European Journal of Operational Research 125 (2000) 205–215

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/dsw

Theory and Methodology

Airline spill analysis – beyond the normal demand

Michael Z.F. Li [a], Tae Hoon Oum [b,*]

[a] *Nanyang Business School, Nanyang Technological University, Nanyang Avenue, Singapore 639798, Singapore*
[b] *Faculty of Commerce and Business Administration, The University of British Columbia, 2053 Main Mall, Vancouver, BC, Canada V6J 1Z2*

Received 1 July 1998; accepted 1 March 1999

Abstract

Most research on airline passenger spill has assumed a normal distribution for the nominal demand. But empirical evidences show that normal distribution is not suitable in many cases, especially for spill analysis of business and first class compartments. In this paper, we derive formulae for calculating the expected number of spilled passengers when the nominal demand is assumed to follow a normal, a logistic, a log-normal, or a gamma distribution. The spills under these alternative distributional assumptions are compared numerically. Finally, the paper demonstrates that, for each of the four distributions, one can construct a generic observed load factor (OLF) table, which does not depend on aircraft seating capacity. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Yield management; Passenger spill modeling; Forecasting

1. Introduction

Spill models predict average lost sales when demand exceeds the flight capacity, and have been used by the airline industry since mid-1970s (Schlifer and Vardi, 1975). These models provide critical information for decisions in selecting aircraft size for a particular market, assigning an airline's existing fleet to various markets and schedules, and planning the aircraft type for future expansion. They are an integral part of modern yield management system. The basic idea behind a spill model is that demand for a group of flights can be represented by a probability distribution. The group of flights can be defined according to the wishes of the analyst or the airline. For example, it can involve just one flight segment, a small group of segments served by single or multiple aircraft types or all the segments served by a single fleet type.

Airline passenger spill analysis has traditionally relied on the normal distribution assumption for the nominal demand. But empirical evidences show that normal distribution does not fit very well in many cases, especially for business and first class compartments (Swan, 1992). In this paper,

* Corresponding author. Tel.: +1-604-822-8320; fax: +1-604-822-8521.

E-mail addresses: zfli@ntu.edu.sg (M.Z.F. Li), tae.oum@commerce.ubc.ca (T.H. Oum).

0377-2217/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S 0 3 7 7 - 2 2 1 7 (9 9) 0 0 1 9 4 - 0

we will address the airline spill problem by going beyond the normally distributed demand. In particular, we will derive formulae for calculating the expected number of spilled passengers when the nominal demand is assumed to follow a normal, a logistic, a log-normal or a gamma distribution. The spill rates under various distributional assumptions will then be compared numerically according to different variability levels of the demand. Finally, the paper will demonstrate that, for each of the four distributions, one can construct a generic observed load factor (OLF) table, which does not depend on seating capacity of the aircraft.

The paper is organized as follows. Section 2 develops the analytical foundation of the airline spill analysis by first establishing an important distribution-free identity relationship among the observed load factor, the nominal load factor and the spill. The main focus of this section is to extend the traditional wisdom of normal distribution assumption for the nominal demand by deriving the spill calculation formulae for three other probability distribution functions, namely, logistic, log-normal, and gamma distribution. Section 3 starts with the parameter conversions and the role of coefficient of variation (CV) in the shape of the demand distributions. We will present a few numerical comparisons on spill calculations under the four distributions and further establish a result on the possibility of constructing a generic OLF table. The last section gives the conclusion.

2. Analytical foundation of spill analysis

2.1. Definitions and notation

Let X be the random demand for a fare class, a flight, or a group of flights; and C the seating capacity level under consideration. X is the true demand for the flight, and thus is frequently called as the *nominal demand* to be distinguished from the *observed demand* which is the number of seats actually sold. During the booking process, once the number of booking requests exceeds the capacity, the airline refuses any additional booking. The passengers who are rejected in the booking process are called *spilled passengers*, or simply *the spill* as commonly used by airlines. However, the airline can not be certain about the exact number of spilled passengers since the demand is random. Also, the random demand is truncated by capacity, making it difficult to examine the true empirical distribution of the demand. On the other hand, for spill analysis and demand forecasting, it is extremely important to accurately identify the true (nominal) demand distribution from observable demand data.

Let us first introduce several concepts that will standardize our presentation.

Definition 1. Let X be the nominal demand for a flight with capacity of C (could the effective capacity, see Swan (1992)). Then

1. Nominal load factor (NLF) is defined as $E(X)/C$.
2. Observed mean load is the expected value of the nominal demand X truncated at the capacity level C, that is, $E(\min(X, C))$; and the mean observed load factor (OLF) is defined as $E(\min(X, C))/C$.
3. The fill rate (FR) for the pth seat is defined as the probability that the demand is equal to or greater than p, that is, $P(X \geqslant p)$.
4. The spilled passengers (SP) is the number of passengers turned away because the flight is fully booked, that is, $\mathrm{SP} = E[(X - C)I_{(X>C)}]$, where $I_{(X>C)}$ is the indicator function defined as $I_{(X>C)}(x) = 1$ if $x > C$ and 0 if $x \leq C$.
5. The spill rate (SR) is defined as the ratio of spilled passengers over the mean of the nominal demand, that is, $\mathrm{SR} = \mathrm{SP}/E(X) = E[(X - C)I_{(X>C)}]/E(X)$.

