IIE Transactions (2005) **37**, 635–649

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170590948495

Empirical Bayes forecasting methods for job flow times

MANSOUR ABDOLI and F. FRED CHOOBINEH*

Department of Industrial and Management Systems Engineering, University of Nebraska-Lincoln, NE 68588, USA
E-mail: fchoobineh@unl.edu

Received December 2003 and accepted December 2004

We consider the problem of forecasting multi-class job flow times in a resource-sharing environment. We assume the deviation of flow times in each class from the class nominal value follows an exponential distribution with its parameter following a gamma distribution. A large simulation experiment is conducted to assess and compare the performance of the Bayes and empirical Bayes forecasting methods under differing model assumptions. Simulation results show that non-parametric empirical Bayes methods are more efficient and robust relative to the parametric empirical Bayes.

1. Introduction

The flow time of an order (a job) in a system is the difference between the release time of the job into the system and the departure time of the job from the system. In a manufacturing environment a job's flow time, sometimes referred to as its cycle time, consists of the actual processing, queuing, and material handling times. In general, the flow time of a job in a system is a random variable. More specifically, in a manufacturing system flow time randomness is caused by: (i) changes in the product mix that affect the queuing and material handling times, (ii) the reliability of processing and material handling devices that affect all components of the flow time; and (iii) natural variations in workers' or machines' production rates. Accordingly, the exact value of a job's flow time becomes known only after the job departs from the system, i.e., a job's flow time is not known *a priori*.

However, prior knowledge (or good estimates) of job flow times is essential to effective planning, control, and management of customer relationships. For example, for planning purposes, estimates of flow times are used for capacity planning and workload leveling; for control purposes, estimates of flow times are used for job scheduling and sequencing; and for customer relationships, estimated flow times are used when promising delivery dates to customers. Since planning, control and customer relationship management are essential functions that are performed for many systems, and these functions rely on estimates of the job flow times, it is essential that these estimates are as accurate as possible.

A forecasting method assigns a value to a random variable (e.g., a flow time) prior to observation of its value. Conventionally, this assigned value is referred to as an estimate or forecast of the random variable's future value. In this paper we use the terms forecast and estimate of the random variable interchangeably.

In general, the complexity of forecasting methods used to estimate job flow times depends on the importance of the flow time estimates in their applications and the sophistication of the user with respect to forecasting methods, as well as the perception of the sources of randomness. For example, a widely used but rudimentary forecast of a job's flow time in manufacturing environments that is utilized in material requirement planning systems is obtained by an empirical rule of thumb. This rule estimates the flow time to be four times the nominal processing time of the job.

In the literature various procedures have been proposed for flow time estimation. A small group of papers assume that the flow times follow a specific distribution with either known or unknown parameters. Examples in this group include the work of Seidmann and Smith (1981) who consider assigning due-dates by forecasting flow times that are identically distributed with a given probability density function, and the work of Chen *et al.* (2001) who assume that the flow times are normally distributed with unknown means and variances. Another example is the work of Hopp and Roof Sturgis (2000) who assume that the flow times are approximately normally distributed with unknown parameters. They treat the unknown parameters as random variables and suggest heuristics to estimate their values based on the number of jobs in the system.

A second group of papers assume that the flow time forecast errors follow a specific distribution. The majority of these papers use regression-based forecasting methods, in which the forecast is a function of some observable attribute plus a normally distributed error. Observable attributes are either job- or system-status-related information. The

*Corresponding author

number of operations and total work content are examples of job characteristics, and the number of jobs in the system and number of jobs in the queues are examples of system status information. Regression-based forecasting methods can be further categorized in terms of the type of observable attributes they use in the forecast. Kaplan and Unal (1993) show that in forecasting flow times, system status information is more useful than job characteristics. Adam *et al.* (1993), Smith *et al.* (1995) and Ruben and Mahmoodi (2000) are other examples of regression-based studies. For more information on such forecasting methods in a due-date assignment context, see Cheng and Gupta (1989).

A third group of papers assume deterministic processing times for multiple classes of jobs and consider the effects of the sequencing of the jobs and their inter-arrival times. Luss and Rosenwein (1993), Bagchi *et al.* (1994), and De *et al.* (1994) are examples of such studies.

Despite the extensive use of Bayesian techniques in the solution of general forecasting problems, they find little application in flow time forecasting problems. An exception is the work of Chen *et al.* (2001) who consider forecasting the mean and variance of normally-distributed flow times when the mean and variance are functions of the work-in-process. They use a two-level prior Bayesian approach to determine function parameters of the mean and variance. However, none of the published flow time forecasting and due-date assignment papers have accounted for process dependencies on different classes of jobs caused by the sharing of resources while still accounting for each class-specific probabilistic flow time.

In this paper we consider the problem of forecasting the flow time of the next job in a production system with m classes of jobs, where for each class i there exist n_i flow time observations such that the mean-squared forecast error is minimized. We use the information contained in the existing observations to estimate the value of the next observation (flow time of the next job leaving the system) and equate that to the flow time forecast of the next job entering the system. This is a valid forecast only under the assumption that the system parameters observed by the entering job will be the same as those parameters observed by the last n_i completed jobs of its class. This assumption is made by most forecasting methods that utilize past observations, but it is particularly reasonable for empirical Bayes methods. The reason is that empirical Bayes methods utilize information from all classes in forecasting the flow time of one class, thus diminishing the need for a large number of historical observations for each class. The validity of this assumption weakens as the length of time to collect the observations increases.

In this paper, we apply Bayes and empirical Bayes methods to obtain *direct* forecasts of the flow times in a multi-product setting, and conduct a simulation study to assess their capabilities and robustness when model assumptions are violated. A direct forecast of a random variable value by using Bayesian methods is novel since, in the reported literature, Bayesian methods are primarily used to estimate the distribution parameters of a random variable. Furthermore, since the flow time forecasting and due-date setting literature have generally ignored the difficult question of how best to account for job flow time dependency among different classes of products, our contributions here provide a venue for considering this dependency while requiring only small observations from each class.

The organization of the paper is as follows. Section 2 describes our flow time forecasting problem in a multi-class production system, and formally defines the problem under investigation. In Section 3, we review the notions of optimality for Bayes and asymptotic optimality for empirical Bayes forecasts. Different forecasting approaches to solve the stated problem are presented in Section 4. In Section 5, we present a simulation experiment to investigate the effectiveness of these approaches, and simulation results are presented and discussed. Finally, Section 6 offers some concluding remarks.

2. Flow times in multi-class production systems

In any production system, the flow time of the jth job of the ith class, T_{ij}, is affected by two sources: (i) controllable factors (such as product mix;) and (ii) uncontrollable factors (such as machine failures). Changes in the controllable factors over time directly affect the flow times by altering resource allocation, and indirectly, by affecting the waiting time distributions of the jobs.

Consequently, the uncontrollable factors that are attributed to natural phenomena, as well as the controllable factors that are attributed to control actions, both contribute to the uncertainty in the job flow times. This uncertainty changes dynamically over time.

Let P_i be the minimum time required to complete a single job of class i in the system. Note that P_i is the summation of the minimal elapsed times in stages of class i routing, which generally differs from the minimum times of other classes. Although flow times are random variables affected by a dynamic environment, P_i is always the same. Therefore, the effect of controllable and uncontrollable factors can be summarized as an increase in the minimum flow time. Accordingly, a random flow time T_{ij} can be expressed in terms of a constant flow time P_i plus a variable increase X_{ij} that summarizes the effects of the variations in the controllable factors and the randomness of the uncontrolled factors i.e., $T_{ij} = P_i + X_{ij}$.

In a well-managed system that experiences little variation, most flow times would be close to their respective P_i. Occasionally some flow times would be way above their minimum value. Thus, it is reasonable to assume that the variable part of the flow time of the jth job of class i, X_{ij}, follows an exponential distribution with parameter θ_i for $i = 1, \ldots, m$, with a caveat that each class of job has its own exponential distribution and the shape of these distributions changes over time.

In a multi-class production system, product families in which products share the same routings can be used as classes of flow times. Since all product families use the same facility, a resource is often shared by more than one product family of the shape parameter, α, allow this model to be adapted to dynamic systems, where there are changes in variability among product classes.

In general, variations in process times for a job affect the waiting times of successive jobs. This follows from the queuing fact that the flow time of a job J is affected by the flow times of all jobs ahead of job J that coexist in the system. Note that the number and type of jobs observed by job J at its release time are random variables. Therefore, when the system traffic intensity is not very high and the system operation is in steady state, it is reasonable to assume that the flow times of jobs are independent.

The following summarizes our model for forecasting flow times in multi-class production systems:

2.1. *The multi-class production model*

Suppose there are m classes of products, and that n_i observations of flow times, denoted by $T_{i1}, \ldots, T_{in_i}$, are given for each class i, $i = 1, \ldots, m$. Find T_i^*, the optimal forecast of the next flow time of class i, i.e., T_{i,n_i+1}, under the squared-error loss where the following holds:

- $T_{ij} = P_i + X_{ij}$, where P_i is the minimum flow time for class i.
- Given θ_i, the X_{ij}s are independent and identically distributed random variables with density f:

$$f(x; \theta_i) = \theta_i e^{-\theta_i x} \mathbf{1}_{[x \geq 0]}. \tag{1}$$

- The θ_is are unobservable random variables, independent and identically distributed with density π.

$$\pi(\theta) = [\Gamma(\alpha)]^{-1} \beta^{\alpha} \theta^{\alpha-1} e^{-\beta\theta} \mathbf{1}_{[\theta > 0]}, \alpha > 1, \beta > 0. \tag{2}$$

In the next section, the above flow time forecasting problem is approached by different forecasting methods.

3. Forecasting techniques: Preliminaries

In this section, we present a formal definition of the optimality criterion under squared-error loss. We also show how forecasting the Xs suffices for finding the optimal forecast of the Ts. Moreover, we briefly review the rudiments of the Bayes and empirical Bayes methods before each forecasting method is presented.

