Journal of Statistical Modelling: Theory and Applications
Vol. 7, No. 1, 2026, pp. 129-158
Yazd University Press 2026

JSMTA

Research Paper

Misalignment between statistical fit and operational performance of Bayesian estimators in automated teller machine queues with fuzzy-censored data

Kazem Fayyaz Heidari, Iman Makhdoom*
Department of Statistics, Payame Noor University, Tehran, Iran

Received: August 10, 2026/ Revised: September 12, 2026/ Accepted: September 25, 2026

Abstract: This paper presents a simulation study inspired by automated teller machine systems to investigate the misalignment between statistical fit measured by the deviance information criterion and operational performance measured by system cost using fuzzy-censored data. The Bayesian, E-Bayesian and hierarchical Bayesian estimators are compared with the maximum likelihood estimator. A misalignment index is defined based on the difference between system cost and the normalized ranks of the deviance information criterion or the Akaike information criterion for the maximum likelihood estimator. Simulation results across various scenarios involving combinations of (ρ, m, K) indicate that misalignment occurs over a wide range of traffic loads ($\rho = 0.15$ to $\rho = 0.90$) with $m = 3$. Moreover, increasing m to 4 does not completely eliminate misalignment, and it is observed in many scenarios with $m = 4$ as well. The hierarchical Bayesian estimator exhibited the most consistent alignment between statistical fit and operational cost within the parameter grid studied. Sensitivity analysis with respect to the loss function parameter shows that the misalignment pattern is stable for $p < 0$ (underestimation penalized more heavily) but changes for $p > 0$ (overestimation penalized more heavily). Credible intervals are provided for the Bayesian estimators. Based on a comprehensive sensitivity map, practical recommendations for selecting an appropriate estimator in automated teller machine-like systems are provided.

Keywords: Bayesian estimation; Credible interval; Fuzzy-censored data; Queueing model; Statistical fit misalignment.
Mathematics Subject Classification (2010): 62A86, 62N86, 60K30.

*Corresponding author: makhdoom@pnu.ac.ir

1 Introduction

In classical statistical theory, the concepts of Fisher information and the Cramer-Rao bound are fundamental tools for evaluating the precision of estimators (Fisher, 1925; Rao, 1945; Cramer, 1946). These theoretical frameworks assume that observations are precise and fully observed. However, in many real-world applications, data are not available in a precise and deterministic manner but are accompanied by a degree of fuzziness due to measurement errors, subjective judgments, or inherent ambiguity in the system. On the other hand, practical limitations such as cost, time, or operational constraints often cause data to be observed in a censored manner, where only partial information is available.

Queueing systems are ubiquitous in service industries, telecommunications, and manufacturing. Among them, the $M/M/m/K$ queueing model with finite capacity is one of the most widely used frameworks for performance analysis and capacity planning. The traffic intensity parameter plays a critical role in determining system performance, as it directly affects key operational metrics such as queue length, effective arrival rate, and blocking probability. Accurate estimation of the traffic intensity parameter is therefore essential for optimal system design and resource allocation.

Several studies have addressed the optimization of queueing systems through cost functions. Makhdoom and Yaghoobzadeh Shahrastani (2023) investigated improving the quality of $M/M/m/K$ queueing systems using system cost function optimization, demonstrating the importance of balancing operational costs with service quality. Yaghoobzadeh Shahrastani et al. (2025) optimizes the quality of the $M/E_r/1$ queueing model family by jointly considering cost, system stationarity probability, and customer satisfaction over a finite time horizon. Shahrastani and Makhdoom (2026) developed a new E2-Bayesian estimator for the efficiency parameter of an infinite-capacity multi-server queueing system. Yaghoobzadeh (2026) proposed a new estimation for the traffic intensity parameter of a single-server queueing system with finite capacity. Bura and Vaish (2026) performed Bayesian analysis of a state-dependent service Markovian queueing model under asymmetric loss functions. Singh Bura and Vaish (2025) conducted Bayesian inference of traffic intensity for an M/M/1 queueing model with state-dependent service under asymmetric loss functions. Additionally, Makhdoom (2023) focused on optimum statistical estimation of traffic intensity parameters in the $M/M/1/K$ queueing model under both fuzzy and non-fuzzy criteria, highlighting the need for robust estimation methods that account for data ambiguity. Previous research in fuzzy statistics has focused on parameter estimation and hypothesis testing using fuzzy data Pak et al. (2013, 2014). Several studies have addressed Bayesian estimation (BE) in queueing systems under various data conditions. Makhdoom and Yaghoobzadeh Shahrastani (2024) performed BE of traffic intensity in the $M/M/1$ model with fuzzy and Type-II censored data. Fayyaz Heidari et al. (2024) presented E-Bayesian (EB) and H-Bayesian (HB) estimations for the $M/M/m/K$ model with a fuzzy index. Heidari et al. (2026) developed fuzzy BE, EB, and HB estimations of the stress-strengh reliability for the Weibull distribution under Type-II censored data. Zarei et al. (2025) performed fuzzy EB and HB estimation of scalar parameter in Gompertz distribution under asymmetric loss functions. Despite these advances, a significant gap remains between statistical fit measured by information criteria such as deviance information criterion (DIC) and operational performance measured by sys-

tem cost in queueing systems with fuzzy-censored data. Most existing studies evaluate estimators solely based on statistical criteria such as bias, mean squared error, or information criteria, without considering whether a statistically well-fitted estimator translates into better operational performance. This gap is particularly critical in automated teller machine (ATM)-like systems, where operational efficiency directly impacts customer satisfaction and revenue. Managers of such systems must balance multiple objectives: minimizing customer waiting time, reducing blocking probability, and controlling system costs. The present study addresses this gap through a controlled Monte Carlo investigation based on the steady-state $M/M/m/K$ model with fuzzy-censored data, rather than a full operational simulation of a real ATM system. However, an estimator that provides superior statistical fit may not necessarily lead to lower system cost, leading to a misalignment between statistical fit and practical performance.

The main objective of this paper is to investigate this gap and provide a comprehensive framework for identifying and quantifying the misalignment between statistical fit and operational performance over a wide range of operational scenarios. We propose a novel inconsistency index (INC) - a rank-based measure of misalignment - based on the difference between system cost and normalized ranks of DIC or Akaike information criterion (AIC) for maximum likelihood estimator (MLE). Through an extensive simulation study inspired by ATM systems with fuzzy-censored data, we compare four estimators (MLE, BE, EB, HB) and evaluate their performance across various combinations of (ρ, m, K). Sensitivity analyses examine the robustness of the results with respect to hyperparameters, cost coefficients, fuzzification level, and loss function. A comprehensive sensitivity map is provided to guide estimator selection in ATM-like systems.

The structure of the paper is as follows. Section 2 defines the basic concepts of $M/M/m/K$ queueing, fuzzy data, and fuzzy censoring. Section 3 introduces the fuzzy-censored likelihood function. Section 4 presents the estimators and evaluation criteria, including the proposed INC index. Section 5 is devoted to the simulation study and sensitivity analyses. Finally, Section 6 provides conclusions and practical recommendations for ATM-like systems.

2 Basic concepts and modeling

2.1 $M/M/m/K$ queueing model

The queueing model under consideration is $M/M/m/K$, in which the customer arrival process follows a Poisson distribution with rate λ and the service time follows an exponential distribution with rate μ. The number of servers is m and the total system capacity is K, where it is assumed that $m \leq K$. The ratio $r = \lambda/\mu$ and the traffic intensity $\rho = r/m = \lambda/(m\mu)$ are the target parameter of estimation in this paper. In the finite-capacity model, the stability condition $\rho < 1$ is not necessary, and the system is stable for any $\rho > 0$ (Allen, 2014). In the steady state, the probability distribution of the number of customers in the system, denoted by P_n, is obtained as follows (Allen, 2014):

$$P_n = \begin{cases} \dfrac{r^n}{n!} P_0, & 0 \le n \le m, \\ \dfrac{r^m}{m!} \rho^{n-m} P_0, & m < n \le K, \end{cases} \tag{1}$$

where $P_0 = \left(\sum_{i=0}^{m} \frac{r^i}{i!} + \frac{r^m}{m!} \sum_{j=1}^{K-m} \rho^j \right)^{-1}$, and P_0 is computed from the condition $\sum_{n=0}^{K} P_n = 1$.

In infinite-capacity $M/M/m$ queues, the stability condition is $\rho < 1$, ensuring that the arrival rate does not exceed the service capacity. However, in the finite-capacity model $M/M/m/K$, the system is stable for any $\rho > 0$. This is because the finite buffer K limits the number of customers in the system, preventing unbounded growth of the queue length even when $\rho \ge 1$. When $\rho \ge 1$, the system operates in an overloaded regime, and the blocking probability P_K becomes significant. The finite capacity effectively acts as a stabilizing mechanism, albeit at the cost of increased customer blocking and potential revenue loss. This characteristic makes finite-capacity queues particularly relevant for ATM systems, where physical space and service capacity are inherently limited.

The average number of customers in the queue (L_q), the effective arrival rate ($\overline{\lambda}$), and the average number of customers in the system (L) are computed as follows:

$$L_q = \sum_{n=m}^{K} (n-m) P_n, \quad \overline{\lambda} = \lambda \left(1 - P_K\right), \quad L = L_q + \frac{\overline{\lambda}}{\mu}.$$

2.2 Fuzzy data

Continuous data (inter-arrival and service times) are modeled as triangular fuzzy numbers $\widetilde{T}_j = (x_{j1}, x_{j2}, x_{j3})$ with the following membership function (Dubois and Prade, 1980) instead of precise numbers:

$$\mu_{\widetilde{T}_j}(x) = \begin{cases} \frac{x - x_{j1}}{x_{j2} - x_{j1}}, & x_{j1} \le x \le x_{j2}, \\ \frac{x_{j3} - x}{x_{j3} - x_{j2}}, & x_{j2} \le x \le x_{j3}, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

It should be noted at the outset that the continuous fuzzy observations defined in (2) are introduced here for completeness of the fuzzy-data framework, but they are not used in the estimation procedure of this paper. The reason is that, in the $M/M/m/K$ model, the traffic intensity parameter ρ is identified through the steady-state distribution of the number of customers in the system, $P_n(\rho)$, and not through the individual inter-arrival or service times. Since the data are simulated directly from $P_n(\rho)$ (see Section 5.1.1), the estimation relies exclusively on the discrete fuzzy observations $\widetilde{N}_j$ defined in (3). The continuous fuzzy framework in (2) would become relevant in extensions where inter-arrival or service times are themselves observed with fuzziness, such as in $M/G/m/K$ or $G/M/m/K$ models; this is left as a direction for future research.

For discrete data (number of customers in the system), the fuzzy membership function $\mu_{\widetilde{N}_j}(n)$ on the set $\{0,\ 1, \ldots, K\}$ is defined as follows:

$$\mu_{\widetilde{N}_j}(n) = \begin{cases} \frac{n-(N_j-d_j)}{d_j}, & N_j - d_j \leq n \leq N_j, \\ \frac{(N_j+d_j)-n}{d_j}, & N_j \leq n \leq N_j + d_j, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

where N is an integer, $\delta = 0.05$ specifies the fuzzification level, and $d_j = \lceil \delta N_j \rceil$ is the integer-valued fuzzification radius. For $N_j = 0$, we set $d_j = 1$. For $n < 0$ or $n > K$, the membership value is considered zero.