Before proceeding further, it is necessary for us to establish a simple but useful relationship between the nominal load factor (NLF), the observed load factor (OLF), and the spilled passengers (SP) in the following lemma.

Lemma 2. *It is always true that* $\mathrm{OLF} = \mathrm{NLF} - \mathrm{SP}/C$ *and consequently*, $\mathrm{OLF} = (1 - \mathrm{SR}) \times \mathrm{NLF}$.

Proof. It is straightforward and left to the reader.

2.2. Spill formula for some common distributions

It is well-known that the traditional spill analysis has focused mainly on normal distributions, and occasionally gamma distributions. On the other hand, empirical distributions for the observed demands are usually more diverse. Therefore, it is interesting and necessary to explore other types of probability distribution for the demand. According to Swan (1992), demand for the first class cabin is usually neither a normal nor a gamma distribution.

In this section, analytical formulae for the expected spill and the spill rate are derived for four alternative demand distributions.

(A) *Normal distribution*: Let X follow a normal distribution $\mathrm{N}(\mu, \sigma^2)$ with the mean μ and the variance σ^2. Let $\phi(x)$ and $\Phi(x)$ be the corresponding probability density function and the cumulative probability distribution function for the standard normal N(0, 1), respectively. Then the expected number of spilled passengers is given by

$$\begin{aligned}
\mathrm{SP} &= E[(X-C)I_{(X>C)}] = 1/(\sqrt{2\pi}\sigma) \\
&\quad \int_C^{\infty} (x-C)\exp\{-(x-\mu)^2/(2\sigma^2)\}\,\mathrm{d}x \\
&= \sigma\int_b^{\infty}(t-b)\phi(t)\,\mathrm{d}t = (\sigma/\sqrt{2\pi}) \\
&\quad \int_b^{\infty} t\exp(-t^2/2)\,\mathrm{d}t - \sigma b\int_b^{\infty}\phi(t)\,\mathrm{d}t \\
&= \sigma(1/\sqrt{2\pi})\big[(-\exp(-t^2/2)\big]_{t=b}^{t=\infty} \\
&\quad - \sigma b(1-\Phi(b)) \\
&= \sigma[\phi(b) - b(1-\Phi(b))],
\end{aligned}$$

where $b = (C-\mu)/\sigma$, commonly known as the *buffer*. Consequently, the spill rate for the normal distribution becomes

$$\mathrm{SR} = \mathrm{SP}/\mu = \mathrm{CV}\times[\phi(b) - b(1-\phi(b))],$$

where $\mathrm{CV} = \sigma/\mu$ is the *coefficient of variation*, measuring the variability of a probability distribution.

The above formula first appeared in Shlifer and Vardi (1975), but in a different context. Swan (1983) was the first to derive the formula in the context of spill analysis.

(B) *Logistic distribution*: The logistic distribution has been frequently used in airline spill analysis. There are mainly two reasons for the popularity: (a) logistic distribution provides a reasonable approximation to normal distribution; and (b) a simple formula for spill calculation can be derived using the logistic distribution. Also, the logistic approximation could be calibrated to fit nicely between the results of a normal distribution and a gamma distribution within the relevant ranges of spill calculation (Swan, 1992). This subsection provides a closed-form expression for the spill formula when a general logistic distribution is used to model the nominal demand.

A random variable X is said to have a logistic distribution with parameters θ and β if it has the following probability density function:

$$f(x) = \frac{\exp\{(x-\theta)/\beta\}}{\beta[1+\exp\{(x-\theta)/\beta\}]^2} \quad \text{for } -\infty < x < \infty,$$

where θ is the location parameter such that $-\infty < \theta < \infty$ and β is a positive scaling parameter. We will denote this distribution by $L(\theta, \beta)$. It is easy to show that a $L(\theta, \beta)$ distribution is symmetric around θ and is bell-shaped similar to a normal distribution. The main difference between the two is that a logistic density function has relatively longer tails and is more peaked in the center than a normal density function.

The mean and the variance of $L(\theta, \beta)$ are given by $E(X) = \theta$ and $\mathrm{Var}(X) = \beta^2\pi^2/3$, respectively. This implies that $\beta = \sqrt{3}\sigma_X/\pi$. Now for the spill, we have

$$\begin{aligned}
\mathrm{SP} &= E[(X-C)I_{(X>C)}] \\
&= \int_C^{\infty}(x-C)f(x)\,\mathrm{d}x \\
&= \int_C^{\infty}(x-C)\frac{\mathrm{e}^{(x-\theta)/\beta}}{\beta(1+\mathrm{e}^{(x-\theta)/\beta})^2}\,\mathrm{d}x \\
&= \int_{(C-\theta)/\beta}^{\infty}[\beta y + \theta - C]\frac{\mathrm{e}^{y}}{(1+\mathrm{e}^{y})^2}\,\mathrm{d}y \\
&\quad \text{(by taking } y = (x-\theta)/\beta)
\end{aligned}$$