Definition 1. The optimal forecast under squared error.

Suppose $X_1, \ldots, X_n, X_{n+1}$ are $n+1$ random variables and that $X_1, \ldots, X_n$ are observed. We define X_{n+1}^*, a function of $X_1, \ldots, X_n$, as the optimal forecast of X_{n+1} under squared-error loss, if the following inequality holds for any other $\tilde{X}_{n+1}$, which is also a function of $X_1, \ldots, X_n$:

$$E[(X_{n+1}^* - X_{n+1})^2] \leq E[(\tilde{X}_{n+1} - X_{n+1})^2]. \tag{3}$$

For simplicity, we use X^* and $\tilde{X}$ instead of X_{n+1}^* and $\tilde{X}_{n+1}$, respectively. Let us define random variables T_js such that $T_j = P + X_j, j = 1, \ldots, n+1$, where P is a given constant. Then, $T^* = P + X^*$ is the optimal forecast of T_{n+1} under squared-error loss because:

$$\begin{aligned} E[(T^* - T_{n+1})^2] &= E[(P + X^* - (P + X_{n+1}))^2], \\ &= E[(X^* - X_{n+1})^2] \leq E[(\tilde{X} - X_{n+1})^2], \\ &= E[(P + \tilde{X} - (P + X_{n+1}))^2], \\ &= E[(\tilde{T} - T_{n+1})^2], \end{aligned}$$

where $\tilde{T} = P + \tilde{X}$. This implies that the optimal forecast of flow time T_{ij} can be found by forecasting X_{ij}. Therefore, we consider a simple form of the flow time forecasting model where $P_i = 0$, for $i = 1, \ldots, m$; i.e., the X_{ij}s are observed and the X_i^* are sought.

The optimal forecast described in Equation (3) is commonly called the *minimum mean-squared error* forecast, and is given by:

$$X^* = E[X_{n+1} \mid X_1, \ldots, X_n]. \tag{4}$$

The above equation is a well-known result in forecasting time-series data (see for example Box and Jenkins (1970, p. 126)). Furthermore, it is similar to the Bayes estimators under squared-error loss (see for example Mood *et al.* (1974, p. 346)). Nonetheless, one can verify Equation (4) by expanding the expected loss $E[(X^* - X_{n+1})^2]$ as:

$$\begin{aligned} &EE[(X^* - E[X_{n+1} \mid X_1, \ldots, X_n] \\ &+ E[X_{n+1} \mid X_1, \ldots, X_n] - X_{n+1})^2 \mid X_1, \ldots, X_n]. \end{aligned}$$

Analogous to Bayesian estimation, we refer to distributions of X_{n+1} and $X_{n+1} \mid X_1, \ldots, X_n$ respectively as the *prior distribution* and the *posterior distribution* of X_{n+1}. Notice that the X^* given by Equation (4) is the expected value of the posterior distribution of X_{n+1}. Therefore, we refer to X^* as the Bayes forecast under the squared-error loss.

Linear Bayes is another possible approach to the flow time forecasting problem, which follows the relationship between the Bayes forecast and the Bayesian estimation context. When $\tilde{X}_{n+1}$ is restricted to linear functions of $X_1, \ldots, X_n$, the optimal forecast defined by Equation (3) is called the linear Bayes forecast. Let X^{L} denote the linear Bayes forecast under squared-error loss. Then:

$$X^{\mathrm{L}} = w_0 + \sum_{j=1}^{n} w_j X_j, \tag{5}$$

where the w_js, $j = 0, \ldots, n$, are to be determined. Linear Bayes forecasts are often used when it is difficult to obtain Bayes forecast using Equation (4), or when only the first and second moments of the prior distribution are available.

Let Θ, a random variable, denote the parameter that identifies the distribution of the X_is, $i = 1, \ldots, n+1$, and let the X_is be independent for a given $\Theta = \theta$. Then, instead of using the posterior distribution of X_{n+1} to evaluate the Bayes forecast provided by Equation (4), it is convenient to use the posterior distribution of Θ through the following equation:

$$X^* = E[\mu(\theta) \mid X_1, \ldots, X_n], \tag{6}$$

where $\mu(\theta)$ is the function that specifies the mean value of the observations (see for example Klugman *et al.* (1998, p. 427)).

Relating the Bayes forecast to the auxiliary random variable Θ using Equation (6) allows us to employ the empirical techniques used in the Bayesian estimation context when parameters of the distribution of Θ are unknown and several instances of $(\Theta, X_1, \ldots, X_n)$ are available. The empirical Bayes (linear empirical Bayes) methods can be used to estimate the unknown parameters of the distribution of Θ and obtain empirical Bayes (linear empirical Bayes) forecasts that are asymptotically as good as Bayes (linear Bayes) forecasts. Let m available instances of $(\Theta, X_1, \ldots, X_n)$ be denoted by $\{(\Theta_i, X_{i1}, \ldots, X_{in_i})\colon i = 1, \ldots, m\}$ and the Bayes forecasts of $X_{i(n_i+1)}$ for $i = 1, \ldots, m$ be denoted by X_i^* as defined by Equation (6). The asymptotic optimality of the empirical Bayes forecasts denoted by X_i^{EB}, which is a function of m, is defined as follows:

$$\lim_{m\to\infty} \left(E\left[\left(X_{i(n_i+1)} - X_i^{\mathrm{EB}}\right)^2\right] - E\left[\left(X_{i(n_i+1)} - X_i^*\right)^2\right]\right) = 0. \tag{7}$$

The asymptotic optimality of the linear empirical Bayes forecasts, denoted by X_i^{LEB}, can be defined similarly by replacing X_i^{EB} and X_i^*, respectively, with X_i^{LEB} and X_i^{L} (the linear Bayes forecast of $X_{i(n_i+1)}$) in Equation (7).

Both the empirical Bayes and linear empirical Bayes methods are capable of dealing with multi-class data structures; thus, they are appropriate methods for our forecasting problem. Note that these empirical methods are based on the assumption that observations are made for different classes with a random parameter vector $\boldsymbol{\Theta}$. Accordingly, by considering $\boldsymbol{\Theta}$ as a constant vector and combining all classes into one class, non-Bayesian forecasts can be applied to the flow time forecasting problem of our interest. In the next section we present forecasting models that are grouped according to the level of information on $\boldsymbol{\Theta}$.

4. Forecasting techniques

Based on our knowledge of the parameters of the flow time forecasting problem, we distinguish three cases: (i) *model I*, where Θ is gamma distributed and both α and β are known; (ii) *model II*, where Θ is gamma distributed but neither α nor β are known; and (iii) *model III*, where Θ is an unknown constant. These models are discussed separately.

4.1. *Model I: Random Θ with known α and β*

When the parameters of the model are known, both the Bayes and linear Bayes methods can be used to determine the optimal forecast, X_i^*, of the next observation in class i. We use Equation (6) to derive the Bayes forecast, and the Buhlmann's credibility approach (Buhlmann, 1976) to find the linear Bayes forecast.

4.1.1. *The Bayes forecast*

Since $\pi(\theta)$ is a gamma distribution with parameters α and β, the posterior distribution of θ_i, denoted by $\pi^{\mathrm{pos}}(\theta_i)$, would be a gamma distribution with $\alpha + n_i$, $\beta + \sum_{j=1}^{n_i} X_{ij}$. This is a standard result when $(X_1, \ldots, X_n)$ is a random sample from an exponential distribution with rate θ, and θ has a prior gamma distribution with parameters $\alpha > 1$ and $\beta > 0$, and thus the posterior distribution of θ is a gamma with parameters $\alpha' = \alpha + n$ and $\beta' = \beta + \sum_{i=1}^{n} X_i$ (see, for example, Klugman *et al.* (1998, p. 481, exercise 5.8)).

According to $\pi^{\mathrm{pos}}(\theta)$, the optimal forecast, denoted by X_i^{B}, would be:

$$X_i^{\mathrm{B}} = E[1/\theta_i] = \left(\beta + \sum_{j=1}^{n_i} X_{ij}\right) \Big/ (\alpha + n_i - 1). \tag{8}$$

The above follows from the facts that for an exponential distribution with rate θ_i, we have $\mu(\theta_i) = 1/\theta_i$ and for a gamma distributed θ_i with parameters $\alpha' = \alpha + n_i$ and $\beta' = \beta + \sum_{j=1}^{n_i} X_{ij}$, we have $E[1/\theta_i] = \beta'/(\alpha' - 1)$.

4.1.2. *The linear Bayes (credibility) forecast*

In an actuarial context, the credibility model of Buhlmann (1976) credibility model is a forecasting model in which the forecast is determined as a linear function of observations. Buhlmann's credibility problem can be expressed as follows:

Suppose we have observed $(X_{i1} = x_{i1}, \ldots, X_{in_i} = x_{in_i})$ for $i \in \{1, \ldots, m\}$. Let Θ_i denote an unobservable random variable for class i such that for given $\Theta_i = \theta_i$, observations of class i, the X_{ij}s, are independent with $E[X_{ij} \mid \theta_i] = \mu(\theta_i)$ and $\mathrm{Var}[X_{ij} \mid \theta_i] = \upsilon(\theta_i)$ for $j \in \{1, \ldots, n_i\}$ and $i \in \{1, \ldots, m\}$, and $E[\mu(\theta_i)] = \mu < \infty$, $\mathrm{Var}[\mu(\theta_i)] = a > 0$, and $0 < E[\upsilon(\theta_i)] = \upsilon < \infty$ for $i \in \{1, \ldots, m\}$. Find $\tilde{X}_i$, a linear forecast of X_{i,n_i+1}, such that $E[(\tilde{X}_i - \mu(\theta_i))^2]$ is minimized.

