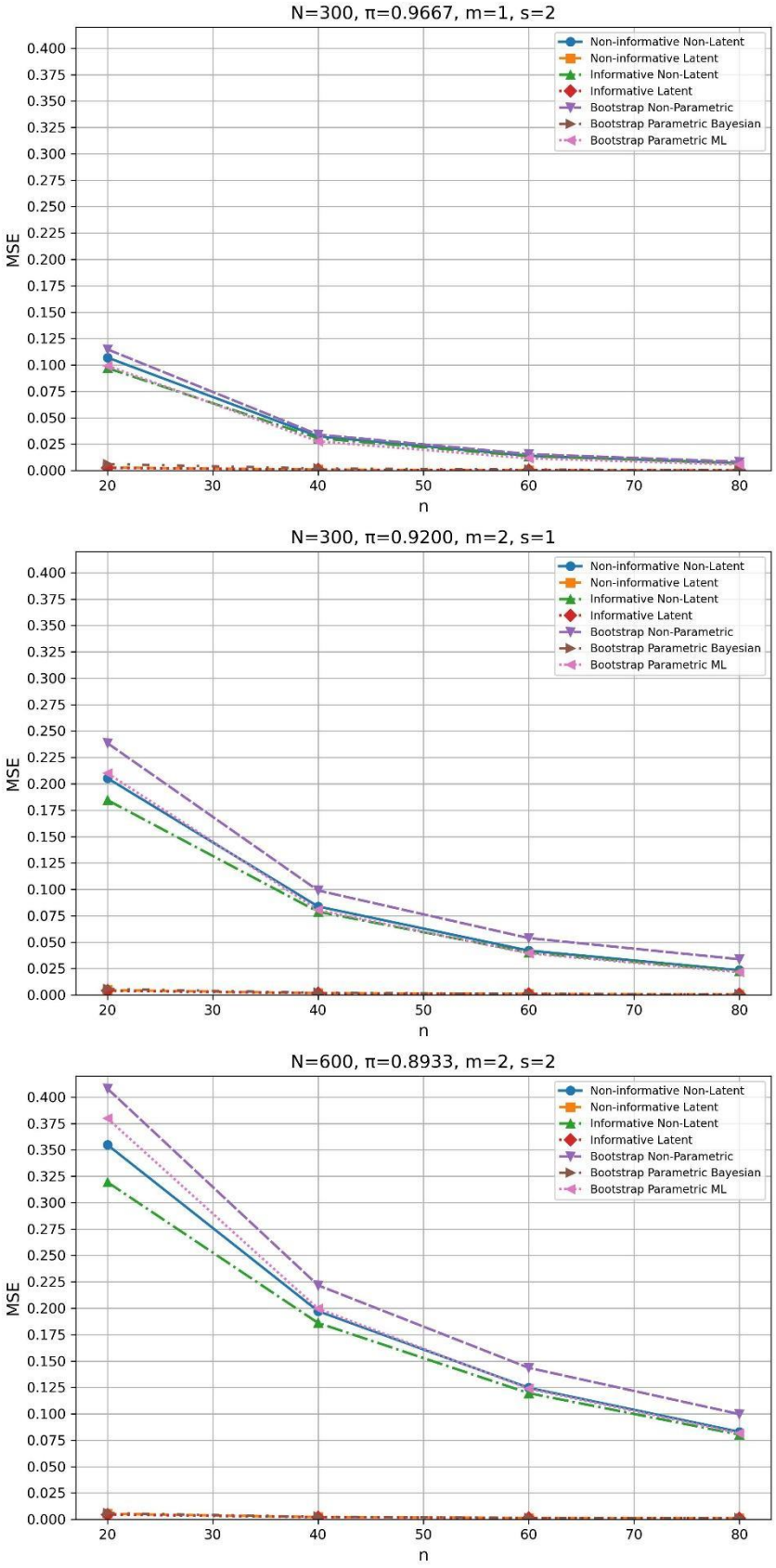


Figure 2. *MSE comparison for the multiple rays technology under different (N, π, m, s, n) configurations*

Figure 2 presents a comparison of the MSE values of various estimators under different parameter configurations (N, π, m, s, n) for multiple-ray technology. Simulation results, based on all estimators considered in previous studies on the incidence of inefficiency, confirm that the parametric bootstrap method provides more accurate estimators than the nonparametric bootstrap method for estimating the incidence of inefficiency. Among the bootstrap methods, the parametric Bayesian approach shows the best performance in terms of MSE for all (N, π, m, s, n) combinations. Furthermore, the Bayesian estimator developed under an informative prior and the latent variable assumption tends to yield superior results. The superiority of the Bayesian estimator is more evident in multiple-ray technology, which identifies more efficient units (and fewer inefficient ones) compared to single-ray technology. In summary, the latent version of the informative Bayesian estimator achieves the lowest MSE across all (N, π, m, s, n) configurations.