$$
\begin{aligned}
&= \beta \int_c^{\infty} (y-c) \frac{e^y}{(1+e^y)^2}\, dy \\
&= \beta \left[(y-c)\left(-\frac{1}{1+e^y} \right) \right]_{y=c}^{y=\infty} \\
&\quad + \beta \int_c^{\infty} \frac{dy}{1+e^y} \quad \left(c \equiv \frac{C-\theta}{\beta} \right) \\
&= 0 + \beta \int_c^{\infty} \frac{e^{-y}}{1+e^{-y}}\, dy \\
&= \beta[-\ln(1+e^{-y})]_{y=c}^{y=\infty} \\
&= \beta \ln(1+e^{-c}),
\end{aligned}
$$

i.e.,

$$
\begin{aligned}
\mathrm{SP} &= \beta \ln(1+\exp(-c)) \\
&= \beta \ln(1+\exp(-(C-\theta)/\beta)).
\end{aligned}
$$

Consequently, the spill rate is

$$
\mathrm{SR} = (\beta/\theta)\ln(1+\exp(-(C-\theta)/\beta)).
$$

The above is a new spill formula derived explicitly from the logistic distribution. In the past, researchers used logistic distribution in spill calculation as an approximation for the normal distribution as was done in Swan (1983). For the purpose of comparison, let us address this issue in more details. Recall that, in deriving the spill formula for the normal distribution, we are lead to

$$
\mathrm{SP} = \sigma \int_b^{\infty} (t-b)\phi(t)\, dt
$$

where $b = (C-\mu)/\sigma$ is the buffer. Instead of using the exact normal distribution function and derive the spill formula, one can use the following logistic density function as an approximation of the normal density function $\phi(t)$:

$$
f_w(t) = \frac{w e^{wt}}{(1+e^{wt})^2},
$$

where $w = 1.7$. This approximation leads to the following formula for SP:

$$
\begin{aligned}
\mathrm{SP} &= \sigma \int_b^{\infty} (t-b) \frac{w e^{wt}}{(1+e^{wt})^2}\, dt \\
&= \sigma \left[(t-b)\left(-\frac{1}{1+e^{wt}} \right) \right]_{t=b}^{t=\infty} \\
&\quad + \sigma \int_b^{\infty} \frac{1}{1+e^{wt}}\, dt \\
&= 0 + \frac{\sigma}{w} \int_{\mathrm{wb}}^{\infty} \frac{1}{1+e^{v}}\, dv \\
&= \frac{\sigma}{w} \ln(1+e^{-\mathrm{wb}}).
\end{aligned}
$$

Then the corresponding formula for the spill rate under this approximation becomes

$$
\mathrm{SR} = \frac{\sigma}{w\mu} \ln(1+e^{-wb}) = \frac{\mathrm{CV}}{1.7} \ln(1+e^{-1.7b}). \tag{1}
$$

For comparison, rewriting the original spill formula under the logistic distribution in terms of the buffer b gives us the following:

$$
\begin{aligned}
\mathrm{SR} &= (\beta/\theta)\ln(1+\exp(-(C-\theta)/\beta)) \\
&= \frac{\sigma}{\mu\pi/\sqrt{3}} \ln\left(1 + \exp\left(-\frac{\pi}{\sqrt{3}} \frac{C-\mu}{\sigma} \right) \right) \\
&\approx \frac{\mathrm{CV}}{1.8138} \ln(1+e^{-1.8138b}).
\end{aligned} \tag{2}
$$

Eq. (2) is different from the spill rate formula in (1). The difference between (1) and (2) is mainly due to the difference in distributional assumption of the nominal demand. In fact, according to Johnson et al. (1995, p. 119), when $w = 1.7017456$ logistic distribution provides the best approximation to the standard normal distribution.

(C) *Log-normal distribution*: Now assume that the nominal demand X follows a log-normal distribution, which implies that there exist some constants γ, δ and θ such that $U = \gamma + \delta \ln(X-\theta)$ follows the standard normal distribution N(0, 1). Then, it can be shown that the probability density function of X is given by

$$
f(x) = \frac{\delta}{\sqrt{2\pi}(x-\theta)} \exp\left(-\frac{(\gamma+\delta\ln(x-\theta))^2}{2} \right), \quad x > \theta.
$$

For our purpose, we assume θ equal to 0. Thus, the rth moment of X is given by

$$E(X^r) = \exp r\mu + (r\sigma)^2/2,$$

where $\mu \equiv -\gamma/\delta = E(\ln(X))$ and $\sigma^2 \equiv 1/\delta^2 = \mathrm{Var}(\ln(X))$. Consequently, the mean and the variance of X are given by:

$$E(X) = e^{\mu} e^{(1/2)\sigma^2} \text{ and } \mathrm{Var}(X) = e^{2\mu} e^{\sigma^2}(e^{\sigma^2} - 1).$$

When $\theta = 0$, the density function of the log-normal distribution can be rewritten as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right\}, \quad \text{for } x > 0.$$