Given that a, υ and μ are known, the credibility forecast of X_{i,n_i+1}, $i \in \{1, \ldots, m\}$, denoted by X_i^{C}, is given by:

$$X_i^{\mathrm{C}} = Z_i \bar{X}_i + (1 - Z_i)\mu, \tag{9}$$

where $Z_i = n_i\eta/(n_i\eta + 1)$, $\eta = a/\upsilon$ and $\bar{X}_i = (1/n_i) \sum_{j=1}^{n_i} X_{ij}$. Notice that Equation (9) is an alternative form of Equation (5), where $w_0 = (1 - Z_i)\mu$ and $w_j = Z_i/n_i$. It is also known that X_i^{C} is the best linear forecast in the sense of Equation (3) (see for example Klugman *et al.* (1998, p. 434)), which means for any other linear forecast $\tilde{X}_i^{\mathrm{L}}$, the inequality $E[(X_{i,n_i+1} - X_i^{\mathrm{C}})^2] \leq E[(X_{i,n_i+1} - \tilde{X}_i^{\mathrm{L}})^2]$ holds. Therefore, the credibility forecast is in fact the linear Bayes forecast.

Note that our flow time forecasting model can be represented in the form of the credibility model by:

$$\mu = \beta(\alpha - 1)^{-1}, \tag{10}$$
$$a = \beta^2[(\alpha - 1)^2(\alpha - 2)]^{-1}, \tag{11}$$
$$\upsilon = \beta^2[(\alpha - 1)(\alpha - 2)]^{-1}. \tag{12}$$

Equations (10)–(12) can be derived by employing definitions of μ, a and υ. Using Equations (10)–(12) and definitions of η and Z, one can show that Equations (9) and (8) generate identical forecasts for the flow time forecasting problem of our interest; i.e., the Bayes and the linear Bayes forecasts in this case are identical. This follows from the fact that the gamma and exponential are conjugates.

4.2. *Model II: Random Θ with unknown α and β*

When the parameters of the distribution of θ are unknown, estimates of these unknown parameters can be used to find empirical Bayes forecasts (linear empirical Bayes forecasts) that are asymptotically as good as the Bayes forecast (the linear Bayes forecast). Different methods can be used to estimate the unknown parameters. Estimates of α and β are used in Equation (8) to generate the empirical Bayes forecasts, and estimates of $\eta(= a/\upsilon)$ and μ are used in Equation (9) to generate the linear empirical Bayes forecasts.

4.2.1. *Empirical Bayes forecasts*

When the parameters α and β are unknown, Equation (8) can be used with estimators of α and β to provide a forecast for the next observation as follows:

$$X_i^{\mathrm{EB}} = \left(\hat{\beta} + \sum_{j=1}^{n_i} X_{ij}\right)(\hat{\alpha} + n_i - 1)^{-1}. \tag{13}$$

The following theorem shows that the empirical Bayes forecast provided by Equation (13) asymptotically minimizes the mean-squared error loss, when $\hat{\alpha}$ and $\hat{\beta}$ are bounded consistent estimators of α and β.

Theorem 1. *Let $\{X_{ij} : i = 1, \ldots m, j = 1, \ldots, n_i, n_i > 0\}$ be a set of random variables satisfying assumptions of model I. For simplicity of notation, let $X_i = X_{i,n_i+1}$. Also, let X_i^{B} denote the Bayes forecast of X_i provided by Equation* (8). *Suppose that $\hat{\alpha}$ and $\hat{\beta}$ are bounded consistent estimators of α and β, and X_i^{EB} is the forecast of X_i provided by Equation* (13). *Then X_i^{EB} is an asymptotically minimum mean squared error forecast of X_i in the following sense*:

$$\lim_{m\to\infty} \left(E\big[(X_i - X_i^{\mathrm{EB}})^2\big] - E\big[(X_i - X_i^{\mathrm{B}})^2\big]\right) = 0. \tag{14}$$

Proof. The proof is provided in the Appendix. ■

To find asymptotically optimal forecasts, we need to use appropriate estimators of α and β. We select four estimators of these parameters that have been proposed and suggested by Thiruvaiyaru and Basawa (1992) and Mashayekhi *et al.* (2004). In the following, let $N1$ and $N2$ be the upper bounds for estimators of α and β, respectively, and let:

$$\bar{X}_{..} = m^{-1}\sum_{i=1}^{m}\left(n_i^{-1}\sum_{j=1}^{n_i} X_{ij}\right),$$
$$\overline{X^2_{..}} = m^{-1}\sum_{i=1}^{m}\left(n_i^{-1}\sum_{j=1}^{n_i} X_{ij}^2\right),$$
$$\bar{Y} = m^{-1}\sum_{i=1}^{m}\left(2[n_i(n_i-1)]^{-1}\sum_{j=1}^{n_i-1}\sum_{k=j+1}^{n_i} X_{ij}X_{ik}\right),$$
$$\bar{Y}_{..}(t) = m^{-1}\sum_{i=1}^{m} n_i^{-1}\sum_{j=1}^{n_i}\left(\mathbf{1}_{[X_{ij}<t]}X_{ij} + \mathbf{1}_{[X_{ij}\geq t]}t\right),$$
$$\bar{\delta}_{..}(z) = m^{-1}\sum_{i=1}^{m} n_i^{-1}\sum_{j=1}^{n_i} 1_{[X_{ij}\geq z]}.$$

Method A: This method is based on Mashayekhi *et al.* (2004):

$$\tilde{\beta} = \mathbf{1}_{[S>0]}(\bar{X}_{..}\bar{Y}/S) + \mathbf{1}_{[S=0]}(N_2),$$
$$\tilde{\alpha} = \mathbf{1}_{[S=0]}N_1 + \mathbf{1}_{[S\neq 0]}T,$$
$$\bar{\alpha} = \max(\tilde{\alpha}, 2.0),$$
$$\hat{\alpha} = \min(N_1, \bar{\alpha}), \tag{15}$$
$$\hat{\beta} = \min(N_2, \tilde{\beta}), \tag{16}$$

where $T = 1 + \overline{Y}/S$ and $S = \bar{Y} - \overline{X^2_{..}}$.

Method B: This method is based on Mashayekhi *et al.* (2004):

$$\tilde{\beta} = \mathbf{1}_{[R\neq 0]}[R^{-1}(\bar{\delta}_{..}(1.0)\bar{Y}_{..}(0.5) - 0.5\,\bar{\delta}_{..}(0.5)\bar{Y}_{..}(1.0))] + \mathbf{1}_{[R=0]}N_2,$$
$$\bar{\beta} = \max(\tilde{\beta}, 0),$$
$$\hat{\beta} = \min(N_2, \bar{\beta}), \tag{17}$$
$$\bar{\alpha} = \mathbf{1}_{[\bar{Y}_{..}(1.0)\neq 0]}[\bar{Y}_{..}(1.0)^{-1}(\hat{\beta} - (\hat{\beta} + 1.0)\bar{\delta}\cdot\cdot(1.0)) + 1.0] + \mathbf{1}_{[\bar{Y}_{..}(1.0)=0]}N_1,$$
$$\hat{\alpha} = \min(N_1, \bar{\alpha}), \tag{18}$$

where $R = (1.0 - \bar{\delta}\cdot\cdot(1.0))\,\bar{Y}_{..}(0.5) - (1.0 - \bar{\delta}\cdot\cdot(0.5))\,\bar{Y}_{..}(1.0)$.

The method of moment estimator: The Method of Moment Estimator (MME) is based on Thiruvaiyaru and Basawa (1992):

$$\begin{aligned}\widehat{\alpha} &= 1 + \overline{X_{..}^2}(\overline{X_{..}^2} - 2\overline{X_{..}^2})^{-1},\\ \hat{\alpha} &= \widehat{\alpha} \times \mathbf{1}_{[2<\widehat{\alpha}<N_1]} + 2 \times 1_{[\widehat{\alpha}\leq 2]} + N_1 \times 1_{[\widehat{\alpha}\geq N_1]}, \end{aligned} \quad (19)$$

$$\begin{aligned}\widehat{\beta} &= (\hat{\alpha} - 1)\bar{X}_{..},\\ \hat{\beta} &= \widehat{\beta} \times \mathbf{1}_{[\widehat{\beta}<N_2]} + N\mathbf{2} \times \mathbf{1}_{[\widehat{\beta}\geq N_2],}\end{aligned} \quad (20)$$

where $\widehat{\alpha}$ and $\widehat{\beta}$ are MME estimators of α and β.

The one-step maximum likelihood estimator: The one-step Maximum Likelihood Estimator (MLE) is based on Thiruvaiyaru and Basawa (1992). Let $\eta_0 = [\alpha_{\text{mme}}\ \beta_{\text{mme}}]^T$. The one-step MLE is then attained by employing one step of the Newton search as follows:

$$[\alpha_{1-\text{mle}}\beta_{1-\text{mle}}]^T | = \boldsymbol{\eta}_0 - (\mathbf{H}^{-1}\mathbf{G}) |_{\eta_0}, \quad (21)$$

where $\mathbf{H}$ and $\mathbf{G}$, respectively, are the Hessian and gradient of the log-likelihood function of $\boldsymbol{\eta}$, with

$$\text{Log(Likelihood)} = \sum_{i=1}^{m} \left[\alpha \log(\beta) - (\alpha + n_i) \times \log\left(\beta + \sum_{j=1}^{n_i} x_{ij}\right) + \sum_{j=1}^{n_i} \log(\alpha + n_i - j) \right].$$

4.2.2. *Empirical credibility forecasts*

When the observations are known to follow the structure of the Buhlmann's credibility model expressed in Section 4.1.2, but parameters $\boldsymbol{\eta} = a/\upsilon$ and μ are unknown, estimates of the parameters can be used in Equation (9) to find an empirical credibility forecast as follows:

$$\hat{X}_i^{\text{C}} = \hat{Z}_i \bar{X}_i + (1 - \hat{Z}_i)\hat{\mu}, \quad (22)$$

where $\hat{Z}_i = n_i\hat{\eta}/(n_i\hat{\eta} + 1)$, and $\hat{\mu}$ and $\hat{\eta}$ are estimators of μ and η.