4. REAL-WORLD DATA APPLICATION

In this section, the performance comparison of Bayesian and bootstrap-based estimators is applied to real-world data. The IOI estimation application is conducted using data from 27 national and international airports operating in Türkiye between 2009 and 2013. In a panel data structure collected over a 5-year period from 27 airports, there are 135 DMUs (DHMI 2022). To assess the performance of the airports in the analysis, three inputs: plane traffic (the total number of takeoffs and landings), passenger traffic (the total number of arriving and departing people), and cargo traffic (the total transported incoming and outgoing cargo quantity), one output as income (million Turkish Liras) are used. This study is applied separately for single-ray technology and multiple-ray technology cases, and the results are reported. In the real-world data application, n DMUs are randomly sampled with replacement using the resampling method. Based on this sample, Bayesian estimators (both latent and non-latent versions) using non-informative prior and informative prior frameworks as described in section 2 are obtained from [1,3,5]. Additionally, non-parametric bootstrap estimator in [5], as well as Bayesian and ML-based parametric bootstrap estimators are derived as proposed in section 3 in this study. Since the calculations are performed on observations obtained from a single real-world dataset, the estimator values are provided, but MSEs cannot be computed. In single ray technology, the number of inefficient units is 132, thus, the population IOI for 135 DMUs and three inputs, one output case is calculated as 0.9778. Then, 40 DMUs are randomly sampled with replacement using the resampling method. Bayesian and bootstrap estimators are calculated. According to the results, Bayesian estimators based on non-informative prior are obtained as 0.904763 and 0.952024 for non-latent and latent versions, respectively. Informative prior-based estimators are 0.906978 (non-latent) and 0.953123 (latent version). Considering the bootstrap-based estimator values, the nonparametric bootstrap estimate is calculated as 0.923077, the Bayesian bootstrap estimator as 0.933524, and the ML based bootstrap estimator as 0.924850. In this example, where all estimators are compared, it is observed that the Bayesian estimator constructed under the informative prior and latent variable assumption provides the best performance in the multidimensional case (three inputs, one output), consistent with the findings of the simulation studies. For multiple-ray technology, the number of inefficient units is 127, then the population IOI for 135 DMUs and three inputs, one output case is calculated as 0.9407. 40 DMUs are randomly sampled with replacement. Bayesian estimators based on non-informative prior are obtained as 0.809524 and 0.904763 for non-latent and latent versions, respectively. Informative prior-based estimators are 0.813953 (non-latent) and 0.906978 (latent version). Considering the bootstrap-based estimator values, the nonparametric bootstrap estimate is calculated as 0.821170, the Bayesian bootstrap estimator as 0.892931, and the ML-based bootstrap estimator as 0.824350. These results support both the findings of the application study for single-ray technology and the simulation results.

5. RESULTS

This study stands as the most extensive comparison study in the literature on the estimation of the population IOI. Simulation and application studies incorporating all estimators from previous research on the inefficiency incidence confirm that the ML-based parametric bootstrap method yields better estimators than the non-parametric bootstrap approach for estimating the inefficiency incidence. When all estimators are evaluated, the latent variable and informative prior based Bayesian estimator is found to be superior in

many trials. While the present study offers a comprehensive comparative analysis, certain limitations can be identified and directions for future research may be outlined. The scope of the study is limited to estimators developed for finite population sizes. Moreover, the Bayesian estimators included in the comparison, which have been proposed in earlier studies, rely on specific prior distributions. Future research may focus on these points to extend the subject.

DECLARATION OF ARTIFICIAL INTELLIGENCE (AI) USE

Generative AI agents or large language models (LLMs) were used only to improve clarity and quality of the English language presentation of the manuscript.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Friesner, D., Mittelhammer, R., and Rosenman, R., "Inferring the incidence of industry inefficiency from DEA estimates", European Journal of Operational Research, 224(2): 414-424, (2013). DOI: 10.1016/j.ejor.2012.08.003

[2] Ünsal, M.G., Friesner, D., and Rosenman, R., "New posterior distributions for the incidence of inefficiency in DEA scores", Communications in Statistics – Theory and Methods, 50(8): 1774-1780, (2021). DOI: 10.1080/03610926.2019.1653920

[3] Ünsal, M.G., Friesner, D., and Rosenman, R., "The curse of dimensionality (COD), misclassified DMUs, and Bayesian DEA", Communications in Statistics – Simulation and Computation, 51(8): 4186-4203, (2022). DOI: 10.1080/03610918.2020.1740260

[4] Ünsal, M.G., "New Bayesian estimators for the incidence of inefficiency in the cases without a rule of thumb", Journal of Statistical Computation and Simulation, 92(3): 437-450, (2022). DOI: 10.1080/00949655.2021.1952204

[5] Ünsal, M.G., Friesner, D., Rosenman, R.E., and Örkcü, M., "Estimation of the incidence of inefficiency using bootstrap and Bayesian estimators", Communications in Statistics – Simulation and Computation, 1-15 (accepted manuscript), (2024). DOI: 10.1080/03610918.2024.2423221

[6] Kneip, A., Simar, L., and Wilson, P.W., "Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models", Econometric Theory, 24: 1663-1697, (2008).