The discrete fuzzy observations $\widetilde{N}_j$ defined in (3) are the only fuzzy data used in the estimation procedure of this paper, as they directly relate to the steady-state distribution $P_n(\rho)$ from which the data are simulated.

2.3 Fuzzy censoring

In classical survival analysis and reliability studies, censoring occurs when the exact observation time is not fully known. Two common types of censoring are Type-I censoring, where the experiment is terminated after a fixed time, and Type-II censoring, where the experiment continues until a predetermined number of failures is observed. However, in many practical queueing scenarios, the determination of the exact censoring point may be subject to ambiguity or imprecision.

Definition 2.1. *Fuzzy Censoring: Fuzzy censoring extends the classical censoring concept by allowing the censoring time to be a fuzzy number rather than a crisp value. In this approach, the end point of the experiment is defined as a fuzzy integer $\widetilde{s}$ with membership function $\mu_{\widetilde{s}}(t)$, representing the degree of plausibility that the experiment ends at observation count t.*

The motivation for using fuzzy censoring in this study is twofold:
1. Practical ambiguity: In ATM-like systems, the observation period may not have a sharply defined end point. For example, data collection may depend on operational constraints, maintenance schedules, or varying business hours, leading to uncertainty about the exact number of complete observations.
2. Robustness: By allowing a range of possible censoring points with associated weights, the fuzzy censoring approach provides more robust inference compared to a single arbitrarily chosen censoring point.

In this paper, the fuzzy censoring distribution is defined over the set $\{33, 34, 35, 36, 37\}$ with weights $\omega = (0.05, 0.20, 0.50, 0.20, 0.05)$. This choice reflects that the most likely number of complete observations is $t = 35$, with decreasing plausibility for values farther from this central value. The normalized weights are computed as in (5).

3 Fuzzy-censored likelihood function

In this study, the data are simulated from independent replications of the system in steady state, not from a time series. This means that each observation $\widetilde{N}_j$ is obtained

from an independent and separate run of the system with the same parameter ρ. Therefore, the independence assumption in relation (4) holds. If the data were in the form of a time series, conditional likelihood or Markov chain-based likelihood would have to be used.

In the standard approach for fuzzy data (Viertl, 2006; Pak et al., 2013), the likelihood function for fuzzy data is defined using Zadeh's extension principle (Zadeh, 1968) as follows:

$$L\left(\rho; \widetilde{N}_1, \ldots, \widetilde{N}_s\right) = \prod_{j=1}^{s} \left(\sum_{n=0}^{K} P_n(\rho) \mu_{\widetilde{N}_j}(n) \right). \tag{4}$$

Equation (4) is the standard fuzzy likelihood under Zadeh's extension principle, as formalized by Viertl (2006) and applied to reliability and queueing problems by Pak et al. (2013, 2014).

In the fuzzy censoring approach, the end point of the experiment is defined as a fuzzy integer $\widetilde{s}$ with membership function $\mu_{\widetilde{s}}$. The normalized weights are defined as follows:

$$\omega_t = \frac{\mu_{\widetilde{s}}(t)}{\sum_{u=0}^{T} \mu_{\widetilde{s}}(u)}, \qquad t = 0, 1, \ldots, T. \tag{5}$$

The fuzzy-censored likelihood function is defined as a weighted average of the fuzzy likelihoods for the number of complete observations t:

$$L_{FC}(\rho) = \sum_{t=0}^{T} \omega_t L\left(\rho; \widetilde{N}_1, \ldots, \widetilde{N}_t\right). \tag{6}$$

Equation (6) defines $L_{FC}(\rho)$ as a fuzzy extension of the likelihood under fuzzy censoring, following the general framework of Viertl (2006) for combining fuzzy observations with fuzzy experimental conditions.

In the simulation study, the maximum number of observations ($t_{\max} = 37$) is generated for each replication, and the fuzzy-censored likelihood in (6) is computed over nested subsets of these observations, corresponding to each possible censoring point $t \in \{33, 34, 35, 36, 37\}$. This ensures that all terms in (6) are well-defined for every replication. The computational cost of relation (6) is of order $O(T \ . \ K)$, which is acceptable for $n = 90$ and $K = 10$. The computational cost of relation (6) is of order $O(T \ . \ K)$, which is acceptable for $n = 90$ and $K = 10$.

Equation (6) defines the fuzzy-censored likelihood as a convex combination of likelihood functions for different complete observation counts t, with weights ω_t. This formulation has the following interpretation: instead of choosing a single censoring point t, we consider all possible censoring points simultaneously, weighted by their plausibility. The fuzzy-censored likelihood $L_{FC}(\rho)$ can thus be viewed as the expected likelihood under the fuzzy censoring distribution:

$$L_{FC}(\rho) = \mathbb{E}_{\widetilde{s}}\left[L\left(\rho; \widetilde{N}_1, \ldots, \widetilde{N}_{\widetilde{s}}\right) \right],$$

where the expectation is taken over the fuzzy censoring weights ω_t. This approach ensures that the inference does not rely on a single arbitrary censoring point, providing robustness against uncertainty in the observation endpoint.

3.1 Interpretation of the fuzzy-censored likelihood

The quantity $L_{FC}(\rho)$ defined in (6) is a fuzzy likelihood in the sense of Viertl (2006), constructed by combining two distinct fuzzy mechanisms. The first mechanism is the fuzzy observation of the system state, expressed in (4) through Zadeh's extension principle Zadeh (1968), which replaces the classical likelihood contribution $P_n(\rho)$ with its fuzzy-weighted average $\sum_{n=0}^{K} P_n(\rho)\mu_{\widetilde{N}_j}(n)$. The second mechanism is fuzzy censoring, expressed in (6) through the expectation over the fuzzy censoring distribution with weights ω_t. Combining these two mechanisms yields a single function $L_{FC}(\rho)$ that is not a classical likelihood in the strict sense, because it does not arise from a probability density over a σ-algebra of crisp events. Rather, it is a fuzzy extension of the likelihood that preserves the essential features required for inference: it is non-negative, continuous in ρ, has a well-defined maximum, and converges to the classical likelihood as the fuzzification level $\delta \to 0$ (Viertl, 2006; Pak et al., 2013).

For the purposes of point estimation and model comparison, $L_{FC}(\rho)$ can be treated as a pseudo-likelihood: it plays the role of a likelihood in the construction of the MLE in (11) and the Bayesian estimators in (7)-(10), and it inherits the asymptotic properties of the classical likelihood under regularity conditions as $\delta \to 0$. The resulting posterior distributions are fuzzy posteriors in the sense of Viertl (2006), and the corresponding credible intervals retain their interpretation as fuzzy credible intervals. The MLE in (11) is therefore a fuzzy MLE, and the Bayesian estimators in (7)-(10) are obtained by replacing the classical likelihood with its fuzzy counterpart in Bayes' theorem.

The use of AIC and DIC with $L_{FC}(\rho)$ requires a brief clarification. Because $L_{FC}(\rho)$ is a fuzzy likelihood rather than a classical one, the deviance $D(\rho) = -2\log L_{FC}(\rho)$ should be interpreted as a fuzzy deviance, and the resulting AIC and DIC values as fuzzy information criteria. Because $L_{FC}(\rho)$ preserves the essential likelihood properties required for model ranking (non-negativity, continuity, well-defined maximum, and convergence to the classical likelihood as $\delta \to 0$), the numerical values of AIC and DIC computed from $L_{FC}(\rho)$ can be compared across estimators in the same spirit as their classical counterparts.

This interpretation is consistent with the broader literature on fuzzy information criteria (Viertl, 2006; Pak et al., 2013, 2014) and with the use of INC as a rank-based measure in Section 4.2, since only the ordering of estimators under AIC/DIC affects INC.

4 Estimators and evaluation criteria

4.1 Estimators

The estimators defined in this paper are derived under the general entropy (GE) loss function, introduced by Calabria and Pulcini (1996). For a parameter ρ and its estimator $\hat{\rho}$, the GE loss function is defined as

$$L_{GE}(\rho, \hat{\rho}) = \left(\frac{\hat{\rho}}{\rho}\right)^{p} - p\log\left(\frac{\hat{\rho}}{\rho}\right) - 1, \quad p \neq 0.$$

The GE loss function is asymmetric for all $p \neq 0$. The parameter p controls both the direction and the degree of asymmetry. For $p < 0$, underestimation ($\hat{\rho} < \rho$) is

penalized more heavily than overestimation; for $p > 0$, overestimation is penalized more heavily than underestimation. The case $p = 0$ is excluded because the GE loss reduces to zero in that limit. In particular, the GE loss is not symmetric under the transformation $u \leftrightarrow 1/u$ for any $p \neq 0$, and it is not equivalent to the squared error loss function in the logarithmic scale, $(\log(\hat{\rho}/\rho))^2$, which is a distinct loss function.

When $p = -1$, the loss function takes the form:

$$L_{GE}(\rho, \hat{\rho}) = (\frac{\hat{\rho}}{\rho})^{-1} + \log(\frac{\hat{\rho}}{\rho}) - 1 = (\frac{\rho}{\hat{\rho}}) + \log(\frac{\hat{\rho}}{\rho}) - 1.$$

This form penalizes relative errors on a multiplicative scale but remains asymmetric: for example, an underestimation by a factor of 2 ($\hat{\rho}/\rho = 0.5$) receives a penalty of approximately 0.307, whereas an overestimation by the same factor of 2 ($\hat{\rho}/\rho = 2$) receives a penalty of approximately 0.193.

The choice $p = -1$ is particularly suitable for queueing parameters such as traffic intensity ρ, for the following reasons:
1. Scale invariance: The relative error is more meaningful than absolute error when ρ ranges over $[0, 1]$.
2. Moderate asymmetry: For $p = -1$, the asymmetry is mild compared with more extreme values such as $p = -3$, and the loss places a slightly heavier penalty on underestimation of ρ. In queueing contexts, underestimating traffic intensity may lead to under-provisioning of service capacity and to higher actual costs, so a mild preference for penalizing underestimation is operationally reasonable.
3. Analytical tractability: For $p = -1$, the Bayesian estimator under GE loss reduces to a closed-form expression given by (7), which simplifies both the theoretical derivation and the numerical implementation. This choice is also consistent with the standard practice in the EB literature (Han, 1997, 2009).

Based on standard definitions in the literature (Han, 1997, 2009; Gelman et al., 2013; Fayyaz Heidari et al., 2024), the following estimators are defined. In all formulas, $p = -1$, which yields the posterior mean as the Bayesian estimator (the same point estimator that would be obtained under squared error loss). It should be emphasized that this choice does not imply that the GE loss with $p = -1$ is itself equivalent to squared error loss; rather, the two loss functions lead to the same Bayesian point estimator. $L_{FC}(\rho)$ is the fuzzy-censored likelihood function from (6).