Let us now derive the formula for the spill under a log-normal distribution:

$$\begin{aligned}
\mathrm{SP} &= E[(X - C)I_{(X>C)}] \\
&= \int_C^\infty (x - C) f(x)\,dx \\
&= \int_C^\infty \frac{x - C}{\sqrt{2\pi}\sigma x} \exp -\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2 dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_C^\infty \exp -\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2 dx \\
&\quad - \frac{C}{\sqrt{2\pi}\sigma} \int_C^\infty \exp -\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2 \frac{dx}{x} \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_c^\infty e^{-(t^2)/2} \sigma e^{\sigma t + \mu}\,dt \\
&\quad - \frac{C}{\sqrt{2\pi}\sigma} \int_c^\infty e^{-(t^2)/2} \sigma\,dt \\
&\quad \left(\text{by taking } t = \frac{\ln x - \mu}{\sigma}\right) \\
&= \frac{1}{\sqrt{2\pi}} \int_c^\infty \exp -\frac{t^2 - 2\sigma t - 2\mu}{2}\,dt \\
&\quad - C(1 - \Phi(c)) \\
&= \frac{e^{\mu + \sigma^2/2}}{\sqrt{2\pi}} \int_c^\infty e^{-((t-\sigma)^2)/2}\,dt - C(1 - \Phi(c)) \\
&= e^{\mu + \sigma^2/2}(1 - \Phi(c - \sigma)) - C(1 - \Phi(c)),
\end{aligned}$$

where $c = (\ln C - \mu)/\sigma$. Consequently, the spill rate is given by

$$\mathrm{SR} = \frac{\mathrm{SP}}{E(X)} = (1 - \Phi(c - \sigma)) - \frac{C}{e^{\mu + \sigma^2/2}}(1 - \Phi(c)).$$

(D) *Gamma distribution*: According to Swan (1983), there are three main reasons why a gamma distribution is attractive for modeling the nominal demand. First, the guarantee of non-negative demand is more realistic than a normal distribution. Second, a gamma distribution appears to be closer to the shape of the observed load distribution, especially for some flights on high demand days since they generally have fatter positive tails than the normal distributions. Finally, the Bernoulli trial component variation is a gamma distribution.

But interestingly, there has not been any formal development of the spill analysis using gamma distributions. Recall that the probability density function of a standard two-parameter gamma distribution is

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad \text{for } x \geqslant 0 \text{ with}$$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\,dx.$$

In this two-parameter gamma distribution, α is the shape parameter and β the scale parameter. It is easy to show that the moment generating function of the Gamma distribution is given by $m(t) = E(\exp(tX)) = (1 - \beta t)^{-\alpha}$. And the corresponding mean and variance are $E(X) = \alpha\beta$ and $\mathrm{Var}(X) = \alpha\beta^2$, respectively. Hence, $\mathrm{CV} = 1/\sqrt{\alpha}$. Now let us derive the spill:

$$\begin{aligned}
\mathrm{SP} = E((X - C)I_{(X>C)}) &= \int_C^\infty (x - C) f(x)\,dx \\
&= \int_C^\infty (x - C)\frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}\,dx \\
&= \int_C^\infty \frac{x^{\alpha} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}\,dx - C\int_C^\infty \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}\,dx \\
&= \alpha\beta[1 - G(C, \alpha + 1, \beta)] - C[1 - G(C, \alpha, \beta)],
\end{aligned}$$

where $G(x, \alpha, \beta)$ is the cumulative probability function of the Gamma distribution with the parameters (α, β):

$$G(x, \alpha, \beta) = \int_0^x \frac{t^{\alpha-1} e^{-t/\beta}}{\beta^\alpha \Gamma(\alpha)}\,dt.$$

Therefore, the spill rate is given by $\mathrm{SR} = [1 - G(C, \alpha+1, \beta)] - (C/\alpha\beta)[1 - G(C, \alpha, \beta)]$. Since the exponential distribution is a special case of Gamma distribution with $\alpha = 1$, it is easy to show that the spill rate under an exponential distribution is given by $\mathrm{SR} = \mathrm{e}^{-C/\beta}$. But in practice, the exponential distribution is rarely used to model the demand.

It is worth noting that spill calculation under a Gamma distribution is more numerically demanding than under other distributions, and using the CV and the buffer b in the spill calculation is possible, but not as natural as under other distributions.

3. Comparing spill values

3.1. Conversion of distributional parameters

From the discussions in the previous section, it is clear that there is no closed-form solution for spill calculation except under a logistic distribution. In the early days, a logistic approximation to the normal distribution was a natural and convenient choice because of its simplicity and the closed-form solution. Direct application of the normal or gamma distribution would certainly involve more numerical coding and become computationally more demanding. Dramatic changes have occurred in spill analysis as the spreadsheet software has become more sophisticated since the early 1990s. The purpose of this section is to perform a few numerical spill calculations and comparisons by using standard spreadsheet software.

Over the years, practitioners in the airline industry have become used to calculating the spill or spill rate by directly using the coefficient of variation (CV) and the expected demand $E(X)$ (i.e., μ_X). To be consistent with this practice, we now reparameterize the distributions and rewrite the corresponding spill formulae, as summarized in Table 1.