We consider one estimator for $\hat{\mu}$ and two estimators for $\hat{\eta}$, as follows:

Let

$$\hat{\upsilon} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2\right)\left(\sum_{i=1}^{m} n_i - m\right)^{-1},$$

$$\text{MSB} = (m-1)^{-1}\sum_{i=1}^{m} n_i(\bar{X}_i - \bar{X}_{..})^2,$$

$$g = \sum_{i=1}^{m} n_i - \left(\sum_{i=1}^{m} n_i\right)^{-1}\left(\sum_{i=1}^{m} n_i^2\right),$$

$$\tilde{a} = g^{-1}(m-1)(\text{MSB} - \hat{\upsilon}),$$

$$\hat{a} = \max(0, \tilde{a}).$$

Then,

$$\hat{\mu} = \sum^{m}\left(\sum_{i=1}^{m}\hat{Z}_i\right)^{-1}\left(\sum_{i=1}^{m}\hat{Z}_i\bar{X}_i\right) + \bar{X}_{..}\mathbf{1}_{[\hat{a}=0]}. \quad (23)$$

Note that in the above equation, when $\hat{a} = 0$ or $n_i = n$ for all $i = 1, \ldots, m$, the equation simplifies to $\hat{\mu} = \bar{X}_{...}$

Method Crd-I: This estimator is based on Sundt (1983):

$$\hat{\eta}_1 = \hat{a}/\hat{\upsilon}. \quad (24)$$

Method Crd-II: This estimator is based on Ghosh and Meeden (1986) and Ghosh and Lahiri (1987):

$$\hat{\eta}_2 = (((m-3)\hat{\upsilon})^{-1}(m-1)\text{MSB} - 1)(m-1)g^{-1}. \quad (25)$$

4.3. *Model III: Unknown constant* $\boldsymbol{\Theta}$

In practice, it is usually assumed that the θ_is are unknown constants. Note that if θ_i is a given constant for class i, then the X_{ij}s are independent and identically distributed random variables. According to Equation (4), we have:

$$X_i^{\text{NB}} = E[X_{i,n_i+1} \mid X_{i,1}, \ldots, X_{i,n_i}] = E[X_{i,n_i+1}] = \mu(\theta_i) = 1/\theta_i, \quad (26)$$

where X_i^{NB} denotes the optimal forecast of X_{i,n_i+1} and $\mu(\theta_i)$ is the expected value of the common distribution of $X_{i,1}, \ldots, X_{i,n_i}, X_{i,n_i+1}$. Therefore, when θ_i is assumed to be an unknown constant, the average value of $X_{i,1}, \ldots, X_{i,n_i}$ can be used as an estimate of $\mu(\theta_i)$ in Equation (26) to provide a non-Bayesian forecast of X_{i,n_i+1}, as follows:

$$\hat{X}_i^{\text{NB}} = \hat{\mu}(\theta_i) = n_i^{-1}\sum_{j=1}^{n_i} X_{ij}. \quad (27)$$

In some situations, observations of all classes are aggregated into one class with an unknown constant θ. In this aggregated case, the forecast is the overall average as follows:

$$\hat{X}^{\text{NB}} = \hat{\mu}(\theta) = m^{-1}\sum_{i=1}^{m} n_i^{-1}\sum_{j=1}^{n_i} X_{ij}. \quad (28)$$

Figure 1 summarizes the forecasts presented above and depicts their relationships. The empirical forecasts are asymptotically optimal, which implies that the expected loss for all methods is the same when the number of classes goes to infinity. In practice, however, the number of classes is not very large nor can it be increased arbitrarily. Therefore, when parameters are unknown, empirical methods may perform differently. In addition, since the credibility method relies on weaker distribution assumptions, it has a wider practical appeal than other methods; thus, it is reasonable to expect its performance to be robust with respect to parameter assumptions.

Statistical model selection (Lahiri, 2001) criteria, such as variations of the Akaike's information criterion or Schwarz's Bayesian information criterion, are often used in the statistical literature to select a model from a given set of models by using available data. Although a model selection procedure could be used to select the forecasting

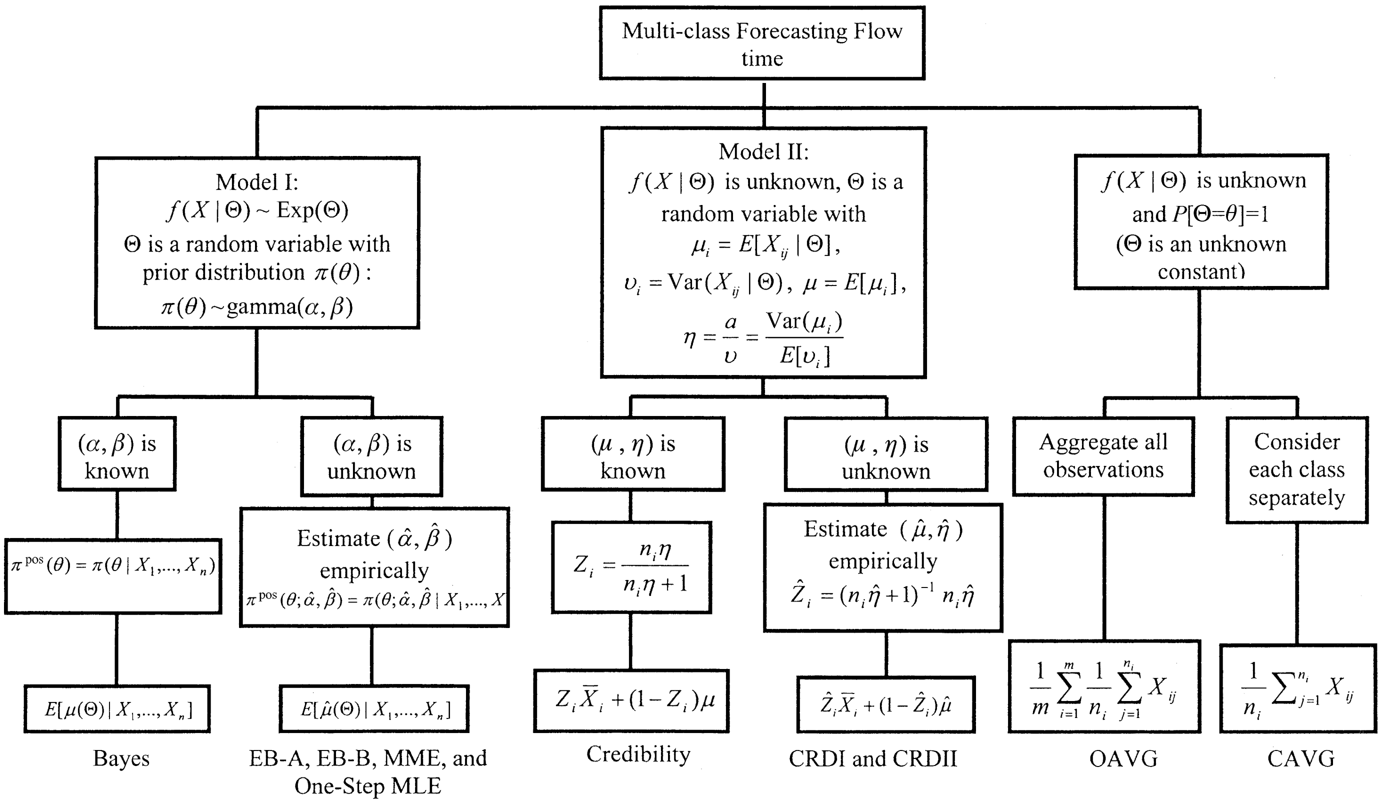

Fig. 1. General relationship among multi-class flow time forecasting problems.

method that best resembles the Bayes method, we chose to employ other specific performance measures and statistical tests because our objective is to understand the behavior of forecasting methods as well as to identify the best method(s). In the next section we describe the elements of our simulation study.

5. Simulation study

The first goal of our simulation study was to compare the performance of different empirical forecasts given the assumptions of the flow time forecasting problem described at the end of Section 3. The simulation study was also conducted to investigate the robustness of each method with respect to violations of the assumption that flow times are distributed exponentially.

To broaden the comparative study, the performance of each method was measured for different parameter values. Four levels of α were considered to account for different flow rates ($\alpha = 3$, 4, 5 and 6). Moreover, three levels of β were considered to expand the variety of the means and variances of production rates ($\beta = 1$, 5 and 10). For simplicity, the number of observations for all classes was assumed to be n, i.e., $n_i = n$ for $i = 1, \ldots, m$, and three levels of n were considered ($n = 5$, 10 and 50). The performance of these methods was observed over five different levels of m ($m = 5$, 10, 20, 50 and 100).

Violation of the assumptions of the flow time forecasting model was achieved by altering the exponential distribution of the X_{ij}s to a gamma distribution with parameters λ and θ_i, as follows:

$$X_{ij} \mid (\lambda, \theta_i) \sim iid \ \mathrm{gamma}(\lambda, \theta_i) \quad \text{for} \quad j \in \{1, \ldots, n_i\} \quad \text{and} \quad \times i \in \{1, \ldots, m\}. \tag{29}$$

Two levels of λ were considered: (i) where $\lambda = 1$ covers the base case (exponentially distributed observations); and (ii) where $\lambda = 10$ offers an extreme violation of the model distribution assumption.

Table 1 summarizes the parameters and their values used in the simulation experiment. Using the values listed in this table, $5 \times 3 \times 4 \times 3 \times 2 = 360$ different situations (cases) were simulated.

Table 1. Parameters of the forecasting problem and their values used in the simulation study

Parameters	*Levels*
M	5, 10, 20, 50, 100
N	5, 10, 50
α	3, 4, 5, 6
β	1, 5, 10
λ	1, 10

The simulation program was developed using Visual® C++ 6.0 and all the random numbers were generated using IMSL® C Library 5.0 routines. The simulation program outputs were processed and analyzed by MATLAB®.