Definition 4.1. *Bayesian Estimator (BE): The Bayesian estimator is defined using a uniform prior distribution* $\pi(\rho) \sim Beta(1,\ 1)$ *as follows:*

$$\hat{\rho}_{BE} = \left[\frac{\int_0^1 \rho^{-p} L_{FC}(\rho)\pi(\rho)\, d\rho}{\int_0^1 L_{FC}(\rho)\pi(\rho)\, d\rho} \right]^{-1/p}. \tag{7}$$

Definition 4.2. *E-Bayesian Estimator (EB): First, the Bayesian estimator with prior parameters* a *and* b *is defined:*

$$\hat{\rho}_{BE}(a, b) = \left[\frac{\int_0^1 \rho^{-p} L_{FC}(\rho)\pi(\rho \mid a, b)\, d\rho}{\int_0^1 L_{FC}(\rho)\pi(\rho \mid a, b)\, d\rho} \right]^{-1/p}, \tag{8}$$

where $\pi(\rho|a,b) \sim Beta(a,b)$. *Then, the EB estimator is obtained by averaging* $\hat{\rho}_{BE}(a,b)$*:*

$$\hat{\rho}_{EB} = \int_0^B \int_0^A \hat{\rho}_{BE}(a,b).\frac{1}{A}.\frac{1}{B}dadb, \tag{9}$$

where $A = B = 10$. *Stability analysis with respect to* $A = B$ *is presented in Section 5-3.*

It should be clarified how the 95% HPD credible interval and the DIC are computed for the EB estimator. Because the EB estimator in (9) is defined as an average of Bayesian estimators over the hyper-parameters a and b, it does not by itself define a single posterior distribution. The natural posterior distribution for EB is therefore the mixture posterior defined by:

$$\pi_{EB}(\rho|data) = \int_0^B \int_0^A \pi(\rho|a,b,data).\frac{1}{A}.\frac{1}{B}dadb,$$

where $\pi(\rho|a,b,data)$ is the posterior distribution of ρ under the $Beta(a,\ b)$ prior. This mixture posterior is the natural Bayesian counterpart of the EB point estimator: its mean coincides with $\hat{\rho}_{EB}$ in (9), and its spread reflects the uncertainty induced by averaging over the hyper-parameter region $[0,\ A] \times [0,\ B]$.

The 95% HPD credible interval for EB is computed from this mixture posterior. The DIC for EB is computed as $DIC_{EB} = \overline{D}_{EB} + p_{D,EB}$, where $\overline{D}_{EB}$ is the posterior mean deviance and $p_{D,EB}$ is the effective number of parameters, both evaluated under $\pi_{EB}(\rho|data)$. These quantities are computed by numerical integration over the hyperparameter region using Gauss-Legendre quadrature with 100 points in each dimension, as described in Section 5.1.3. Because the integration is performed over the hyperparameter region $[0,\ A] \times [0,\ B]$ and not at a single point, the resulting interval and DIC fully account for the averaging that defines the EB estimator.

Definition 4.3. *Hierarchical Bayesian estimator (HB):*

$$\hat{\rho}_{HB} = \left[\frac{\int_0^1 \rho^{-p} L_{FC}(\rho)\pi_H(\rho)\,d\rho}{\int_0^1 L_{FC}(\rho)\pi_H(\rho)\,d\rho}\right]^{-1/p}, \tag{10}$$

where

$$\pi_H(\rho) = \int_0^{10} \int_0^{10} \pi(\rho|a,b).\frac{1}{10}.\frac{1}{10}dadb.$$

For the computation of HB, MCMC with 50,000 iterations, 10,000 burn-in periods, and 5 chains is used. The $\hat{R} < 1.05$ *criterion is confirmed for all parameters. 95% Highest Posterior Density (HPD) credible intervals are computed from the MCMC posterior samples.*

The EB and HB estimators are closely related but conceptually distinct:

• The EB estimator averages the Bayesian estimator $\hat{\rho}_{BE}(a,b)$ over the prior hyper-parameters a and b with a uniform weight over the region $[0,\ A] \times [0,\ B]$. This can be interpreted as empirical Bayes at the hyper-parameter level, where instead of estimating a and b from the data, we integrate them out with a flat prior. This approach

reduces the sensitivity to the choice of prior parameters.
• The HB estimator uses a hierarchical prior structure where ρ is assigned a Beta prior with parameters a and b, and a and b themselves are assigned hyper-priors (uniform over $[0,\ 10]\ \times [0,\ 10]$). The prior $\pi_H(\rho)$ in Equation (11) is the marginal prior for ρ obtained by integrating out a and b. The HB estimator is the Bayes estimator under this marginal prior.

The key difference is that EB averages the posterior means obtained from different priors, while HB computes the posterior mean under a single hierarchical prior. Both methods aim to reduce sensitivity to prior specification, but through different mechanisms. The stability analysis in Section 5.3 investigates the sensitivity of both methods to the choice of $A = B$.

Definition 4.4. *Maximum Likelihood Estimator (MLE):*

$$\hat{\rho}_{MLE} = \arg \max_{\rho \in [0,1]} L_{FC}(\rho). \tag{11}$$

4.2 Evaluation criteria

Information Criterion (DIC for Bayesian estimators, AIC for MLE): For evaluating the fit of Bayesian models, the DIC given by (Spiegelhalter et al., 2002) is used. The deviance is defined as $D(\rho) = -2 \log L_{FC}(\rho)$, and DIC is computed as $DIC = \overline{D} + p_D$. For the MLE, AIC is used: $AIC = -2 \log L_{FC}\left(\hat{\rho}_{MLE}\right) + 2k$.

Although AIC is used for the MLE and DIC is used for the Bayesian estimators, both criteria share a common theoretical foundation as estimators of the expected predictive deviance under the Kullback-Leibler divergence (Spiegelhalter et al., 2002; Gelman et al., 2013). Specifically, $AIC = -2 \log L\left(\hat{\rho}_{MLE}\right) + 2k$ and $DIC = \overline{D} + p_D$ both take the form of a deviance term plus a complexity penalty. Under regularity conditions and a sufficiently large number of observations, DIC converges to AIC as the posterior distribution concentrates around the MLE. More importantly, because INC is a rank-based measure (12), only the relative ordering of estimators under each criterion affects INC, not the numerical scale of the criterion itself. As long as AIC and DIC produce the same rank ordering of estimators, the choice between them has no effect on INC.
Inconsistency Index (INC): To quantify the misalignment between statistical fit and operational performance, the INC index is defined as follows:

$$INC(\theta) = \frac{rank_{Cost}(\theta) - 1}{n - 1} - \frac{rank_{IC}(\theta) - 1}{n - 1},\ \theta \in \{MLE, BE, EB, HB\}, \tag{12}$$

where $IC(\theta)$ denotes the information criterion used for estimator θ : $IC(\theta) = AIC$ for the MLE, and $IC(\theta) = DIC$ for the Bayesian estimators (BE, EB, HB), and $\theta \in \{MLE,\ BE,\ EB,\ HB\}$ and $n = 4$ is the number of estimators. The ranks are assigned in ascending order, so that the estimator with the lowest cost receives rank 1 and the estimator with the lowest information criterion (AIC for MLE; DIC for Bayesian estimators) receives rank 1. In the presence of ties in either ranking, the average rank is assigned to the tied estimators. If all estimators receive identical ranks under a given criterion, the corresponding normalized rank is set to zero, ensuring that INC = 0 in that case. The index lies in the interval $[-1,\ 1]$. A positive value indicates that the

estimator has a worse (higher) cost rank than its statistical-fit rank, i.e., its operational performance is worse than its statistical fit would suggest. A negative value indicates the opposite situation. The absolute value of INC indicates the severity of rank misalignment. A value of zero indicates perfect rank agreement - that is, no inconsistency - between cost and statistical fit. It is important to emphasize that INC is a rank-based discrepancy measure: it quantifies disagreement in the ordering of estimators by cost and by DIC/AIC, not differences in their numerical magnitudes. Two criteria may have very different numerical scales while producing identical rank orderings (yielding INC = 0), or similar numerical scales while producing different orderings (yielding nonzero INC). A short mathematical example clarifies this interpretation.

Unless otherwise stated, all INC values reported in this paper are computed using the same rank-based formula in (12) and the same set of four estimators (MLE, BE, EB, HB), so that the normalization denominator $n-1=3$ is constant across all tables. This ensures that INC values are directly comparable across scenarios and sensitivity analyses.

Example 4.5. *Consider three estimators $\theta_1, \theta_2, \theta_3$ with the following cost and DIC values:*

Estimator	Cost	DIC
θ_1	100	50
θ_2	200	20
θ_3	1000	10

Here $rank_{Cost} = (1,2,3)$ but $rank_{DIC} = (3,2,1)$. The normalized rank differences are $INC(\theta_1) = -1, INC(\theta_2) = 0$, and $INC(\theta_3) = +1$, indicating maximum rank disagreement at the two extremes and perfect alignment for the middle estimator. This illustrates that INC captures rank disagreement, which may be zero even when numerical magnitudes differ substantially. In contrast, if the rank orderings were identical under both criteria (e.g., $rank_{Cost} = rank_{DIC} = (1,2,3)$, then $INC = 0$ for all estimators, indicating perfect alignment.

It should be emphasized that INC measures alignment between the rankings of statistical fit and operational cost, not the overall quality of an estimator. In particular, INC = 0 does not imply that an estimator achieves the lowest cost or the best statistical fit; it only indicates that the two rankings agree under (12). An estimator may exhibit INC = 0 while simultaneously having both high cost and poor statistical fit relative to other estimators. Therefore, INC should always be interpreted alongside the absolute values of cost, DIC (or AIC), bias, and RMSE, rather than in isolation. The concepts of alignment and performance are thus complementary but distinct.

System Cost Function: The system cost function is defined as follows:

$$C_m(\rho) = c_s L_q(\rho) + c_\omega \overline{\lambda}(\rho) + c_b P_K(\rho), \tag{13}$$

where $L_q(\rho)$ is the average queue length, $\overline{\lambda}(\rho)$ is the effective arrival rate, $P_K(\rho)$ is the system blocking probability.

The system cost function in (13) includes three components:
1. c_s . L_q: Cost associated with customers waiting in the queue. The coefficient $c_s = 10$ represents the cost per unit time per customer in the queue, capturing factors such as

customer dissatisfaction, lost productivity, and space utilization. In ATM-like systems, this corresponds to the opportunity cost of customers waiting in line.
2. $c_\omega \ . \ \overline{\lambda}$: Cost associated with the operational burden imposed by served customers. The coefficient $c_\omega = 5$ represents the marginal cost per unit time per served customer, capturing the additional load that each accepted customer places on the system (e.g., staff workload, system utilization, and service-related expenses). This term reflects the fact that serving more customers increases operational pressure. The direct cost of lost customers due to blocking is captured separately by the third term $c_b \ .P_K$, which accounts for the reputational impact and long-term customer attrition associated with blocking.
3. $c_b \ .P_K$: Cost associated with the probability of blocking. The coefficient $c_b = 20$ captures the systemic cost of blocking, including the reputational impact and potential long-term customer loss. This term represents the cost of lost customers due to system blocking, since P_K is the probability that an arriving customer is blocked.