3.2. The shape of demand distribution and the value of CV

One important task in airline spill analysis is to accurately model the demand distribution. As pointed out earlier, Swan (1992) indicates that the normal distribution does not fit all situations. One of the key issues is that the shape of the demand distribution is skewed to the right for small cabins, implying a relatively large value of CV. Fig. 1 illustrates this point graphically, where all of the four distributions are assumed to have the same mean and CV values.

One can make a number of general observations from Fig. 1. First, for small CV values, all of the three non-normal distributions are close to the normal distribution, implying that there will not be any significant difference in spill calculations under the four distributional assumptions. Second, a normal distribution becomes increasingly inappropriate to model the nominal demand as the CV value increases. The spill calculated by assuming a normal distribution will clearly *overestimate* the true spill when the value of CV becomes large. Third, the difference in spill between a log-normal distribution and a Gamma distribution is surprisingly small for a large CV even though they behave quite differently for smaller values. Fourth, as the value of CV becomes larger, the Gamma distribution is getting closer to the shape of an exponential distribution, which is

Table 1
Reparameterization of demand distributions by using CV and μ_X

Distribution	Reparameterization	Spill formula (SP)
Normal	$\mu = \mu_X$; $\sigma = \mathrm{CV} \times \mu_X$	$\sigma[\phi(b) - b(1 - \Phi(b))]$, $b = (C - \mu_X)/\sigma_X$
Logistic	$\theta = \mu_X$; $\beta = \frac{\mathrm{CV} \times \mu_X}{\pi/\sqrt{3}}$	$\beta \ln(1 + \mathrm{e}^{-(C-\theta)/\beta})$
Log-normal	$\mu = \ln \frac{\mu_X}{\sqrt{1+\mathrm{CV}^2}}$; $\sigma = \sqrt{\ln(1 + \mathrm{CV}^2)}$	$\mu_X(1 - \Phi(c - \sigma)) - C(1 - \Phi(c))$, $c = \frac{\ln C - \mu}{\sigma}$
Gamma	$\alpha = \frac{1}{\mathrm{CV}^2}$; $\beta = \mathrm{CV}^2 \times \mu_x$	$\mu_X[1 - G(C, \alpha+1, \beta)] - C[1 - G(C, \alpha, \beta)]$

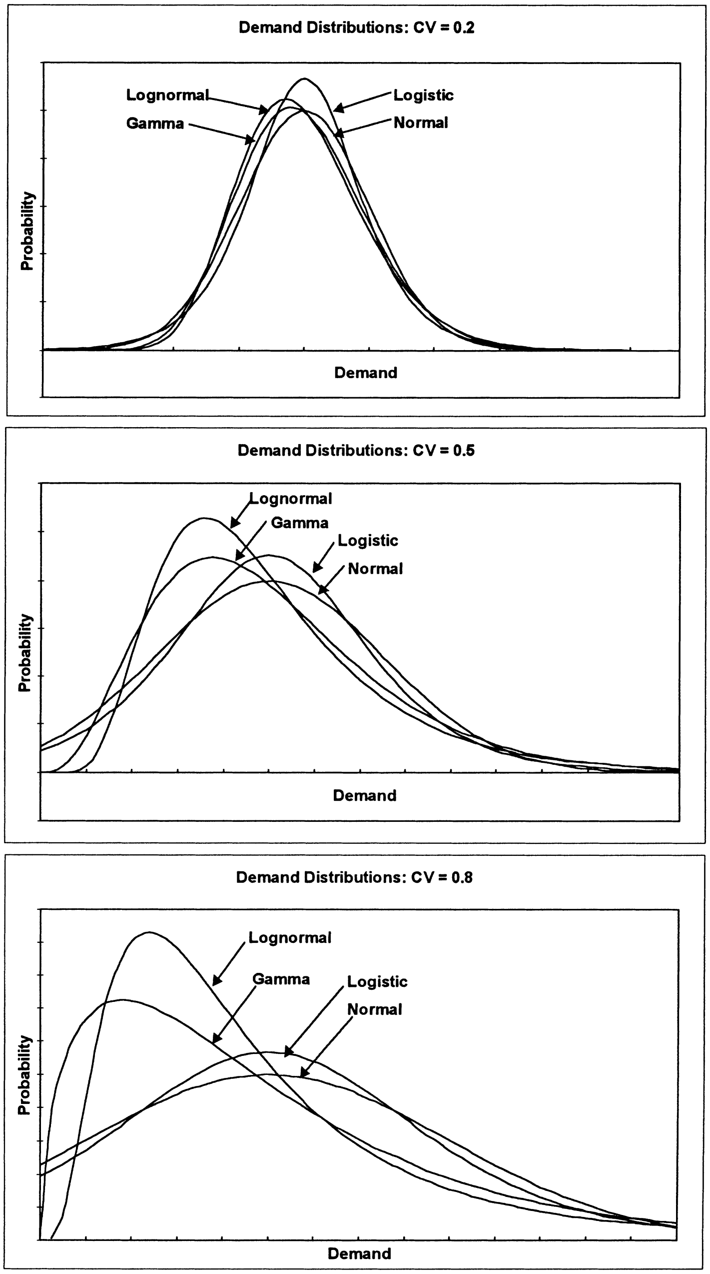

Fig. 1. The shape of demand distribution and the CV.

usually not a good shape for demand. On the other hand, the log-normal distribution behaves much more "robust" for the lower portion of the demand. Finally, for large values of CV, neither a normal nor a logistic distribution appears to be appropriate to model the nominal demand because these distributions will have a relatively high probability of taking a negative demand while a negative value is impossible under a log-normal or a gamma distribution.