The simulation program generates a random instance for a problem case. This instance is an array of m groups of $n + 1$ random numbers, where each group represents observations of one class. The first n numbers of each class are used for forecasting its $(n + 1)$th observation. For each class, the forecasting losses are calculated based on the $(n + 1)$th generated number and its calculated forecast. Since most empirical forecasting methods use the first n generated numbers of all m classes to estimate their required unknown parameters, they do not generate independent forecasting losses for different classes. Thus, for each method and from each random instance of a problem case, only one realization (out of m realizations) of the forecasting losses was used in our analysis.

In each simulation run, $K = 2000$ instances of each problem case were generated. Figure 2 shows the flowchart for one replication of a simulation run. Each simulation run was replicated 10 times, and outcomes of $10 \times K$ random instances were then used in our analysis. The performance measures and statistical tests used for interpreting outcomes are reviewed in the next section.

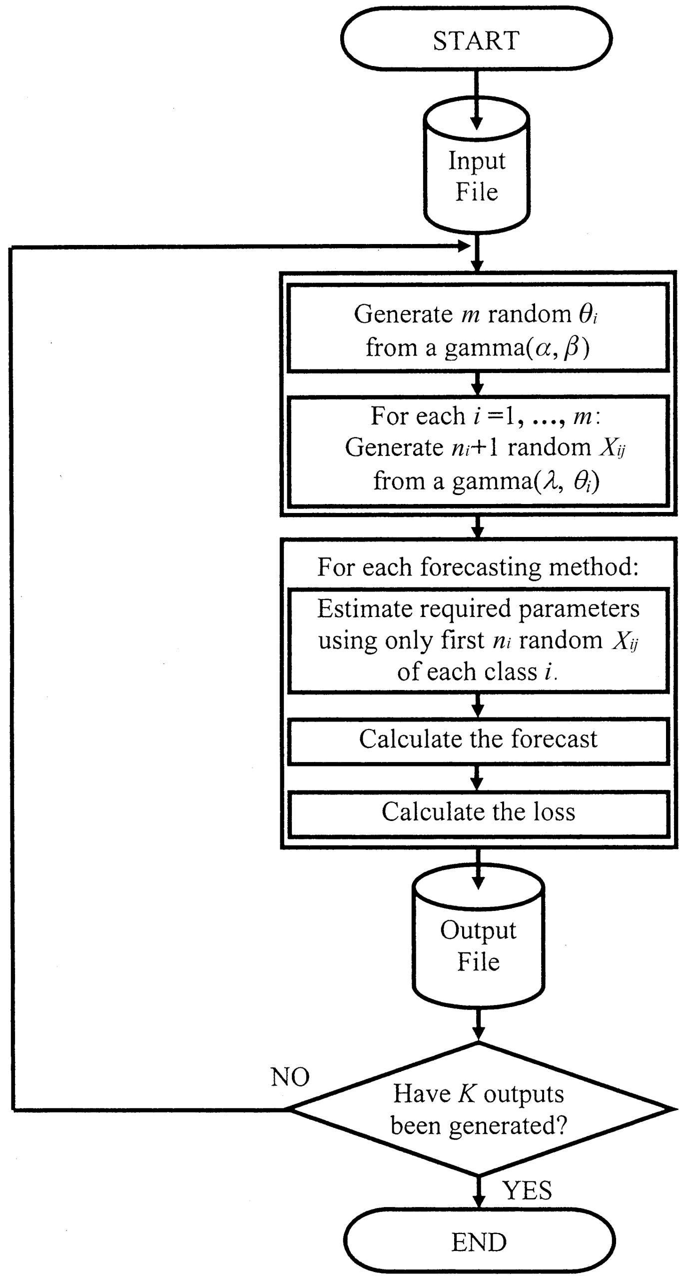

Fig. 2. Flowchart for one replication of the simulation program.

5.1. *Performance measures and statistical tests*

Let L_M denote the loss of forecasting method M. Since forecasts are functions of observations, the loss and the average loss are random variables. The optimality criterion given by Equation (3) requires the forecasting methods to be compared by their expected value of the loss. Therefore, the average loss, $\bar{L}_M$, is the main measure of performance for forecasting method M in our study. Hence, forecasting methods are compared based on their average losses over $10 \times K$ randomly generated cases for each combination of parameters. For this purpose, and to show the effect of parameters m, n, α and β, average losses are plotted separately for each parameter.

The squared-error loss function is symmetric; i.e., it assumes the same loss for both over- and under-forecasted errors. Moreover, the average loss does not provide much information about the skewness and peakedness of the loss distribution. Note that in this application peakedness needs to be evaluated with respect to a loss threshold that specifies the desired level of accuracy. Therefore, to provide some insight about skewness and peakedness of the loss distribution of each forecasting method, the probability of the loss which is less than a given threshold level c, $P[L \le c]$, and the expected value of losses more than c, $E[L \mid L > c]$, are calculated. $P[L \le c]$ provides an indication of the forecasting method relative peakedness (larger is better) and $E[L \mid L > c]$ provides an indication of the tail behavior of L (smaller is better). To summarize the information of these two performance measures, $P[L \le c]$ is plotted versus $\mathrm{E}[L \mid L > c]$ for different values of m. We refer to these graphs as *probability-loss* graphs. For each m, $P[L \le c]$ and $E[L \mid L > c]$ are estimated by averaging the relative frequency of losses less than c and the average of losses greater than c over 10 replications of all cases generated by different values of α, β and n.

Furthermore, the medians of the average losses of all methods are compared using a *pair-wise sign test*. The *paired t-test* for the mean was not used since $\bar{L}_{M1}$ and $\bar{L}_{M2}$, losses of methods $M1$ and $M2$, did not satisfy the normality requirement of the paired t-test. Also, since $\bar{L}_{M1} - \bar{L}_{M2}$ is not symmetric, the *paired signed-rank test* for comparing

Table 2. List of forecasting methods used in the comparative study

Forecasting method	*Abbreviation*
Bayes	BAYES
Empirical Bayes using estimation method A	EB-A
Empirical Bayes using estimation method B	EB-B
Empirical Bayes using the MME	MME
Empirical Bayes using the one-step MLE	MLE
Empirical credibility using Equation (24) for $\hat{\eta}$	CRDI
Empirical credibility using Equation (25) for $\hat{\eta}$	CRDII
Multi-class non-Bayesian Forecast (Equation 27)	CAVG
Single, aggregated class non-Bayesian forecast (Equation 28)	OAVG

means is not applicable. Nonetheless, the sign test can provide reasonable information since distributions of the $\bar{L}_M$s are almost the same. Here, we consider the following hypotheses by using the *pair-wise sign test* on the median, R, of methods $M1$ and $M2$:

$$H_0 : R_{M1} = R_{M2}, \tag{30}$$

$$H_a : R_{M1} > R_{M2}. \tag{31}$$

For each test, the probability of a type-I error is reported, where H_0 is rejected in favor of H_a when that probability is small.

5.2. *Simulation results*

The simulation results are presented in two main parts. In the first part, the effectiveness of the forecasting methods and their sensitivity to the problem parameters (m, n, α and β) are demonstrated when the observations of the X_{ij}s are exponentially distributed ($\lambda = 1$). In the second part, we consider the robustness of the forecasting methods under violation of the model distribution assumptions (by comparing results of $\lambda = 10$ with $\lambda = 1$).

Table 2 shows the abbreviations used for identifying forecasting methods in this study.

5.2.1. *Performance of forecasting methods when $\lambda = 1$*

Figures 3 and 4 show the average losses of different methods as functions of the parameters m and n. As expected, the Bayes method (BAYES) has the smallest loss. Figure 3 shows that the overall average (OAVG) method has the worst average loss, and the other methods are clumped into two groups. The best group consists of methods CRDI, CRDII, MME and EB-A. The MLE method joins the best clump as m increases. Note that the asymptotic optimality of these methods is not visible for $m \leq 100$. Figure 4 shows that as the number of class observations (n) increases, the average loss of all methods decreases.

Figures 5 and 6 show the effect of changes in values of α and β on the average losses and the relative performances of the methods. According to these graphs, the average loss increases as β increases and α decreases, and the relative performance of forecasting methods is not sensitive to changes in values of these parameters.

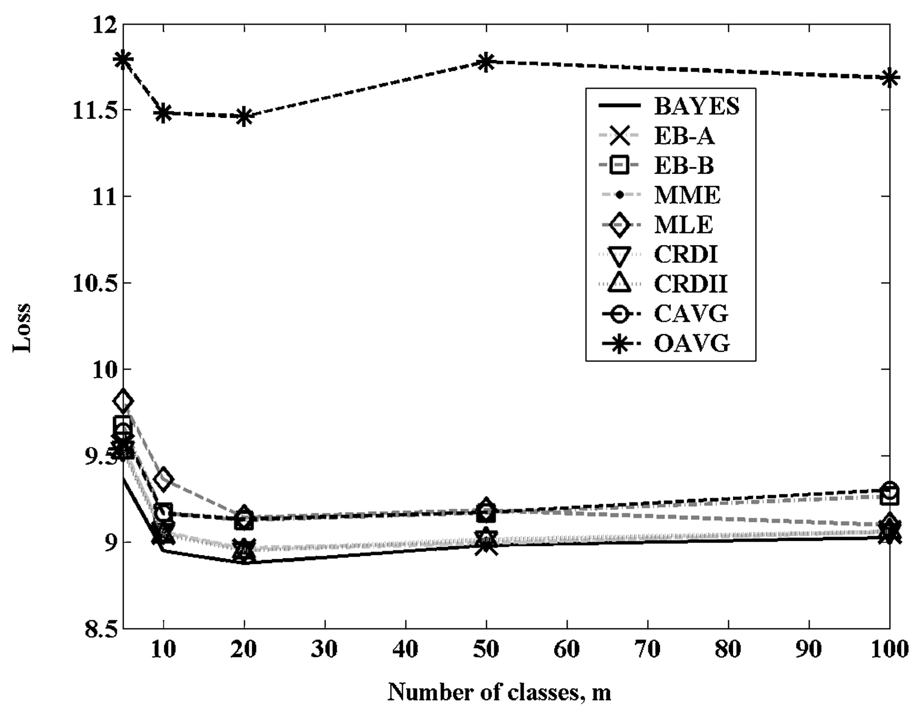

Fig. 3. Average losses against m for the $\lambda = 1$ model.