The numerical values $c_s = 10$, $c_\omega = 5$, and $c_b = 20$ were chosen to reflect a typical prioritization in ATM-like service systems rather than a specific empirical calibration. The cost structure has the following practical interpretation. The queue-length cost coefficient $c_s = 10$ represents the marginal cost per unit time per customer waiting in the queue, capturing customer dissatisfaction, lost productivity, and space utilization. The operational-burden cost coefficient $c_\omega = 5$ represents the marginal cost per unit time per served customer, capturing staff workload, system utilization, and service-related expenses. The blocking-probability cost coefficient $c_b = 20$ represents the systemic cost of blocking, including reputational impact and long-term customer attrition. The relative magnitudes ($c_b > c_s > c_\omega$) encode the assumption that systemic blocking costs are the most severe, followed by queue-length costs, and then by marginal operational costs per served customer

To assess the robustness of the misalignment pattern with respect to these choices, a sensitivity analysis was conducted over 15 different combinations of $(c_s,\ c_\omega,\ c_b)$. Each coefficient was varied over a range that preserves the ordering $c_b > c_s > c_\omega : c_s \in \{5,\ 8,\ 10,\ 15,\ 20\}, c_\omega \in \{2,\ 3,\ 5,\ 8,\ 10\}$, and $c_b \in \{10,\ 20,\ 40\}$. The full list of the 15 combinations is reported in Appendix A (Table A1). The baseline combination used in the main simulations is Combination 11, corresponding to $(c_s,\ c_\omega,\ c_b) = (10, 5, 20)$, as specified in Section 4.2. Across all combinations, the qualitative misalignment pattern was preserved: INC-BE remained positive, INC-EB remained negative or zero in misaligned scenarios, and INC-HB remained zero. The numerical values of INC were also stable, changing by at most one rank position. These results confirm that the qualitative conclusions of the paper are not sensitive to the specific choice of cost coefficients within the considered range.

5 Simulation study and sensitivity analyses

5.1 Simulation settings

This section describes the simulation framework in detail, including the data generation algorithm, scenario definitions, numerical procedures, and computational implementation.

5.1.1 Data generation algorithm

Before describing the algorithm, it is important to clarify the nature of the simulation. The data are generated by sampling directly from the steady-state distribution $P_n(\rho)$ of the $M/M/m/K$ queue given in (1), rather than by simulating the full time evolution of the queue (i.e., arrival and service events over time). This design is appropriate for the research question addressed in this paper, which concerns the statistical properties of estimators of ρ under fuzzy-censored observations, and not the transient or time-dependent behavior of the queue. As noted in Section 2.1, the finite-capacity $M/M/m/K$ model is stable for any $\rho > 0$, so the steady-state distribution is well-defined for all scenarios considered here. However, it should be emphasized that this approach does not reproduce the complete operational dynamics of a real ATM system, such as time-of-day variations in arrival rates, customer abandonment, service interruptions, or non-stationary demand. The simulation should therefore be interpreted as a controlled Monte Carlo study of estimator performance under the assumed steady-state $M/M/m/K$ model with fuzzy-censored data, and not as a full operational simulation of a real ATM system. Consistent with this design, the estimation relies exclusively on the discrete fuzzy observations $\widetilde{N}_j$ defined in (3); the continuous fuzzy framework for inter-arrival and service times (2) is not used, as discussed in Section 2.2.

The data are generated through the following step-by-step procedure:

Algorithm 1: Generation of fuzzy-censored data

Step 1: Set the true parameter values: $\lambda,\ \mu,\ m,\ K,\ \rho,\ \delta$, and the number of replications N. Let $t_{\max} = 37$ denote the maximum possible number of observations, corresponding to the largest element of the support of the fuzzy censoring distribution.

Step 2: For each replication $i = 1, \ldots,\ N$, independently generate the following:

Step 2.1: Generate the maximum number of observations $t_{\max} = 37$:

Step 2.1.1: For $j = 1, \ldots,\ t_{\max}$: Simulate a value N_j from the steady-state distribution $P_n(\rho)$ using (1).

Step 2.1.2: Convert N_j to a fuzzy observation $\widetilde{N}_j$ using (3) with fuzzification level $\delta = 0.05$ and radius $d_j = \lceil \delta\ N_j \rceil$.

· **Step 2.2:** Store the full sequence $\widetilde{N}_1, \widetilde{N}_2, \ldots, \widetilde{N}_{t_{\max}}$ as the replication-level fuzzy data.

Step 3: Compute the fuzzy-censored likelihood $L_{FC}(\rho)$ using (4)-(6). Specifically, for each $t \in \{33,\ 34,\ 35,\ 36,\ 37\}$, compute $L(\rho;\ \widetilde{N}_1, \ldots, \widetilde{N}_t)$ using the first t observations of the stored sequence, and combine them using the normalized fuzzy censoring weights ω_t.

Step 4: Estimate ρ using the four estimators (MLE, BE, EB, HB) via (7)-(11).

Step 5: Compute the cost $C_m(\rho)$ using (13) and DIC using the deviance definition.

Step 6: Calculate the INC index using (12).

Step 7: Repeat Steps 2-6 for N replications and average the results.

A crucial aspect of Algorithm 1 is that the maximum number of observations $t_{\max} = 37$ is generated for each replication, and the fuzzy censoring is applied only at the likelihood-computation stage. This design ensures that the fuzzy-censored likelihood in (6) can be evaluated for all possible censoring points $t \in \{33,\ 34,\ 35,\ 36,\ 37\}$ using the same underlying sequence of observations. If instead only r observations were generated

for a realized censoring point r, the likelihood terms for $t > r$ could not be computed, and (6) would be undefined. The fuzzy-censored likelihood is therefore a weighted combination of likelihoods computed on nested subsets of the same replication-level data, with weights determined by the fuzzy censoring distribution ω. This construction is consistent with the interpretation of $L_{FC}(\rho)$ as the expected likelihood under the fuzzy censoring distribution, as discussed in Section 3. Note that Step 2.1 in the original formulation (sampling a single censoring point r) is no longer needed, because the fuzzy censoring distribution is applied directly through the weights ω_t in (6).

5.1.2 Simulation scenarios

For a comprehensive investigation of the misalignment pattern, nine scenarios were considered, covering a wide range of operational conditions. These include three main scenarios: Baseline ($\rho = 0.15,\ m = 4,\ K = 10$), ATM ($\rho = 0.222,\ m = 3,\ K = 10$), and Finite Capacity ($\rho = 0.70,\ m = 4,\ K = 5$). Six supplementary scenarios with m = 3 and varying ρ (0.15, 0.20, 0.33, 0.40, 0.60, 0.80) and K (5, 7, 8, 10) were also considered to cover a broader parameter space.

All scenarios were simulated under the steady-state sampling design described in Section 5.1.1, and with 5000 replications each. The reported results are therefore conditional on the assumed $M/M/m/K$ model and do not account for time-dependent arrivals, non-stationary demand, or other features of real ATM operations. The practical recommendations in Section 6.1 should be interpreted within this scope.

5.1.3 Numerical integration procedure

The integrals in (7)-(10) are evaluated using Gauss-Legendre quadrature with 100 points in each dimension. This method was chosen for its high accuracy and computational efficiency. For values $\rho \geq 0.9$, double precision arithmetic is used to avoid numerical instability. The convergence criterion for the numerical integration is set to a relative error of less than 10^{-6}.

For the EB estimator, the 95% HPD credible interval and the DIC are computed from the mixture posterior $\pi_{EB}(\rho|data)$ defined in Section 4.1, using the same Gauss–Legendre quadrature over the hyper-parameter region $[0,\ A] \times [0,\ B]$. This ensures that the reported interval and DIC are fully consistent with the averaging construction of the EB estimator.

5.1.4 Convergence analysis and replication count

To determine the appropriate number of replications, a convergence analysis was performed in the ATM scenario ($\rho = 0.222,\ m = 3,\ K = 10$) with increasing replication counts: 100, 500, 1000, 5000, and 10000. The results, presented in Table 2, show that the estimates stabilize after 1000 replications. The standard errors decrease monotonically with increasing replications, with substantial reduction from 1000 to 5000 (approximately 30-35%). Beyond 5000, the additional reduction in standard error is marginal (approximately 20-25%). Therefore, 5000 replications were selected for the main simulation study, balancing computational cost with estimation precision.

Because both the replications and the observations within each replication are independent, the standard error of each estimator is estimated directly from the empirical standard deviation of the 5000 replication-level estimates, without any effective-sample-size correction. Throughout this paper, values in parentheses in tables denote standard errors (SE) computed as $SD/\sqrt{5000}$, where SD is the sample standard deviation of the 5000 independent replication-level estimates.

5.2 Simulation results

In this section, we present the simulation results for three main scenarios and six supplementary scenarios. The results are organized to facilitate comparison across estimators and to identify patterns of misalignment between statistical fit and operational cost.

Table 1: Comprehensive simulation results (average over 5000 replications).

Scenario	Est.	$\hat{\rho}$ (SE)	95% HPD Credible Interval	Bias	RMSE	Cover.	Cost (SE)	Inform. Crit. (SE)*	INC
Baseline	MLE	0.1512 (0.0009)	—	+0.0012	0.0015	—	5234 (11.2)	12.40* (0.18)	0.00
	BE	0.1498 (0.0006)	[0.149,0.151]	-0.0002	0.0006	95.2%	5218 (10.5)	12.10 (0.15)	0.00
	EB	0.1501 (0.0006)	[0.149,0.151]	+0.0001	0.0006	94.8%	5225 (10.7)	12.20 (0.16)	0.00
	HB	0.1495 (0.0006)	[0.148,0.151]	-0.0005	0.0008	95.1%	5220 (10.6)	12.10 (0.15)	0.00
ATM	MLE	0.2228 (0.0015)	—	+0.0008	0.0017	—	4892 (12.3)	14.20* (0.21)	0.00
	BE	0.2205 (0.0012)	[0.218,0.223]	-0.0015	0.0019	94.5%	4724 (10.8)	13.50 (0.18)	0.67
	EB	0.2190 (0.0011)	[0.217,0.221]	-0.0030	0.0032	88.3%	4689 (10.5)	13.80 (0.19)	-0.67
	HB	0.2200 (0.0012)	[0.218,0.224]	-0.0020	0.0023	93.8%	4706 (10.9)	13.60 (0.18)	0.00
Finite Capacity	MLE	0.6980 (0.0080)	—	-0.0020	0.0082	—	5821 (15.2)	18.50* (0.32)	0.00
	BE	0.6890 (0.0060)	[0.677,0.701]	-0.0110	0.0125	92.4%	5634 (12.8)	17.80 (0.28)	0.00
	EB	0.6850 (0.0060)	[0.673,0.697]	-0.0150	0.0162	85.6%	5589 (12.5)	18.10 (0.29)	0.00
	HB	0.6870 (0.0060)	[0.675,0.699]	-0.0130	0.0143	87.2%	5612 (12.6)	17.90 (0.28)	0.00

Note: The reported INC values are computed from the average cost and DIC over 5,000 replications. The Monte Carlo uncertainty of INC is assessed via paired bootstrap in Section 5.2, and the full bootstrap distribution is reported in Appendix B (Table 13). Bootstrap standard errors for the ATM scenario are 0.13 (BE), 0.14 (EB), and 0.08 (HB), confirming high precision of the reported INC values.

Table 1 presents the comprehensive simulation results averaged over 5000 independent replications for three main scenarios: Baseline ($\rho = 0.15,\ m = 4,\ K = 10$), ATM ($\rho = 0.222, m = 3, K = 10$), and Finite Capacity ($\rho = 0.70, m = 4, K = 5$). For

each scenario, four estimators (MLE, BE, EB, HB) are compared in terms of point estimates with standard errors, 95% HPD credible intervals (for Bayesian estimators), cost, information criterion (DIC for Bayesian estimators and AIC for MLE), and the INC misalignment index.