3.3. Numerical comparisons of spill values

In this section, the spill values computed under the four alternative distributions are compared numerically. For each of the four distributions, we calculate the spill for two capacity levels: $C = 150$ and 30, and three CV values: 0.2, 0.5 and 0.8.

Tables 2 and 3 report spill values at capacity levels 150 and 30 seats per aircraft, respectively. Each of these tables are arranged so as to make it easy to compare spill values for the four alternative distributions at different levels of mean and coefficient of variation (CV) of the nominal demand.

From the above two tables, we can make the following general observations.

(1) It is clear that differences in spills among the four alternative distributions are quite small when CV = 0.2, which indicates that distributional assumption does not play a pivotal role in spill calculation when the demand is not very volatile, i.e., small CV value. This observation is consistent with the fact that, for small CV, all of the three non-normal distributions are close to a normal distribution.

(2) On the other hand, the differences in spill values between a normal distribution and a log-normal distribution become rather large for both

Table 2
Spill table – capacity = 150 [a]

	CV = 0.2				CV = 0.5				CV = 0.8			
μ_X	A	B	C	D	A	B	C	D	A	B	C	D
115	0.6	0.8	1.1	0.9	9.6	9.1	11.0	11.0	21.8	20.6	21.4	23.3
120	1.2	1.3	1.7	1.6	11.9	11.2	12.9	13.1	25.2	23.8	23.9	26.2
125	2.1	2.1	2.6	2.5	14.4	13.6	15.0	15.4	28.6	27.1	26.5	29.1
130	3.3	3.2	3.8	3.7	17.1	16.2	17.3	17.9	32.3	30.6	29.2	32.2
135	4.9	4.6	5.3	5.2	20.1	19.0	19.8	20.5	36.0	34.2	32.1	35.3
140	6.9	6.5	7.1	7.1	23.2	22.1	22.4	23.3	39.9	38.0	35.0	38.6
145	9.2	8.8	9.3	9.3	26.5	25.3	25.1	26.2	43.8	41.9	38.1	42.0
150	12.0	11.5	11.8	11.9	29.9	28.7	28.0	29.3	47.9	45.9	41.2	45.4
155	15.0	14.5	14.7	14.9	33.5	32.2	31.0	32.5	52.0	49.9	44.5	48.9
160	18.4	17.9	17.9	18.1	37.2	35.9	34.2	35.8	56.2	54.1	47.8	52.5
165	22.0	21.6	21.3	21.6	41.0	39.6	37.5	39.3	60.5	58.3	51.2	56.2
170	25.8	25.5	25.0	25.3	44.8	43.5	40.9	42.8	64.8	62.6	54.7	59.9

[a] A = Normal; B = Logistic; C = Log-normal; D = Gamma.

Table 3
Spill table – capacity = 30 [a]

	CV = 0.2				CV = 0.5				CV = 0.8			
μ_X	A	B	C	D	A	B	C	D	A	B	C	D
20	0.0	0.0	0.0	0.0	0.8	0.8	1.2	1.2	2.6	2.5	2.9	3.1
22	0.1	0.1	0.1	0.1	1.5	1.4	1.8	1.8	3.7	3.5	3.8	4.1
24	0.2	0.3	0.3	0.3	2.4	2.2	2.6	2.6	5.0	4.8	4.8	5.2
26	0.7	0.6	0.8	0.7	3.4	3.2	3.5	3.6	6.5	6.1	5.8	6.4
28	1.4	1.3	1.4	1.4	4.6	4.4	4.5	4.7	8.0	7.6	7.0	7.7
30	2.4	2.3	2.4	2.4	6.0	5.7	5.6	5.9	9.6	9.2	8.2	9.1
32	3.7	3.6	3.6	3.6	7.4	7.2	6.8	7.2	11.2	10.8	9.6	10.5
34	5.2	5.1	5.0	5.1	9.0	8.7	8.2	8.6	13.0	12.5	10.9	12.0
36	6.8	6.8	6.6	6.7	10.6	10.3	9.6	10.0	14.7	14.3	12.4	13.5
38	8.6	8.6	8.4	8.4	12.2	12.0	11.1	11.6	16.5	16.1	13.8	15.1
40	10.4	10.4	10.2	10.3	14.0	13.7	12.7	13.2	18.4	17.9	15.4	16.7
42	12.3	12.3	12.1	12.2	15.7	15.5	14.3	14.8	20.3	19.8	17.0	18.3

[a] A = Normal; B = Logistic; C = Log-normal; D = Gamma.