[7] Çolak, A.B., Sindhu, T.N., Lone, S.A., Shafiq, A., and Abushal, T.A., "Reliability study of generalized Rayleigh distribution based on inverse power law using artificial neural network with Bayesian regularization", Tribology International, 185: 108544, (2023). DOI: 10.1016/j.triboint.2023.108544

[8] Sindhu, T.N., and Hussain, Z., "Objective Bayesian analysis for the power function II distribution under doubly type II censored data", Thai Statistician, 20(3): 710-731, (2022).

[9] Sindhu, T.N., Çolak, A.B., Lone, S.A., Shafiq, A., and Abushal, T.A., "A decreasing failure rate model with a novel approach to enhance the artificial neural network's structure for engineering and disease data analysis", Tribology International, 192: 109231, (2024). DOI: 10.1016/j.triboint.2023.109231

[10] Heng, Q., and Lange, K., "Bootstrap estimation of the proportion of outliers in robust regression", Statistics and Computing, 35(1): 3, (2025). DOI: 10.1007/s11222-024-10526-1

[11] Wu, J.C., Martin, A.F., and Kacker, R.N., "Validation of nonparametric two-sample bootstrap in ROC analysis on large datasets", Communications in Statistics – Simulation and Computation, 45(5): 1689-1703, (2016). DOI**:** 10.1080/03610918.2015.1065327

[12] Zhuang, L., Xu, A., Fang, G., and Tang, Y., "Multivariate reparameterized inverse Gaussian processes with common effects for degradation-based reliability prediction", Journal of Quality Technology, 57(1): 51-67, (2025). DOI: 10.1080/00224065.2024.2402850

[13] Davison, A.C., and Hinkley, D.V., Bootstrap Methods and Their Application, Cambridge Univ. Press, No.1, (1997).

[14] Hall, P., The Bootstrap and Edgeworth Expansion, Springer Sci. Bus. Media, (2013).

[15] Efron, B., and Tibshirani, R.J., An Introduction to the Bootstrap, Chapman and Hall, (1993).

[16] Kreiss, J.P., and Paparoditis, E., "Bootstrap methods for dependent data: A review", Journal of the Korean Statistical Society, 40(4): 357-378, (2011). DOI: 10.1016/j.jkss.2011.08.009

[17] Lemonte, A.J., Simas, A.B., and Cribari-Neto, F., "Bootstrap-based improved estimators for the two-parameter Birnbaum–Saunders distribution", Journal of Statistical Computation and Simulation, 78(1): 37-49, (2008). DOI: 10.1080/10629360600903882

[18] Moretti, A., Shlomo, N., and Sakshaug, J.W., "Parametric bootstrap mean squared error of a small area multivariate EBLUP", Communications in Statistics – Simulation and Computation, 49(6): 1474-1486, (2020). DOI: 10.1080/03610918.2018.1498889

[19] Van Den Noortgate, W., and Onghena, P., "Parametric and nonparametric bootstrap methods for meta-analysis", Behavior Research Methods, 37: 11-22, (2005). DOI: 10.3758/BF03206394

[20] Van Pul, M., "Simulations on the Jelinski-Moranda model of software reliability, application of some parametric bootstrap methods", Statistics and Computing, 2: 121-136, (1992). DOI: 10.1007/BF01891204

[21] Politis, D.N., and Romano, J.P., "The stationary bootstrap", Journal of the American Statistical Association, 89(428): 1303-1313, (1994). DOI: 10.1080/01621459.1994.10476870

[22] Giraleas, D., Emrouznejad, A., and Thanassoulis, E., "Productivity change using growth accounting and frontier-based approaches—Evidence from a Monte Carlo analysis", European Journal of Operational Research, 222(3): 673-683, (2012). DOI: 10.1016/j.ejor.2012.05.015

[23] Banker, R.D., Charnes, A., Cooper, W.W., Swarts, J., and Thomas, D., "An introduction to data envelopment analysis with some of its models and their uses", Research in Governmental and Nonprofit Accounting, 5(1): 125-163, (1989).

[24] Simar, L., and Wilson, P.W., "Statistical inference in nonparametric frontier models: The state of the art", Journal of Productivity Analysis, 13: 49-78, (2000). DOI: 10.1023/A:1007864806704

[25] Agasisti, T., "The efficiency of Italian secondary schools and the potential role of competition: A data envelopment analysis using OECD-PISA2006 data", Education Economics, 21(5): 520-544, (2013). DOI: 10.1080/09645292.2010.511840