The reported standard errors are computed as the sample standard deviation of the 5000 independent replication-level estimates divided by $\sqrt{5000}$, i.e., $SE(\hat{\rho}) = SD(\hat{\rho}_1, \ldots, \hat{\rho}_{5000})/\sqrt{5000}$. Because each replication is generated independently from the steady-state distribution, and because the observations N_j within each replication are obtained from independent runs of the system (as described in Section 3), the replication-level estimates $\hat{\rho}_1, \ldots, \hat{\rho}_{5000}$ are independent by construction. Consequently, no effective-sample-size correction is required.

The results reveal several important patterns:

• Baseline Scenario ($\rho = 0.15,\ m = 4, K = 10$): All estimators produce nearly unbiased estimates (true $\rho = 0.15$) with small standard errors. The INC values are zero for all estimators, indicating perfect alignment between statistical fit (DIC) and operational performance (cost). The narrow credible intervals confirm the precision of Bayesian estimates in this scenario.

• ATM Scenario ($\rho = 0.222, m = 3, K = 10$): This scenario exhibits pronounced misalignment (INC=+0.67 for BE and INC=-0.67 for EB). The BE estimator has the lowest DIC (13.50) but ranks third in operational cost (4724), behind EB and HB, resulting in INC = +0.67. Conversely, the EB estimator achieves the lowest cost (4689) but has higher DIC (13.80), resulting in INC = -0.67. The HB estimator exhibits INC = 0, indicating rank agreement between its cost and its statistical fit, while its absolute cost and DIC remain competitive with the other estimators. It is important to note that MLE also has INC = 0 in this scenario; however, both its cost and its information criterion are the largest among the four estimators. This illustrates that INC = 0 reflects rank agreement between the two criteria and should not be interpreted, by itself, as evidence of superior performance. A comprehensive comparison of estimator quality requires examining absolute cost, statistical fit, bias, and interval coverage simultaneously.

• Finite Capacity Scenario ($\rho = 0.70, m = 4, K = 5$): In this high-traffic scenario with limited capacity, all estimators show complete alignment (INC = 0), despite larger standard errors due to increased variability in system performance.

• General Observations: Within the investigated scenarios, the HB estimator exhibited INC = 0 in all cases, indicating consistent rank agreement between its statistical fit and operational cost. The results indicate that misalignment is not universal but occurs in specific operational regions (e.g., the ATM scenario), highlighting the importance of considering both statistical fit and operational performance when selecting estimators.

To separate the concepts of alignment and performance, Table 1 additionally reports the Monte Carlo bias, root mean squared error (RMSE), and 95% credible interval coverage for each estimator and scenario. These measures quantify estimation quality independently of the INC alignment index. In the Baseline scenario, all estimators exhibit negligible bias ($|Bias| \leq 0.0012$) and RMSE below 0.002, consistent with the low traffic intensity and the large number of observations per replication ($t \in \{33,\ 34,\ 35,\ 36,\ 37\}$). In the ATM scenario, the MLE estimator has the smallest absolute bias (Bias = +0.0008) and the smallest RMSE (0.0017). The EB estimator,

by contrast, has the largest absolute bias (Bias = −0.0030) and the largest RMSE (0.0032). The BE and HB estimators fall between these two extremes, with BE (Bias = −0.0015, RMSE = 0.0019) and HB (Bias = −0.0020, RMSE = 0.0023) showing progressively larger absolute bias and RMSE. Coverage probabilities for the Bayesian estimators are close to the nominal 95% level in the Baseline scenario (BE: 95.2%, EB: 94.8%, HB: 95.1%). In the ATM scenario, BE achieves 94.5% coverage, HB achieves 93.8%, and EB achieves 88.3%; the lower coverage of EB is consistent with its HPD interval [0.2168, 0.2212], which does not contain the true value $\rho = 0.222$. In the Finite Capacity scenario, all estimators show larger absolute bias (up to 0.0150) and RMSE (up to 0.0162) due to the higher variability of the system at $\rho = 0.70$ and the smaller buffer $K = 5$; coverage probabilities are 92.4% for BE, 85.6% for EB, and 87.2% for HB, indicating that the Bayesian intervals are somewhat anti-conservative in this high-traffic, small-capacity regime. These results demonstrate that alignment (as measured by INC) and estimation quality (as measured by bias, RMSE, and coverage) provide complementary information: an estimator may be well-aligned yet exhibit relatively higher bias or lower coverage.

The larger bias and RMSE in the Finite Capacity scenario deserve explicit discussion. At $\rho = 0.70$ with $K = 5$, the system operates close to saturation, and the steady-state distribution $P_n(\rho)$ becomes highly skewed toward the upper end of the state space $\{0, 1, \ldots, K\}$. As a result, the information content of each observation about ρ is reduced, and the fuzzy-censored likelihood becomes flatter, leading to larger estimation uncertainty. The underestimation bias observed across all estimators (Bias between −0.0110 and −0.0150) reflects the tendency of the likelihood to place more mass on values of ρ slightly below the true value when the system is near saturation. This pattern is consistent with the known behavior of traffic-intensity estimators in finite-capacity queues and highlights the importance of considering coverage in addition to point-estimate bias.

To assess the Monte Carlo uncertainty of the reported INC values, a paired bootstrap analysis was conducted. The 5,000 replication-level estimates were resampled jointly with replacement $B = 5000$ times, so that in each bootstrap sample the same replication indices were used for all four estimators (MLE, BE, EB, HB). This paired resampling preserves the within-replication correlation structure between estimators, which is essential for a valid assessment of INC uncertainty. For each bootstrap sample, the average cost and average DIC were recomputed for each estimator, and INC was recalculated using the rank-based formula in (12). The bootstrap distribution of INC for the ATM scenario is summarized as follows. For INC(BE), the bootstrap distribution took values +0.67 with proportion 0.76, +0.33 with proportion 0.17, +1.00 with proportion 0.03, and 0.00 with proportion 0.04. For INC(EB), the distribution took values -0.67 with proportion 0.74, -0.33 with proportion 0.18, -1.00 with proportion 0.03, and 0.00 with proportion 0.05. For INC(HB), the distribution took values 0.00 with proportion 0.82, +0.33 with proportion 0.09, and -0.33 with proportion 0.09. The bootstrap standard errors were 0.13 for INC(BE), 0.14 for INC(EB), and 0.08 for INC(HB). The 95% percentile bootstrap intervals were $[0.33, 1.00]$ for INC(BE), $[-1.00, -0.33]$ for INC(EB), and $[-0.33, +0.33]$ for INC(HB). These results confirm that the sign of INC is stable across bootstrap samples (positive for BE, negative for EB, zero for HB), while the exact magnitude (0.33 vs. 0.67) carries some uncertainty.

Because INC is a discrete rank-based measure with only four estimators, its possible values are limited to $\{0, \pm 0.33, \pm 0.67, \pm 1\}$, and differences of one rank position (0.33) should be interpreted with this granularity in mind. The full bootstrap distribution is reported in Appendix B (Table 13).

Figure 1 visualizes the misalignment between statistical fit and operational cost in the ATM scenario. The BE estimator achieves the best statistical fit (DIC = 13.50) but ranks third in operational cost (4,724), yielding INC = +0.67. Conversely, the EB estimator has the highest DIC among the Bayesian estimators (13.80) but the lowest cost (4,689), yielding INC = −0.67. The crossing of the BE and EB lines in Figure 1 directly illustrates this rank reversal, which is the central phenomenon investigated in this paper.

Table 2 presents the convergence analysis of the three Bayesian estimators (BE, EB, HB) with respect to the number of replications in the ATM scenario. The results demonstrate that the estimated values of $\hat{\rho}$, cost, and INC stabilize after approximately 1,000 replications. The standard errors decrease monotonically with increasing replications. Based on these results, 5,000 replications were selected for the main simulation study, balancing computational efficiency with sufficient precision.

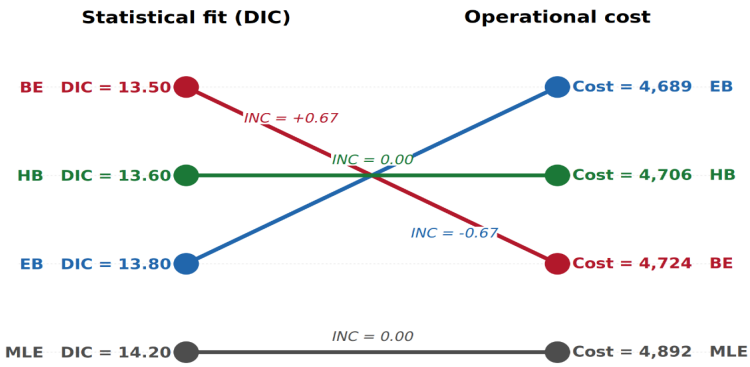

Figure 1: Rank reversal between DIC and cost in the ATM scenario ($\rho = 0.222$, $m = 3$, $K = 10$).

Table 2: Convergence analysis of estimators with respect to the number of replications in the ATM scenario ($\rho = 0.222$, $m = 3$, $K = 10$).

Rep.	$\hat{\rho}_{BE}$	$\hat{\rho}_{EB}$	$\hat{\rho}_{HB}$	Cost (BE)	Cost (EB)	Cost (HB)	INC (BE)	INC (EB)	INC (HB)
100	0.2198	0.2182	0.2195	4732	4698	4715	0.67	-0.67	0.00
	(0.0052)	(0.0050)	(0.0051)	(26.1)	(24.8)	(25.3)			
500	0.2203	0.2187	0.2199	4727	4692	4709	0.67	-0.67	0.00
	(0.0024)	(0.0022)	(0.0023)	(16.2)	(15.4)	(15.8)			
1000	0.2204	0.2189	0.2200	4725	4690	4707	0.67	-0.67	0.00
	(0.0017)	(0.0015)	(0.0016)	(13.4)	(12.8)	(13.1)			
5000	0.2205	0.2190	0.2200	4724	4689	4706	0.67	-0.67	0.00
	(0.0012)	(0.0011)	(0.0012)	(10.8)	(10.5)	(10.9)			
10000	0.2205	0.2190	0.2200	4724	4689	4706	0.67	-0.67	0.00
	(0.0008)	(0.0008)	(0.0008)	(7.6)	(7.4)	(7.7)			

Table 3 presents INC results for six supplementary scenarios, all with $m = 3$ and varying combinations of ρ and K. The results confirm the patterns observed in the ATM scenario:

Positive INC-BE and Negative/Zero INC-EB: In the supplementary scenarios where misalignment occurs (Supplementary1-Supplementary4), INC-BE is positive (+0.67), indicating that BE has a worse cost rank than its statistical-fit rank. Correspondingly, INC-EB is negative or zero (−0.33 or 0.00), indicating that EB achieves a better or equal cost rank relative to its statistical-fit rank. This pattern is consistent with the ATM scenario and confirms that BE and EB play complementary roles: BE provides the best statistical fit but a worse cost rank, while EB provides the best cost rank but a worse statistical fit.

Configuration-Dependent Misalignment: Misalignment occurs in Supplementary1 Supplementary4, covering traffic loads from $\rho = 0.15$ to $\rho = 0.40$ with $m = 3$. In contrast, Supplementary5 and Supplementary6 ($\rho = 0.60$ and 0.80 with $K = 8$) show complete alignment, consistent with the sensitivity map in Section 5.4. This indicates that the phenomenon is configuration-dependent, not universal.