Table 4
Difference in spills between $C_1 = 130$ and $C_2 = 150$ [a]

	CV = 0.2				CV = 0.5				CV = 0.8			
μ_X	A	B	C	D	A	B	C	D	A	B	C	D
115	2.9	2.6	2.9	2.9	6.7	6.3	5.2	5.7	7.9	7.6	5.3	6.0
120	4.2	3.8	3.9	4.0	7.4	7.1	5.8	6.3	8.4	8.1	5.7	6.4
125	5.6	5.2	5.1	5.3	8.1	7.9	6.4	6.9	8.8	8.7	6.1	6.8
130	7.1	6.8	6.5	6.7	8.8	8.6	7.0	7.5	9.2	9.1	6.5	7.2
135	8.6	8.4	7.9	8.1	9.4	9.3	7.6	8.1	9.6	9.6	6.9	7.5
140	10.0	10.0	9.3	9.5	10.0	10.0	8.2	8.7	10.0	10.0	7.3	7.9
145	11.3	11.5	10.6	10.9	10.5	10.6	8.7	9.2	10.3	10.4	7.6	8.2
150	12.6	12.9	12.0	12.1	11.1	11.2	9.3	9.8	10.7	10.8	8.0	8.5
155	13.7	14.0	13.2	13.3	11.5	11.7	9.8	10.3	11.0	11.1	8.4	8.8
160	14.6	15.0	14.3	14.4	12.0	12.2	10.4	10.7	11.2	11.4	8.7	9.1
165	15.4	15.9	15.3	15.3	12.4	12.7	10.9	11.2	11.5	11.7	9.1	9.4
170	16.2	16.6	16.1	16.1	12.8	13.1	11.4	11.6	11.7	12.0	9.4	9.7

[a] A = Normal; B = Logistic; C = Log-normal; D = Gamma.

capacity levels, and increase with the value of CV. Therefore, the choice of a demand distribution becomes a much more serious issue when the demand is volatile. Furthermore, the capacity level plays virtually no role in deciding which distribution should be used to model the demand. This is contrary to the findings of other studies on first class or business class spill analysis (Swan, 1992).

At this juncture, it is important to reiterate that whenever a spill model is used, the decision variable usually is not the spilled demand volume itself, but the difference between the spill volumes for two alternative scenarios. Consider the decision whether to assign a 130-seat (C_1) or a 150-seat (C_2) aircraft to a flight leg. The relevant question is, how many extra passengers the extra seats will accommodate, and whether or not the extra revenue would cover the extra cost of using a larger aircraft? The answer involves evaluating the difference in spill between the two alternatives. For a single incremental seat, this is the *fill rate* for the flight. For 5 incremental seats, it is simplest to take the difference of the two spill calculations, which is summarized in Table 4.

Table 4, together with Tables 2 and 3, suggests that compared with the log-normal and Gamma distributions, using a normal distribution will not only over-estimate the spill at each capacity level, but also over-estimate the difference in spill when used to evaluate two alternative capacity levels. This over-estimation becomes increasingly serious as CV increases.

3.4. Generic OLF table

For practical reasons, it is often important to have information on the OLF associated with spill calculations. Before dealing with the numerical issues related to OLF, let us first establish the following surprising result.

Lemma 3. *If the demand follows a normal, a logistic, a log-normal or a gamma distribution, then the observed load factor (OLF) depends only on the nominal load factor (NLF) and the value of coefficient of variation (CV) of the distribution.*

Proof. By Lemma 1, we know that OLF = (1 − SR) × NLF. Therefore, to prove the Lemma 2, we only need to show that the spill rate (SR) under each of the four distributions can be expressed in terms of NLF($= \mu_X/C$) and CV. This is summarized in Table 5.

It is clear from Table 5 that the spill rate under each of the four distributions is a function of CV and NLF, as required.

An important consequence of this result is that it is possible to generate a generic OLF Table,

Table 5
Representation of spill rate by CV and NLF

Demand distribution	Spill rate formula
Normal	$\mathrm{CV} \times [\phi(b) - b(1 - \Phi(b)];\ b = (\frac{1}{\mathrm{NLF}} - 1)/\mathrm{CV}$
Logistic	$\frac{\pi}{\sqrt{3}}\frac{1}{\mathrm{CV}} \ln[1 + \exp(-\frac{\pi}{\sqrt{3}\mathrm{CV}}(\frac{1}{\mathrm{NLF}} - 1))]$
Log-normal	$(1 - \Phi(c - \sigma)) - \frac{1}{\mathrm{NLF}}(1 - \Phi(c));\quad c = \frac{\ln(\sqrt{1+\mathrm{CV}^2})/(\mathrm{NLF})}{\sqrt{\ln(1+\mathrm{CV}^2)}}, \sigma = \sqrt{\ln(1 + \mathrm{CV}^2)}$
Gamma	$[1 - G(\frac{C}{\beta}, \alpha + 1, 1)] - \frac{1}{\mathrm{NLF}}[1 - G(\frac{C}{\beta}, \alpha, 1)];\ \alpha = \frac{1}{\mathrm{CV}^2},\ \frac{C}{\beta} = \frac{1}{\mathrm{CV}^2 \times \mathrm{NLF}}$