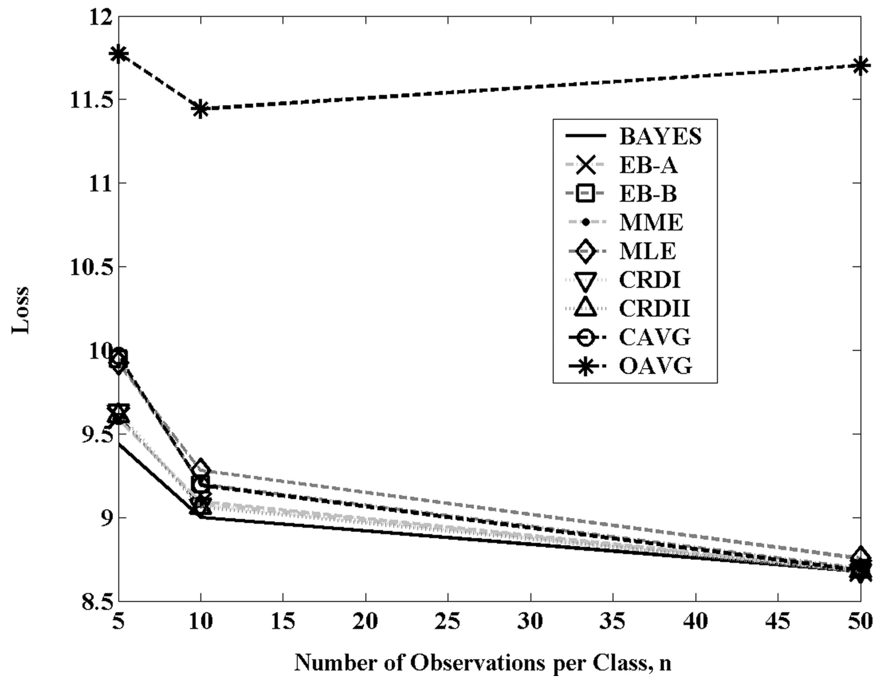

Fig. 4. Average losses against n for the $\lambda = 1$ model.

Figure 7 is the probability-loss graph for BAYES and methods in the best group, i.e., CRDI, CRDII, MME and EB-A. The methods closest to the northwest corner of the graph have the best performance. The graph shows that for $m \geq 10$, the five methods perform approximately the same.

Table 3 summarizes the probability of type-I errors for the hypotheses of Equations (30) and (31), where $M1$ is a column method and $M2$ is a row method. This table shows the pair-wise significance among the methods' median of losses, where methods are ordered based on their relative

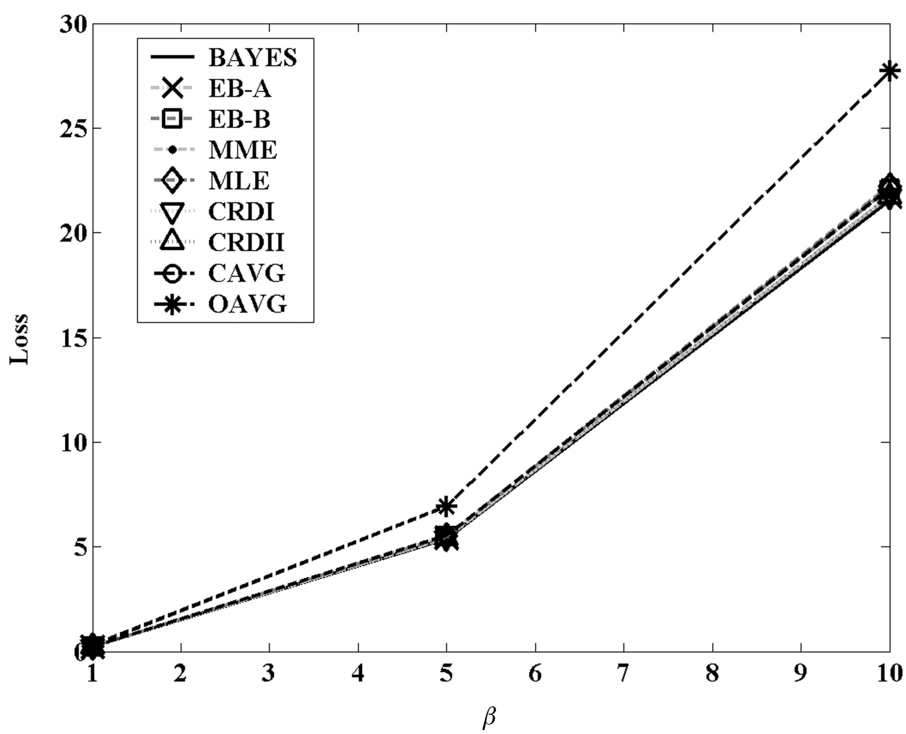

Fig. 5. Average losses against β for the $\lambda = 1$ model.

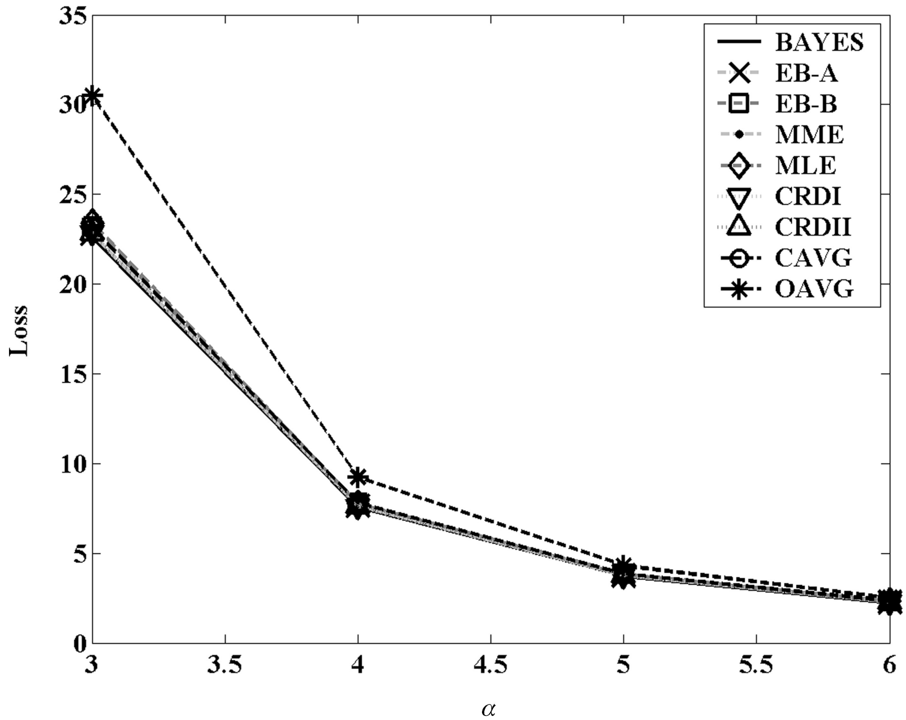

Fig. 6. Average losses against α for the $\lambda = 1$ model.

performance. The best five methods are BAYES, CRDII, CRDI, EB-A and MME.

The above results show that although model parameters affect the performance of each method, they have little effect on the methods' relative performance. The results also support the fact that more information in both class and observation dimensions (m and n) increases the performance of forecasting methods. Finally, the results rank the

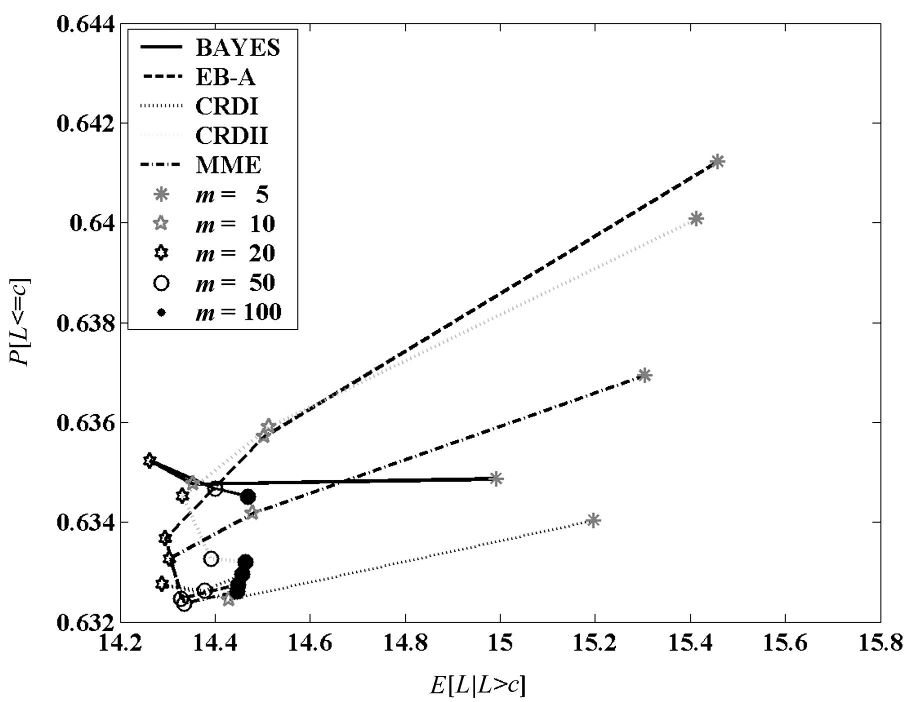

Fig. 7. Probability-loss graph averaged on n, α, and β for the various methods in medium performance cluster for the $\lambda = 1$ model, when $c = 1$.