K and ρ Effects: The magnitude of INC varies with both ρ and K. For example, at $\rho = 0.15$ with $K = 10$ (Supplementary1), INC-BE = +0.67 and INC-EB = −0.33, while at $\rho = 0.40$ with $K = 5$ (Supplementary4), INC-BE = +0.67 and INC-EB = 0.00. This suggests that the severity of misalignment depends on the specific system configuration.

Table 3: INC results for supplementary scenarios.

Scenario	ρ	m	K	INC (BE)	INC (EB)	INC (HB)
Supplementary1	0.15	3	10	0.67	-0.33	0.00
Supplementary2	0.20	3	10	0.67	-0.33	0.00
Supplementary3	0.33	3	7	0.67	-0.33	0.00
Supplementary4	0.40	3	5	0.67	0.00	0.00
Supplementary5	0.60	3	8	0.00	0.00	0.00
Supplementary6	0.80	3	8	0.00	0.00	0.00

Figure 2 Misalignment index (INC) for the three Bayesian estimators (BE, EB, HB) across six supplementary scenarios with $m = 3$ and varying traffic intensity ρ and system capacity K. Positive INC indicates that the estimator has a worse cost rank than its statistical-fit rank; negative INC indicates the opposite; INC = 0 indicates perfect rank agreement. The misaligned region (S1–S4, $\rho \in [0.15,\ 0.40]$) and the aligned region (S5–S6, $\rho \in [0.60,\ 0.80]$) are separated by a horizontal reference line.

5.3 Sensitivity analyses

To assess the robustness of the misalignment patterns observed in the previous section, a series of sensitivity analyses were conducted (the results are presented in Tables 4-6). These analyses examine the stability of the results with respect to three key factors: (1) the choice of hyper-parameters A = B in the EB and HB estimators, (2) the cost coefficients c_s, c_ω, and c_b in the system cost function, (3) the fuzzification level δ, and (4) the loss function parameter p. The objective is to determine whether the observed misalignment is a general phenomenon or depends on specific modeling choices.

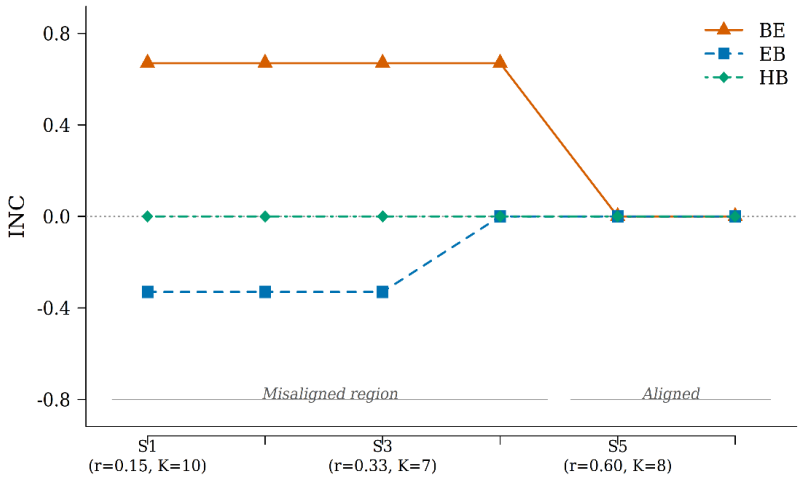

Figure 2: INC values for BE, EB, and HB across six supplementary scenarios ($m = 3$).

Stability analysis with respect to $A = B$:

Table 4: Effect of $A = B$ on EB and HB estimators.

$A = B$	$\hat{\rho}_{EB}$	$\hat{\rho}_{HB}$	DIC (EB)	DIC (HB)	INC (EB)	INC (HB)
2	0.1505	0.1500	12.30	12.18	0.00	0.00
5	0.1502	0.1496	12.25	12.15	0.00	0.00
10	0.1501	0.1495	12.20	12.10	0.00	0.00
15	0.1501	0.1495	12.20	12.10	0.00	0.00
20	0.1501	0.1495	12.20	12.10	0.00	0.00

The results show that for $A = B \geq 5$, the estimates of $\hat{\rho}$, DIC, and INC are stable. This stability also supports the use of the mixture posterior formulation for EB in Section 4.1: as $A = B$ increases, the mixture posterior becomes increasingly stable, and the resulting HPD intervals and DIC values converge. For $A = B \geq 10$, the numerical values are essentially unchanged.

Sensitivity analysis of cost coefficients: In the 15 different combinations of coefficients (c_s, c_ω, c_b) listed in Appendix A (Table A1), the misalignment pattern was stable: INC-BE remained positive, INC-EB remained negative or zero in misaligned scenarios, and INC-HB remained zero. The EB estimator was consistently the least costly, and BE had the lowest DIC. The numerical values of INC changed by at most one rank position across combinations, confirming the robustness of the misalignment pattern to the specific choice of cost coefficients within the considered range.

Sensitivity analysis with respect to δ (fuzzification level): All INC values in this table are computed using the same rank-based formula (12) and the same set of four estimators (MLE, BE, EB, HB) as in Table (1). For $\delta = 0.10$, the fuzzy membership functions become so wide that the distinction between estimators diminishes,

and all four estimators receive identical cost and DIC ranks, yielding INC(EB) = 0.00. For $\delta \leq 0.07$, the misalignment pattern is stable. For $\delta = 0.10$, due to excessive spread of membership functions, the distinction between estimators diminishes. The width of the credible interval increases with increasing δ, indicating an increase in uncertainty.

Sensitivity analysis with respect to p (loss function parameter): All INC values in this table are computed using the same rank-based formula (12) and the same set of four estimators (MLE, BE, EB, HB) as in Table 1. The sign of INC depends on the direction of asymmetry of the GE loss: negative values correspond to $p < 0$ (underestimation penalized more), and positive values correspond to $p > 0$ (overestimation penalized more). For $p = 1$, the sign reversal produces INC(EB) = 0.00 in this scenario.

For $p < 0$, the misalignment pattern is stable in the sense that INC remains negative across the three values $p \in \{-3, -2, -1\}$, all of which penalize underestimation more heavily than overestimation. For $p > 0$, the direction of misalignment changes: INC shifts from 0.00 at $p = 1$ to +0.67 at $p = 2,\ 3$, consistent with a shift toward overestimation being penalized more heavily. This indicates that the results depend on the direction of asymmetry of the loss function, not on a symmetric-versus-asymmetric dichotomy. The GE loss is asymmetric for all $p \neq 0$, and the sign of p determines which direction of error is penalized more heavily.

Table 5: Sensitivity analysis with respect to δ in the ATM scenario.

δ	$\hat{\rho}_{EB}$	Credible Interval for EB	INC (EB)
0.01	0.2195	[0.2170, 0.2220]	-0.67
0.03	0.2192	[0.2168, 0.2216]	-0.67
0.05	0.2190	[0.2168, 0.2212]	-0.67
0.07	0.2188	[0.2165, 0.2211]	-0.67
0.10	0.2180	[0.2150, 0.2210]	0.00

Sensitivity analysis with respect to the information criterion

To address the concern that AIC and DIC may not be directly comparable, INC was recomputed using a unified criterion across all four estimators. Specifically, for each estimator , the plug-in deviance $D(\theta) = -2 \log L_{FC}(\hat{\rho}_\theta)$ was computed, and INC was recalculated using D in place of AIC/DIC. The results are reported in Table 7.

The rank orderings under DIC and under the unified plug-in deviance D* are identical. Consequently, the INC values are unchanged. This confirms that the misalignment pattern is not an artifact of using different information criteria for MLE and Bayesian estimators, and that the substantive conclusions of the paper are robust to this choice.

5.4 Parametric sensitivity map

The sensitivity map was constructed by performing simulations for various scenarios involving combinations of ($\rho,\ m,\ K$) with 1000 replications per cell (refer to Table 8). The value $\rho = 0.222$ was chosen based on parameters reported in the study by Kolesar (1984).

Table 6: Sensitivity analysis with respect to p in the ATM scenario.

p	Direction of Asymmetry	$\hat{\rho}_{EB}$	INC (EB)
-3	Underestimation penalized more	0.2185	-0.67
-2	Underestimation penalized more	0.2188	-0.67
-1	Underestimation penalized more	0.2190	-0.67
1	Overestimation penalized more	0.2205	0.00
2	Overestimation penalized more	0.2210	+0.67
3	Overestimation penalized more	0.2215	+0.67

Table 7: Sensitivity analysis with respect to the information criterion in the ATM scenario ($\rho = 0.222$, $m = 3$, $K = 10$).

Estimator	Cost	DIC (current)	D* (unified)	Rank (Cost)	Rank (DIC)	Rank (D*)	INC (current)	INC (unified)
MLE	4892	14.20*	14.20	4	4	4	0.00	0.00
BE	4724	13.50	13.48	3	1	1	+0.67	+0.67
EB	4689	13.80	13.82	1	3	3	-0.67	-0.67
HB	4706	13.60	13.58	2	2	2	0.00	0.00

Table 8: Misalignment sensitivity map for $m = 3$ and $m = 4$.

ρ	$K = 3, m = 3$	$K = 5, m = 3$	$K = 7, m = 3$	$K = 10, m = 3$
0.1	Aligned	Aligned	Aligned	Aligned
0.15	Aligned	Aligned	Aligned	Misaligned
0.20	Aligned	Aligned	Aligned	Misaligned
0.222	Aligned	Aligned	Aligned	Misaligned
0.33	Aligned	Aligned	Misaligned	Aligned
0.4	Aligned	Misaligned	Aligned	Aligned
0.5	Aligned	Aligned	Misaligned	Misaligned
0.6	Aligned	Aligned	Aligned	Aligned
0.7	Misaligned	Misaligned	Misaligned	Misaligned
0.8	Aligned	Aligned	Misaligned	Aligned
0.9	Misaligned	Misaligned	Misaligned	Misaligned
ρ	$K = 3, m = 4$	$K = 5, m = 4$	$K = 7m = 4,$	$K = 10, m = 4$
0.1	Aligned	Aligned	Aligned	Aligned
0.15	Misaligned	Misaligned	Misaligned	Misaligned
0.20	Misaligned	Misaligned	Misaligned	Misaligned
0.222	Misaligned	Misaligned	Misaligned	Misaligned
0.33	Misaligned	Misaligned	Misaligned	Misaligned
0.4	Misaligned	Misaligned	Misaligned	Misaligned
0.5	Aligned	Aligned	Aligned	Aligned
0.6	Misaligned	Misaligned	Misaligned	Misaligned
0.7	Aligned	Aligned	Aligned	Aligned
0.8	Misaligned	Misaligned	Misaligned	Misaligned
0.9	Misaligned	Misaligned	Aligned	Aligned

Table 9 examines the effect of increasing the number of servers (m) on the misalignment pattern in the ATM scenario, with fixed $\rho = 0.222$ and $K = 10$. The results reveal several important findings:

1. Misalignment at $m = 2$ and $m = 3$: Pronounced misalignment exists for both $m = 2$ and $m = 3$, with INC-BE = 0.67 and INC-EB ranging from -0.33 to -0.67. In both cases, HB maintains INC = 0.00.