Table 6
A generic OLF table[a]

	CV = 0.2				CV = 0.5				CV = 0.8			
NLF	A	B	C	D	A	B	C	D	A	B	C	D
0.667	0.666	0.666	0.665	0.666	0.639	0.639	0.626	0.628	0.580	0.585	0.569	0.563
0.733	0.731	0.730	0.729	0.730	0.683	0.685	0.672	0.673	0.609	0.616	0.607	0.596
0.800	0.792	0.791	0.789	0.790	0.721	0.725	0.714	0.713	0.632	0.641	0.641	0.626
0.867	0.845	0.846	0.841	0.842	0.752	0.759	0.751	0.748	0.652	0.663	0.672	0.652
0.933	0.887	0.890	0.886	0.886	0.779	0.786	0.784	0.778	0.668	0.680	0.700	0.676
1.000	0.920	0.924	0.921	0.920	0.801	0.809	0.813	0.805	0.681	0.694	0.725	0.697
1.067	0.944	0.947	0.948	0.946	0.819	0.828	0.839	0.828	0.692	0.706	0.748	0.716
1.133	0.961	0.963	0.966	0.964	0.834	0.843	0.861	0.848	0.701	0.716	0.769	0.734
1.200	0.973	0.974	0.979	0.977	0.847	0.856	0.880	0.865	0.709	0.724	0.788	0.750
1.267	0.981	0.981	0.987	0.985	0.859	0.866	0.896	0.880	0.715	0.730	0.805	0.764
1.333	0.987	0.986	0.993	0.991	0.868	0.875	0.911	0.894	0.721	0.736	0.821	0.777
1.400	0.990	0.989	0.996	0.994	0.876	0.883	0.923	0.905	0.725	0.740	0.835	0.789

[a] A = Normal; B = Logistic; C = Log-normal; D = Gamma.

which is not related to the capacity level. Table 6 below is such a generic OLF table.

This kind of tables is of great importance to practitioners in airline industry since only the value of OLF is observable. This implies that one can quickly obtain value of the unobservable NLF from a generic OLF table if the demand follows one of the four distributions considered in this paper. With the values of OLF and NLF, one can easily get the spill rate (SR) value by using Lemma 1.

4. Summary and conclusions

In this paper, we re-examined airline spill problem and went beyond the traditional assumption of normal distribution for the nominal demand. We first established a distribution-free multiplicative relationship between the observed load factor (OLF), the nominal load factor (NLF), and the spill rate (SR). In addition to deriving the spill formula for the normal distribution again, new spill formulae were derived under a logistic, a log-normal, and a gamma distribution. Furthermore, each of the four spill formulae is rewritten as a function of the mean demand and the coefficient of variation (CV), which are the common inputs used to calculate the spill.

In our numerical example, the spills under three different values of CV are calculated at two capacity levels for each of the four distributions. It is found that for relatively small values of CV, there is no significant difference in the value of the spills across all four alternative distributions. As the demand become less stable, or equivalently, more volatile, the use of a normal distribution becomes much more problematic because of the increasing probability that the demand will assume a negative value. The numerical examples show that the normal demand will not only over-estimate the spill at a given capacity level, but also over-estimate the difference in incremental spill when two capacity levels are considered.

This paper also demonstrate that, for each of the four distributions, one can create a generic OLF table which does not depend on aircraft seating capacity. This table can be very useful to practitioners in spill analysis as it allows them to infer the value of the nominal load factor from the observable OLF and subsequently the value of the spill.

The primary goal of this paper is to address some technical issues in airline spill analysis, especially in deriving the spill formulae. There are a few areas that need further research. First, the estimation issue is non-trivial because of the fact that the observed demand was truncated at the capacity. Second, it will be quite interesting to study the implication of yield management system on the spill and vice versa. It is well known that modern yield management models typically use the nested booking policy, implying that many low fare classes are *closed* before the flight departure time. The fact is that whenever some class is closed, there will be spill. It is not clear yet how to integrate these two. Finally, it is important to use real booking data to empirically characterize various demand structures so that useful guidelines for spill analysis can be established.

Acknowledgements

The authors are grateful to Bill Swan and the referees for helpful comments and suggestions on the earlier version of this paper. The authors bear the responsibility for all remaining errors. The authors acknowledge gratefully the research grant support of the Nanyang Technological University through RG1/95 and the research grant support of the National Science and Engineering Research Council (NSERC) of Canada.

References

Johnson, N.L., Kutz, S., Balakrishnan, N., 1995. Continuous Univariate Distributions, vol. 2, 2nd Ed. Wiley, New York.

Shlifer, R., Vardi, Y., 1975. An airline overbooking policy. Transportation Science 9, 101–114.

Swan, W.H., 1983. Traffic losses at high load factors. AGIFORS Symposium Proceedings 23, 155–178.

Swan, W.M., 1992. Spill Model Revisited. Unpublished Staff Working Paper, Boeing Company.