Table 3. Pair-wise sign test for $\lambda = 1$; H_0: Median (Col) = Median (Row) vs. H_a: Median (Col) > Median (Row)

	BAYES	*CRDII*	*CRDI*	*EB-A*	*MME*	*MLE*	*EB-B*	*CAVG*	*OAVG*
BAYES		0	0	0	0	0	0	0	0
CRDII	1		0	0	0	0	0	0	0
CRDI	1	1		0.0001	0	0	0	0	0
EB-A	1	1	0.9999		0	0	0	0	0
MME	1	1	1	1		0	0	0	0
MLE	1	1	1	1	1		0.7529	0.0001	0
EB-B	1	1	1	1	1	0.2623		0	0
CAVG	1	1	1	1	1	0.9999	1		0
OAVG	1	1	1	1	1	1	1	1	

best four empirical methods as CRDII, CRDI, EB-A and MME.

5.2.2. *Relative robustness of forecasting methods when* $\lambda = 10$

Equation (8) provides a Bayes estimate of the flow time when the flow time follows an exponential distribution. However, when the exponential distribution assumption is not satisfied, Equation (8) is no longer the minimum loss estimator. We are interested in investigating the performance of the other methods under such conditions. The results of a comparative simulation, when the flow time of each class

Table 4. Pair-wise sign test for $\lambda = 10$; H_0: Median (Col) = Median (Row) vs. H_a: Median (Col) > Median (Row)

	CRDII	*CRDI*	*CAVG*	*EB-A*	*MME*	*MLE*	*EB-B*	*OAVG*
CRDII		0	0	0	0	0	0	0
CRDI	1		0	0	0	0	0	0
CAVG	1	1		0	0	0	0	0
EB-A	1	1	1		0	0	0	0
MME	1	1	1	1		0	0	0
MLE	1	1	1	1	1		0	0
EB-B	1	1	1	1	1	1		0
OAVG	1	1	1	1	1	1	1	

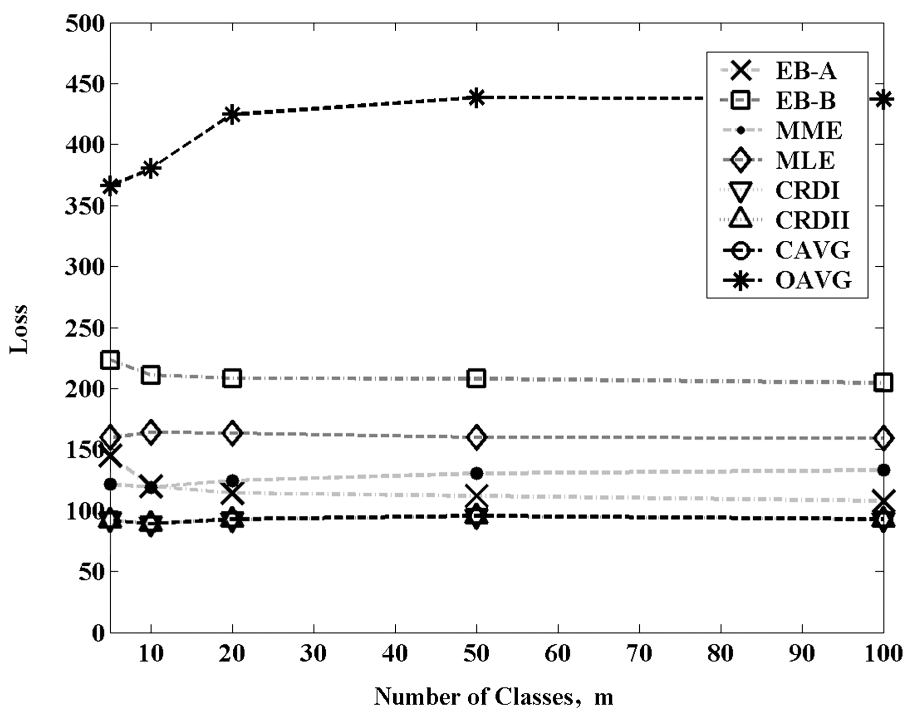

Fig. 8. Average losses against m for the $\lambda = 10$ model.

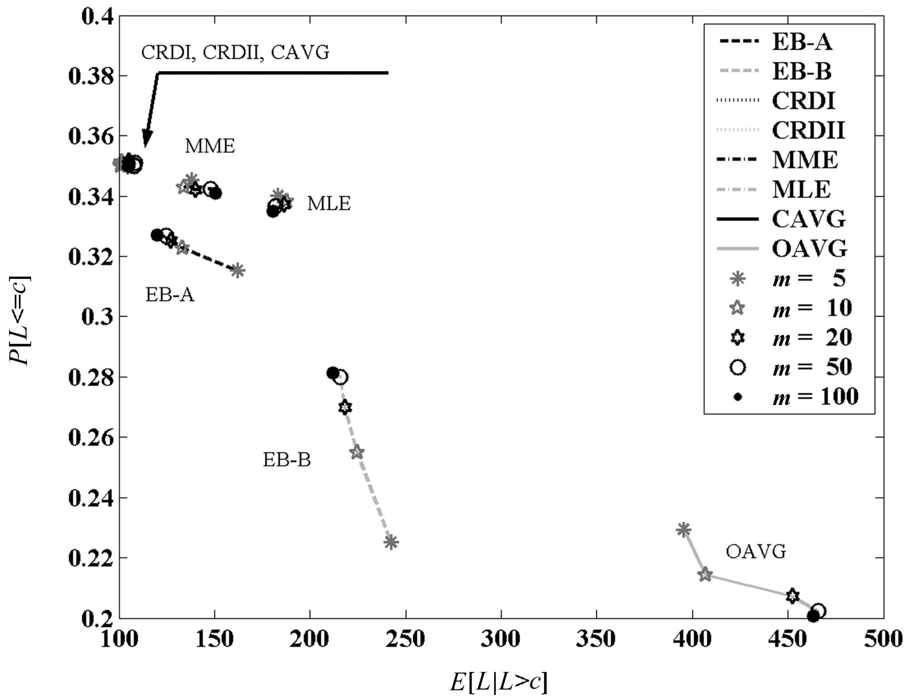

Fig. 9. Probability-loss graph averaged on n, α, and β, for the various methods in medium performance cluster for the $\lambda = 10$ model when $c = 1$.

follows a gamma distribution with parameters $\lambda = 10$ and θ_i, are shown in Figs. 8 and 9 and listed in Table 4.

Figure 8 shows the average loss as a function of m. Methods CRDI, CRDII and CAVG have the smallest loss. Figure 9 depicts the probability-loss graph, and shows that these three methods also dominate the other methods in terms of our probability-loss measure. Table 4 summarizes the probability of type-I errors for the hypotheses of Equations (30) and (31), where $M1$ is a column method and $M2$ is a row method. This table shows the pair-wise significance among the methods' median of losses, where the methods are ordered based on their relative performance. This table shows that CRDII performs better than the other empirical forecasting methods in this study.

According to the results of this experiment, the credibility methods (CRDI and CRDII) are robust and dominate other methods in both experiments, whereas the empirical Bayes methods (EB-A, EB-B, MME and one-step MLE) are flimsy with respect to the change of distribution. The fact that credibility forecasts do not rely on the distribution form of the flow times justifies this outcome. The results also suggest that a careless use of complex models may result in a poor forecast quality, which can even be of a poorer quality than simple models such as CAVG can provide.

6. Conclusions

We addressed the problem of forecasting flow times of multi-class jobs in a system where limited prior flow time observations for each class were available. Our objective was to use different Bayesian forecasting methods to directly forecast job flow times when minimizing loss as defined by the sum of squared errors. The methods were grouped according to the type and availability of information on the parameters of the problem assumptions, from which optimal forecasts were determined.

We argued that the combined effects of controllable and uncontrollable factors on the flow time result in a random increase above the minimal flow time for each job class. We assumed this random increase follows an exponential distribution with its parameter following a gamma distribution, the standard conjugate of exponential distribution. This assumption is very reasonable when modeling practical situations, and it provided us with some mathematical tractability in our derivations. However, when the mathematical tractability provided by the standard conjugate does not exist, Markov chain Monte Carlo simulation (see for example Brooks (1998)) may be used as a general procedure for estimating flow times.

A simulation study was conducted to compare the performances of these methods when model assumptions are satisfied or violated. Our simulation study showed that ignoring class structure by combining all observations into one class provides the worst forecasts. It also showed that when the class structure is considered, the choice of estimators employed by different methods had an impact on the quality of the forecast. For example, methods EB-A and EB-B, which differ in the choice of estimators, did not provide the same quality of forecasts.

Similar observations were made for methods CRDI and CRDII.

In practice, the degree to which the underlying theoretical assumptions of a forecasting method are satisfied changes over its application life. Thus, it behooves the practitioners to select a forecasting method that is simple yet robust. Our simulation results show that when both simplicity of the forecasting method and the quality of its forecast are important, using credibility (CRDI and CRDII) methods ranks high on the list of candidate forecasting methods. The results also show that complex models with inaccurate assumptions can produce forecasts with lower quality than those provided by simple models such as CAVG.

Acknowledgements

The authors acknowledge the support of NSF grant ESP-0091900, and appreciate the constructive comments and suggestions of Professor Mashayekhi and an anonymous referee.

References

Adam, N.R., Bertrand, J.W.M., Morehead, D.C. and Surkis, J. (1993) Due date assignment procedures with dynamically updated coefficients for multi-level assembly job shops. *European Journal of Operational Research*, **68**, 212–227.

Bagchi, U., Julien, F.M. and Magazine, M.J. (1994) Note: Due-date assignment to multi-job customer orders. *Management Science*, **40**(10), 1389–1392.

Box, G.E.P. and Jenkins, G.M. (1970) *Time Series Analysis; Forecasting and Control*, Holden-Day, San Francisco, CA.