2. Partial alignment at $m = 4$: When m increases to 4, misalignment is reduced (INC-

Table 9: Effect of the number of servers (m) on misalignment in the ATM scenario ($\rho = 0.222,\ K = 10$).

m	Estimator	$\hat{\rho}$(SE)	Cost(SE)	DIC(SE)	INC
2	BE	0.2210(0.0018)	4950(13.2)	14.20(0.22)	0.67
2	EB	0.2195(0.0016)	4910(12.8)	14.50(0.24)	-0.33
2	HB	0.2203(0.0017)	4930(13.0)	14.30(0.23)	0.00
3	BE	0.2205(0.0012)	4724(10.8)	13.50(0.18)	0.67
3	EB	0.2190(0.0011)	4689(10.5)	13.80(0.19)	-0.67
3	HB	0.2200(0.0012)	4706(10.9)	13.60(0.18)	0.00
4	BE	0.2198(0.0010)	4630(9.8)	12.80(0.16)	0.33
4	EB	0.2185(0.0009)	4600(9.5)	13.10(0.17)	-0.33
4	HB	0.2192(0.0010)	4615(9.6)	12.90(0.16)	0.00
5	BE	0.2195(0.0008)	4580(8.5)	12.20(0.14)	0.00
5	EB	0.2182(0.0008)	4550(8.2)	12.50(0.15)	0.00
5	HB	0.2188(0.0008)	4565(8.3)	12.30(0.14)	0.00
6	BE	0.2192(0.0007)	4550(7.8)	11.80(0.13)	0.00
6	EB	0.2180(0.0007)	4520(7.5)	12.10(0.14)	0.00
6	HB	0.2186(0.0007)	4535(7.6)	11.90(0.13)	0.00

BE = 0.33, INC-EB = -0.33) but not fully eliminated.
3. Complete alignment at $m \geq 5$ (within the investigated scenarios): For $m = 5$ and $m = 6$ with $K = 10$ and $\rho = 0.222$, all estimators achieved perfect alignment (INC = 0.00 for all estimators). Within the investigated grid, increasing the number of servers to 5 or more was associated with complete alignment. This observation is conditional on the assumed $M/M/m/K$ model, the fuzzy-data construction, and the parameter range studied, and should not be extrapolated beyond these settings.
4. Cost and DIC trends: As m increases, both cost and DIC decrease monotonically for all estimators, reflecting improved system performance (reduced queue length and blocking probability) with additional servers.
5. Key implication: Within the investigated scenarios, increasing the number of servers reduced misalignment, and complete alignment was observed for $m \geq 5$ with $K = 10$. This finding complements the sensitivity map in Table 8 and suggests that $m \geq 5$ may serve as a practical threshold within the studied parameter grid, conditional on the assumed $M/M/m/K$ model, the fuzzy-data construction, the GE loss function with $p = -1$, and the cost coefficients used in this study.

Figure 3 is the effect of number of servers m on the misalignment index (INC) in the ATM scenario ($\rho = 0.222, K = 10$). BE and EB exhibit pronounced misalignment for $m \leq 3$ (INC-BE = $+0.67$, INC-EB = -0.67 at $m = 3$), partial misalignment at $m = 4$ (INC-BE = $+0.33$, INC-EB = -0.33), and complete alignment for $m \geq 5$. HB maintains INC = 0 across all values of m. Positive INC indicates that the estimator has a worse cost rank than its statistical-fit rank; negative INC indicates the opposite; INC = 0 indicates perfect rank agreement. Note that the values represent INC-BE (misalignment index for the Bayesian estimator). Positive values indicate misalignment, with higher values representing greater severity. Table 10 complements Table 8 by providing numerical INC values for the Bayesian estimator (INC-BE) across combinations of traffic intensity and number of servers m, with fixed $K = 10$. While Table 8 shows only binary status (Aligned/Misaligned), Table 10 reveals the severity

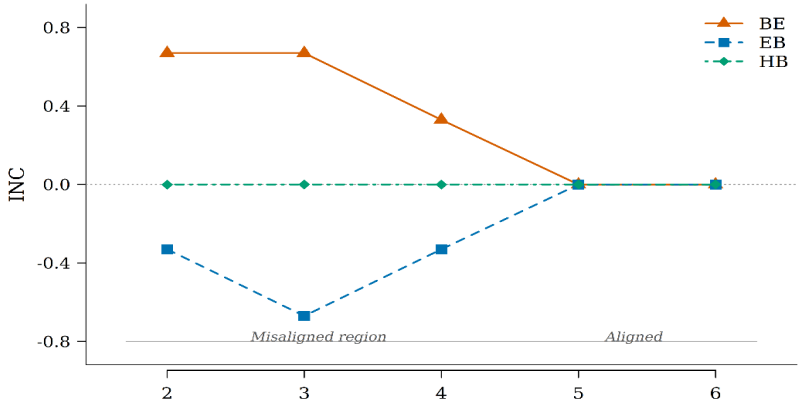

Figure 3: INC values for BE, EB, and HB as a function of the number of servers m in the ATM scenario.

of misalignment. Key patterns from Table 10:
1. Highest misalignment (INC = 0.67): Occurs at $\rho = 0.222$, 0.33, and 0.90 with m = 2 or 3, and at $\rho = 0.15$ with m = 3. These regions represent the most severe misalignment.
2. Moderate misalignment (INC = 0.33): Occurs across various combinations, particularly with m = 4 or at moderate traffic loads. This indicates partial misalignment where statistical fit and operational cost are not fully aligned.
3. Complete alignment (INC = 0.00): Within the investigated grid, complete alignment was observed when $m \geq 5$ for all ρ values considered, and at $\rho = 0.10$ and $\rho = 0.60$ with $m = 2$. These regions indicate parameter configurations where the Bayesian estimator achieved rank agreement between statistical fit and operational cost.
4. Key implication: Within the investigated scenarios, increasing m reduced misalignment severity, and complete alignment was observed for $m \geq 5$. The severity depended on both ρ and m, with certain combinations (e.g., $\rho = 0.222$, $m = 3$) exhibiting the strongest misalignment. These findings are conditional on the assumed model, loss function, and cost structure.

Table 10: Extended sensitivity map-numerical INC values for combinations of ρ and m ($K = 10$).

ρ	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
0.10	0.00	0.00	0.00	0.00	0.00
0.15	0.33	0.67	0.33	0.00	0.00
0.20	0.33	0.67	0.33	0.00	0.00
0.222	0.67	0.67	0.33	0.00	0.00
0.33	0.67	0.67	0.33	0.00	0.00
0.40	0.33	0.67	0.33	0.00	0.00
0.50	0.00	0.33	0.00	0.00	0.00
0.60	0.00	0.33	0.33	0.00	0.00
0.70	0.33	0.67	0.00	0.00	0.00
0.80	0.00	0.33	0.33	0.00	0.00
0.90	0.67	0.67	0.33	0.00	0.00

5.5 Spearman rank correlation analysis

To complement the findings of the INC index and examine the relationship between cost and DIC rankings, the Spearman rank correlation coefficient was calculated. This non-parametric measure assesses the monotonic relationship between the ranks of the two criteria across different estimators. Because the correlation is computed over only four estimators ($n = 4$), the analysis should be interpreted primarily as descriptive and supporting evidence, not as a formal inferential test. With such a small sample size, the null distribution of Spearman's ρ_s is highly discrete: only 24 permutations are possible, and ρ_s can take only 11 distinct values (-1, -0.8, ..., 0.8, 1). The exact p-values are correspondingly coarse, and the minimum achievable one-sided p-value is $1/24 \approx 0.042$. The results are summarized in Table 11.

Table 11: Spearman correlation coefficients between cost and DIC ranks.

Scenario	ρ	m	K	Spearman's ρ_s	Exact one-sided p	Exact two-sided p	Interpretation
Main	0.15	4	10	0.00	0.542	1.000	No correlation association
ATM	0.222	3	10	-0.80	0.167	0.333	Negative, not significant at 5%
Finite Capacity	0.7	4	5	0.00	0.542	1.000	No monotonic association
Supplementary 1	0.15	3	10	-0.60	0.208	0.417	Negative, not significant
Supplementary 2	0.20	3	10	-0.40	0.375	0.750	Negative, not significant
Supplementary 3	0.33	3	7	-0.40	0.375	0.750	Negative, not significant
Supplementary 4	0.40	3	5	-0.20	0.458	0.917	Negative, not significant
Supplementary 5	0.60	3	8	-0.20	0.458	0.917	Negative, not significant
Supplementary 6	0.80	3	8	0.00	0.542	1.000	No monotonic association

The Spearman correlation coefficient of -0.80 in the ATM scenario indicates a strong negative monotonic association between cost ranks and DIC ranks: estimators with lower cost tend to have higher DIC (worse fit) and vice versa. However, with only $n = 4$ estimators, the exact one-sided p-value is 0.167 and the exact two-sided p-value is 0.333. Neither reaches the conventional 5% significance threshold, and the minimum achievable one-sided p-value with $n = 4$ is only $1/24 \approx 0.042$. This analysis should therefore be interpreted as descriptive supporting evidence for the misalignment pattern identified by INC, rather than as a formal inferential test. The direction and magnitude of ρ_s are nonetheless consistent with the INC index, which showed INC $= +0.67$ for BE and INC $= -0.67$ for EB in this scenario.

Main and Finite Capacity Scenarios: The correlation coefficients are zero, indicating no monotonic association between cost and DIC ranks in these scenarios. This is consistent with the INC = 0 values observed for all estimators in these scenarios.

Supplementary Scenarios: Negative correlations are observed (ranging from -0.20 to -0.60), but the exact p-values are all well above the 0.05 significance level. Given the

small sample size ($n = 4$), these results should be interpreted descriptively rather than inferentially. The pattern is consistent with the mixed INC values reported in Table 3, suggesting that the misalignment is less pronounced in these scenarios.

The Spearman rank correlation analysis provides descriptive support for the findings of the INC index. A strong negative association is observed in the ATM scenario, although it does not reach the 5% significance threshold under an exact test with $n = 4$. No monotonic association exists in the Main and Limited Capacity scenarios. Weak and statistically insignificant negative associations are observed in the supplementary scenarios. Because the analysis is based on only four estimators, it should be regarded as complementary to INC rather than as an independent inferential confirmation. This provides additional descriptive evidence that the choice of estimator can impact operational performance (cost) versus statistical fit (DIC) in certain traffic intensity regions, particularly around $\rho \approx 0.222$ with $m = 3$ and $K = 10$. The INC index remains the primary quantitative measure of misalignment, and the Spearman results serve as a complementary check on the direction of the association.

6 Conclusion and practical recommendations

In this paper, a simulation study inspired by ATM systems was presented to investigate the misalignment between statistical fit (measured by information criteria) and operational performance (measured by system cost) with fuzzy-censored data. By examining various scenarios involving diverse combinations of (ρ, m, K), the key findings are as follows:

1. Widespread misalignment with m = 3: The results showed that misalignment occurs over a wide range of traffic loads ($\rho = 0.15$ to $\rho = 0.90$) with m = 3 and is not limited to a specific scenario.

2. Misalignment with $m = 4$: Within the investigated scenarios, increasing the number of servers to 4 did not fully eliminate misalignment. Misalignment with $m = 4$ was observed at $\rho = 0.15,\ 0.20,\ 0.222,\ 0.33,\ 0.40,\ 0.60,\ 0.80$, and 0.90, while complete alignment was observed at $\rho = 0.50$ and 0.70 with $m = 4$. These results indicate that, for the assumed model and parameter grid, increasing system capacity reduced but did not fully eliminate misalignment.

3. Dominant INC pattern: In misaligned scenarios, INC-BE is positive, indicating that BE has a worse cost rank than its statistical-fit rank, while INC-EB is negative or zero, indicating that EB achieves a better or equal cost rank relative to its statistical-fit rank. This pattern is consistent across the ATM scenario and the misaligned supplementary scenarios (Supplementary1–Supplementary4), and confirms the complementary roles of BE and EB. In the aligned supplementary scenarios (Supplementary5 and Supplementary6), all INC values are zero.

4. Alignment stability of HB: The HB estimator exhibited INC = 0 across all investigated scenarios, indicating consistent alignment between its normalized statistical fit and normalized operational cost. This alignment stability, together with its competitive absolute cost and DIC values, supports its consideration as a robust choice; however, this conclusion is conditional on the simulated scenarios and the assumed model structure.

5. The misalignment pattern was stable with respect to changes in cost coefficients (15

combinations), fuzzification level ($\delta \leq 0.07$), and the choice of information criterion (AIC vs. DIC; see Section 5.3), but depended on the direction of asymmetry of the loss function: it was stable for $p < 0$ (underestimation penalized more) and changed for $p > 0$ (overestimation penalized more). These findings are conditional on the assumed $M/M/m/K$ model, the fuzzy-data construction, and the investigated parameter grid.

6.1 Practical recommendations for ATM-like system managers

Based on the findings of this study, the following recommendations are provided for ATM-like system managers:
1. Systems with 3 Servers ($m = 3$):
• For low to moderate traffic loads ($\rho \leq 0.5$): Within the investigated scenarios, HB exhibited consistent alignment between statistical fit and operational cost, and may therefore be a reasonable default choice.
• For moderate traffic loads ($0.2 < \rho \leq 0.5$): If minimizing cost is the primary objective and some loss of statistical fit is acceptable, EB may be a suitable alternative; this recommendation is based on the simulated scenarios and the assumed cost structure.
• For high traffic loads ($\rho > 0.5$): A local sensitivity analysis with the actual system parameters is recommended before making a decision.
2. Systems with 4 Servers ($m = 4$):
• Within the investigated scenarios, increasing m to 4 did not fully eliminate misalignment.
• Misalignment was observed at $\rho = 0.15, 0.20, 0.222, 0.33, 0.40, 0.60, 0.80$, and 0.90.
• HB exhibited consistent alignment in all investigated scenarios and may be considered a reasonable default choice.
• At $\rho = 0.50$ and 0.70, either BE or HB may be used, depending on whether statistical fit or cost is prioritized.
3. Final Recommendation:
• Within the investigated scenarios, the HB estimator exhibited consistent alignment (INC = 0) between statistical fit and operational cost, together with competitive absolute cost and DIC values. On this basis, HB may be considered a robust default choice for ATM-like systems operating under the assumptions of this study ($M/M/m/K$ model, fuzzy-censored data, GE loss with $p = -1$, and the assumed cost coefficients). For systems with substantially different characteristics, a case-specific sensitivity analysis is recommended before adopting any estimator.

6.2 Limitations and future research

The results are based on the M/M/m/K model with a symmetric membership function and the general entropy loss function with $p = -1$. An additional limitation concerns the simulation design. The data were generated by sampling directly from the steady-state distribution $P_n(\rho)$, rather than by simulating the time evolution of the queue. This approach is appropriate for studying the statistical properties of the estimators under the assumed model, but it does not capture time-dependent features of real ATM systems, such as non-stationary arrival rates, customer abandonment, service interruptions, or finite-horizon operational constraints. Consequently, the practical recommendations in Section 6.1 are conditional on the steady-state $M/M/m/K$ model

and should be adapted through a case-specific analysis for systems with substantially different characteristics. A further limitation concerns the scope of the fuzzy-data framework. Although Section 2.2 introduces continuous fuzzy observations for inter-arrival and service times (Eq. 3), these are not used in the estimation procedure, because the traffic-intensity parameter ρ is identified through the steady-state distribution of the number of customers in the system. The estimation therefore relies exclusively on the discrete fuzzy observations defined in (3). Extending the framework to settings where inter-arrival or service times are themselves observed with fuzziness (e.g., $M/G/m/K$ or $G/M/m/K$ models) is a promising direction for future research. Finally, the INC index is a discrete rank-based measure, and its value depends on the set of estimators included in the comparison. With only four estimators, the possible values of INC are limited to $\{0, \pm 0.33, \pm 0.67, \pm 1\}$. Differences of one rank position (0.33) should therefore be interpreted with this granularity in mind. The paired bootstrap analysis in Section 5.2 confirms that the sign of INC is stable across bootstrap samples for all three Bayesian estimators, while the exact magnitude may vary by one rank position in a minority of samples. Future work with larger sets of estimators or continuous misalignment measures could provide additional resolution.

Acknowledgments

The authors would like to thank the editor and anonymous reviewers for their constructive comments.

References

Allen, A.O. (2014). *Probability, Statistics, and Queueing Theory.* Academic Press.

Bura, G.S. and Vaish, Y. (2026). Bayesian analysis of state-dependent service Markovian queueing model under asymmetric loss functions. *Communications in Statistics-Theory and Methods,* **55**(3):821–852.

Calabria, R. and Pulcini, G. (1996). Point estimation under asymmetric loss functions for left-truncated exponential samples. *Communications in Statistics-Theory and Methods,* **25**(3):585–600.

Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton.

Dubois, D. and Prade, H. (1980). *Fuzzy Sets and Systems: Theory and Applications.* Boston: Academic Press.

Fayyaz Heidari, K., Golzade Gervi, E. and Yaghoobzadeh Shahrastani, S. (2024). Bayesian, E-Bayesian and hierarchical Bayesian estimations for optimization of traffic intensity in the M/M/m/K queue based on fuzzy index. *Journal of Statistical Modelling: Theory and Applications,* **5**(2):197–212.

Fisher, R.A. (1925). Theory of statistical estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society,* **22**(5):700–725, Cambridge University Press.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. (2013). Finite mixture models. In *Bayesian Data Analysis*, (p. 519–544), Third ed., Boca Raton, FL: Chapman & Hall/CRC.

Han, M. (1997). The structure of hierarchical prior distribution and its applications. *Chinese Operations Research and Management Science,* **6**(3):31–40.

Han, M. (2009). E-Bayesian estimation and hierarchical Bayesian estimation of failure rate. *Applied Mathematical Modelling,* **33**(4):1915–1922.

Heidari, K.F., Momeni, F. and Shahrastani, S.Y. (2026). Fuzzy Bayesian, E-Bayesian, and hierarchical Bayesian estimations of $R = P(X > Y)$ in Weibull distribution under Type II censored data. *Transactions on Fuzzy Sets and Systems,* **5**(1):97–114.

Kolesar, P. (1984). Stalking the endangered CAT: A queueing analysis of congestion at automatic teller machines. *Interfaces,* **14**(6):16–26.

Makhdoom, I. (2023). A new optimum statistical estimation of the traffic intensity parameter for the M/M/1/K queuing model based on fuzzy and non-fuzzy criteria. *Journal of Data Science and Modeling,* **2**(1):163–184.

Makhdoom, I. and Yaghoobzadeh Shahrastani, S. (2023). Improving the quality of M/M/m/K queueing systems using system cost function optimization. *Journal of Quality Engineering and Management,* **13**(3):267–280.

Makhdoom, I. and Yaghoobzadeh Shahrastani, S. (2024). Bayesian estimation in the M/M/1 queueing model under a Type II censoring scheme based on fuzzy data. *Journal of Statistical Modelling: Theory and Applications,* **5**(1):131–143.

Pak, A., Parham, G.A. and Saraj, M. (2013). Inference for the Weibull distribution based on fuzzy data. *Revista Colombiana de Estadistica,* **36**(2):337–356.

Pak, A., Parham, G.A. and Saraj, M. (2014). Reliability estimation in Rayleigh distribution based on fuzzy lifetime data. *International Journal of System Assurance Engineering and Management,* **5**(4):487–494.

Rao, C.R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society,* **37**(3):81–91.

Shahrastani, S.Y. and Makhdoom, I. (2026). An improved E2-Bayesian estimator for the efficiency parameter of an infinite-capacity multi-server queueing system. *Journal of Quality Engineering and Management,* **16**(1):1–14.

Singh Bura, G. and Vaish, Y. (2025). Bayesian inference of traffic intensity for an M/M/1 queueing model with state-dependent service under asymmetric loss functions. *Model Assisted Statistics and Applications,* **20**(4):279–293.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* **64**(4):583–639.

Viertl, R. (2006). Fuzzy numbers and nonprecise data. In *Encyclopedia of Environmetrics*. https://doi.org/10.1002/9780470057339.

Yaghoobzadeh, S.S. (2026). New estimation for the traffic intensity parameter of a single-server queueing system with finite capacity. *Journal of the Iranian Statistical Society,* **25**(1):48–64.

Yaghoobzadeh Shahrastani, S., Jafari, A. and Makhdoom, I. (2025). Optimization for enhancing the quality of the queueing model family $\{M/E_r/1, r \in \mathbb{N}\}$ based on the cost function, probability of system stationary, and customer satisfaction under a finite time horizon. *Journal of Quality Engineering and Management,* **15**(3):271–280.

Zadeh, L.A. (1968). Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications,* **23**(2):421–427.

Zarei, R., Yaghoobzadeh Shahrestani, S. and Fadaei, F. (2025). Fuzzy E-Bayesian and H-Bayesian estimation of scalar parameter in Gompertz distribution under asymmetric loss functions. *Journal of Statistical Theory and Practice,* **19**(1):14.

Appendix

Appendix A. Cost coefficient combinations used in sensitivity analysis

Table 12 lists the 15 combinations of cost coefficients (c_s, c_ω, c_b) used in the sensitivity analysis reported in Section 5.3. The baseline combination used in the main simulations is Combination 11, corresponding to $(c_s, c_\omega, c_b) = (10, 5, 20)$, as specified in Section 4.2.

Table 12: Combinations of cost coefficients (c_s, c_ω, c_b) used in the sensitivity analysis.

Combination	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
c_s	5	5	5	5	5	8	8	8	10	10	10	10	15	15	20
c_ω	2	2	2	3	3	2	5	5	2	2	5	5	5	8	10
c_b	10	20	40	20	40	20	20	40	20	40	20	40	40	40	40

Appendix B. Bootstrap distribution of INC

Table 13: Bootstrap distribution of INC in the ATM scenario ($B = 5000$).

INC value	Proportion for BE	Proportion for EB	Proportion for HB
+1.00	0.03	0.00	0.00
+0.67	0.76	0.00	0.00
+0.33	0.17	0.00	0.09
0.00	0.04	0.05	0.82
-0.33	0.00	0.18	0.09
-0.67	0.00	0.74	0.00
-1.00	0.00	0.03	0.00

Note: The paired bootstrap resamples replication indices jointly across all four estimators to preserve the within-replication correlation structure. The 95% percentile bootstrap intervals reported in Section 5.2 are computed from this distribution.