Brooks, S.P. (1998) Markov chain Monte Carlo method and its application. *The Statistician*, **47**(1), 69–100.

Buhlmann, H. (1976) Experience rating and credibility. *ASTIN Bulletin*, **4**, 199–207.

Chen, C.Y.J., George, E.I. and Tardif, V. (2001) A Bayesian model of cycle time prediction. *IIE Transactions*, **33**, 921–930.

Cheng, T.C.E. and Gupta, M.C. (1989) Survey of scheduling research involving due date determination decisions. *European Journal of Operational Research*, **38**, 156–166.

De, P., Ghosh, J.B. and Wells, C.E. (1994) Due-date assignment and early/tardy scheduling on identical parallel machines. *Naval Research Logistics*, **41**, 17–32.

Ghosh, M. and Lahiri, P. (1987) Robust empirical Bayes estimation of means from stratified samples. *Journal of the American Statistics of Association*, **82**, 1153–1162.

Ghosh, M. and Meeden, G. (1986) Empirical Bayes estimation in finite population sampling. *Journal of the American Statistical of Association*, **81**, 1058–1062.

Hopp, W.J. and Roof Sturgis, M.L. (2000) Quoting manufacturing due dates subject to a service level constraint. *IIE Transactions*, **32**, 771-784.

Kaplan, A.C. and Unal, A.T. (1993) A probabilistic cost-based due date assignment model for job shops. *International Journal of Production Research*, **31**(12), 2817–2834.

Klugman, S.A., Panjer, H.H. and Wilmot, G.E. (1998) *Loss models: From Data to Decisions*, Wiley, New York.

Lahiri, P. (ed.) (2001) Model Selection, Institute of Mathematical Statistics, Lecture Notes-Monograph Series, **38**, Beach Wood, Ohio.

Luss, H. and Rosenwein, M.B. (1993) A due date assignment algorithm for multiproduct manufacturing facilities. *European Journal of Operational Research*, **65**, 187–198.

Mashayekhi, M., Abdoli, M. and Choobineh, F. (2004) Parametric empirical Bayes forecasting of waiting times under a non-symmetric loss. Working paper, Department of Industrial and Management Systems Engineering, University of Nebraska-Lincoln, NE 68588.

Mood, A.M., Graybill, F.A. and Boes, D.C. (1974) *Introduction to the Theory of Statistics*. 3rd edn., McGraw-Hill, New York.

Ruben, R.A. and Mahmoodi, F. (2000) Lead time prediction in unbalanced production systems. *International Journal of Production Research*, **38**(7), 1711–1729.

Seidmann, A. and Smith, M.L. (1981) Due date assignment for production systems. *Management Science*, **27**(5), 571–581.

Smith, C.H., Minor, E.D. and Wen, H.J. (1995) Regression-based due date assignment rules for improved assembly shop performance. *International Journal of Production Research*, **33**(9), 2375–2385.

Sundt, B. (1983) Parameter estimation in some credibility models. *Scandinavian Actuarial Journal*, 239–255.

Thiruvaiyaru, D. and Basawa, I.V. (1992) Empirical Bayes estimation for queuing systems and networks. *Queuing Systems*, **11** 179–202.

Appendix

Proof of Theorem 1. Let $D(m) = E[(X_i - X_i^{\text{EB}})^2] - E[(X_i - X_i^{\text{B}})^2]$. Notice that since the conditional expectation is a projection in the L_2-space, the following equality holds:

$$E\big[\big(X_i - X_i^{\text{EB}}\big)^2\big] - E\big[\big(X_i - X_i^{\text{B}}\big)^2\big] = E\big[\big(X_i^{\text{B}} - X_i^{\text{EB}}\big)^2\big].$$

The reason for this is that the above equality is a consequence of Pythagoras' theorem. A more detailed proof of the above equality is as follows:

$$\begin{aligned} D(m) &= E\big[\big(X_i - X_i^{\text{EB}}\big)^2 - \big(X_i - X_i^{\text{B}}\big)^2\big],\\ &= E\big[(X_i - X_i^{\text{B}} + X_i^{\text{B}} - X_i^{\text{EB}}\big)^2 - \big(X_i - X_i^{\text{B}}\big)^2\big],\\ &= E\big[\big(X_i - X_i^{\text{B}}\big)^2 + \big(X_i^{\text{B}} - X_i^{\text{EB}}\big)^2\\ &\quad + 2\big(X_i - X_i^{\text{B}}\big)\big(X_i^{\text{B}} - X_i^{\text{EB}}\big) - \big(X_i - X_i^{\text{B}}\big)^2\big],\\ &= E\big[\big(X_i^{\text{B}} - X_i^{\text{EB}}\big)^2\big] + 2E\big[\big(X_i - X_i^{\text{B}}\big)\big(X_i^{\text{B}} - X_i^{\text{EB}}\big)\big],\\ &= E\big[\big(X_i^{\text{B}} - X_i^{\text{EB}}\big)^2\big], \end{aligned}$$

where the last equality follows from the fact that $E[(X_i - X_i^{\text{B}})(X_i^{\text{B}} - X_i^{\text{EB}})] = 0$ as shown below:

$$\begin{aligned} &E\big[\big(X_i - X_i^{\text{B}}\big)\big(X_i^{\text{B}} - X_i^{\text{EB}}\big)\big]\\ &\quad = EE\big[\big(X_i - X_i^{\text{B}}\big)\big(X_i^{\text{B}} - X_i^{\text{EB}}\big) \mid \text{all } X_{ij}'\text{s}\big]\\ &\quad = E\big[E\big[\big(X_i - X_i^{\text{B}}\big) \mid \text{all } X_{ij}\text{s}\big]\big(X_i^{\text{B}} - X_i^{\text{EB}}\big)\big] = 0, \end{aligned}$$

since $(X_i^{\text{B}} - X_i^{\text{EB}})$ is a function of the X_{ij}s, and

$$\begin{aligned} E\big[\big(X_i - X_i^{\text{B}}\big) \mid \text{ all } X_{ij}\text{s}\big] &= E[X_i \mid \text{ all } X_{ij}\text{s}] - X_i^{\text{B}}\\ &= E[X_i \mid X_{ij} : j = 1, \ldots, n_i] - X_i^{\text{B}}\\ &= X_i^{\text{B}} - X_i^{\text{B}} = 0. \end{aligned}$$

Moreover, we have that:

$$
\begin{aligned}
X_i^{\mathrm{B}} - X_i^{\mathrm{EB}} &= \left(\beta + \sum_{j=1}^{n_i} X_{ij}\right)(\alpha + n_i - 1)^{-1} \\
&\quad - \left(\hat{\beta} + \sum_{j=1}^{n_i} X_{ij}\right)(\hat{\alpha} + n_i - 1)^{-1}, \\
&= \frac{\beta(\hat{\alpha} - \alpha) + \alpha(\beta - \hat{\beta})}{(\alpha + n_i - 1)(\hat{\alpha} + n_i - 1)} \\
&= \frac{\beta\hat{\alpha} - \alpha\hat{\beta}}{(\alpha + n_i - 1)(\hat{\alpha} + n_i - 1)},
\end{aligned}
$$

which means that $(X_i^{\mathrm{B}} - X_i^{\mathrm{EB}}) \xrightarrow{p} 0$ as $m \to \infty$, since by assumptions $\hat{\alpha} \xrightarrow{p} \alpha$ and $\hat{\beta} \xrightarrow{p} \beta$. Consequently, $(X_i^{\mathrm{B}} - X_i^{\mathrm{EB}})^2 \xrightarrow{p} 0$ as $m \to \infty$, because convergence in probability is preserved under continuous transformation. Also note that:

$$
\begin{aligned}
\left|X_i^{\mathrm{B}} - X_i^{\mathrm{EB}}\right| &= \left|\frac{\beta\hat{\alpha} - \alpha\hat{\beta}}{(\alpha + n_i - 1)(\hat{\alpha} + n_i - 1)}\right| \\
&\leq \frac{\beta\hat{\alpha} + \alpha\hat{\beta}}{(\alpha + n_i - 1)(\hat{\alpha} + n_i - 1)} \leq \frac{\beta N_1 + \alpha N_2}{(n_i - 1)^2}.
\end{aligned}
$$

Therefore, $D(m) = E[(X_i^{\mathrm{B}} - X_i^{\mathrm{EB}})^2] \to 0$ as $m \to \infty$ by the bounded convergence theorem, since $(X_i^{\mathrm{B}} - X_i^{\mathrm{EB}})^2 \xrightarrow{p} 0$ and $|X_i^{\mathrm{B}} - X_i^{\mathrm{EB}}|$ is bounded.

Biographies

Mansour Abdoli is a Ph.D. candidate in the Department of Industrial and Management Systems Engineering at the University of Nebraska—Lincoln where he received his M.S. in Industrial Engineering in 2002. He also holds a MS degree in Electrical Engineering from the Iran University of Science and Technology and a B.S. in Electrical Engineering from Sharif University of Technology. His research interests include stochastic modeling of manufacturing systems and e-commerce applications such as virtual sales agents. He has 4 years of experience in design and implementation of electrical systems (drive and control), and 3 years of experience in analysis and design of database information systems. He is a member of IIE and INFORMS.

F. Fred Choobineh is a Professor of Industrial and Management Systems Engineering and the Milton E. Mohr Distinguished Professor of Engineering at the University of Nebraska—Lincoln where he also holds a courtesy appointment as a Professor of Management. He received his B.S.E.E., M.S.I.E, and Ph.D. from Iowa State University. His research interests are in productivity improvement of manufacturing systems and the use of approximate reasoning methods such as fuzzy logic and rough sets in industrial decision models. He is a Fellow of the Institute of Industrial Engineers, a member of IEEE and IIE, a registered Professional Engineer in Nebraska and serves on the state licensing Board for Engineers and Architects.

Contributed by the Engineering Statistics and Applied Probability Department