Queuing Models To Balance Systems With Excess Supply

Gastón Mendoza, Fairleigh Dickinson University, USA
Mohammad Sedaghat, Fairleigh Dickinson University, USA
K. Paul Yoon, Fairleigh Dickinson University, USA

ABSTRACT

Many manufacturing and service activities can be modeled using queuing theory. The optimization of the long-run solution to imbalances between supply and demand is very important to established businesses. This paper presents a family of queuing models that minimize the expected total cost incurred when restoring equilibrium to a stochastic system that has become unstable due to changes in the environmental parameters affecting its behavior. Analytical expressions for the expected total cost in terms of a policy parameter are derived from which numerically-savvy users can obtain the policy that minimizes the expected total cost. To determine the model parameters that most affect the optimal policy and to facilitate the determination of near-optimal policies, exact solutions were found for a large number of scenarios and then used to fit a regression model. The resulting regression equation can be used by practitioners to find policy parameters that approximately minimize the expected total cost due to imbalances in supply and demand.

Keywords: Queuing systems; Markov chains; Double-Ended queue; steady state probabilities.

INTRODUCTION

In business organizations, the task of satisfying the demand without oversupplying is often a difficult one. If the demand is higher than the supply, a policy maker has two choices: 1) increase the supply or 2) decrease the demand. On the other hand, if the supply is higher than the demand, the policy maker may choose to lower the supply or promote more demand. In either case, there is a cost associated with the balancing and the optimum decision then depends on the total cost which may be incurred as a result of selecting one of the above policies.

Supply and demand of goods and services has been modeled by many researchers who analyzed and described their behavior and proposed various ways to control supply and demand imbalances. In particular, double-ended models with finite queues have been considered by Brant and Brandt (1999, 2004), Connolly (2002), Takahashi et al (2000), Kendall (1951), Parra and Gallego (1999), Perry and Stadje (1999), Sasieni (1961), and Zenios (1999).

We present a double-ended queuing model for stochastic supply/demand systems where supply and demand queues have finite maximum possible lengths k' and k", respectively. We can describe the state of the system by a one-dimensional index. Excess supply results in a positive index while excess demand results in a negative index. If instant pairing off is assumed, the queue can be either positive or negative, but not both at the same time. By associating costs per time unit due to a unit of excess of supply or demand, we can express a total cost due to imbalance of demand and supply. We examine the queuing behavior and how to minimize the above total cost by advanced planning aimed to hold imbalance costs at a minimum. Mendoza and Sedaghat (1999) derived exact closed-form solutions for the case where k' = k". In this paper, we removed such restriction and found the exact analytical expression for the cost of restoring balance to the system. Finding the policy that minimizes the cost function requires familiarity with numerical optimization techniques. To facilitate the use of the proposed model by practitioners and to determine which of the model parameters are most important in the determination of the optimal policy, we generated a number of scenarios and for each we found the optimal policy factor. The results of the analyses of those scenarios were used to find regression equations that express the corresponding policy factor in terms of the most relevant model parameters.

THE MODEL

Initially, we consider a simple system with only one kind of commodity and many consumers. Both supply and demand are assumed to take place one unit at a time. When there is a demand of one unit, it will be satisfied by commodities in stock, and if there is no item in stock, the demand joins the queue in the demand side and will wait until the commodity becomes available. On the other hand, when the commodity is available and there is no immediate demand for it, it will join the queue on the supply side and will wait until the arrival of the next demand. If a consumer (demand) arrives while k' consumers are already in the queue, it leaves the system. Similarly, when k" units of the commodity (supply) are already in the queue, there will be no more supply to the system.

Our model can be considered as a double queuing system consisting of the servers (supplier) and the arrivals (consumers). In other words, there may be a queue of available suppliers waiting for a consumer or a queue of consumers each waiting to be satisfied by a supplier. We assume that units of supply (of commodity, personnel, service, etc.) arrive according to a Poisson process with average rate λ while units of demand of the same kind arrive according to a Poisson process with average rate μ. A unit of demand (supply) would be instantly paired off, at time of arrival, with a unit of supply (demand), provided that there is at least one unit of supply (demand) in the system waiting to be distributed. Otherwise, that unit of demand (supply) would join the queue on the demand (supply) side. The system is said to be in state "-m", m = 0, 1, …, k', if there are m units of excess demand in the queue, and in state "m", m = 1, 2, …, k", if there are m units of excess supply. Both k' and k" are assumed to be finite.

Figure 1 illustrates the model with four supply units waiting for a demand unit to serve. To help visualizing the model imagine that randomly arriving taxis with a finite number of waiting spaces are available to serve randomly arriving passengers.

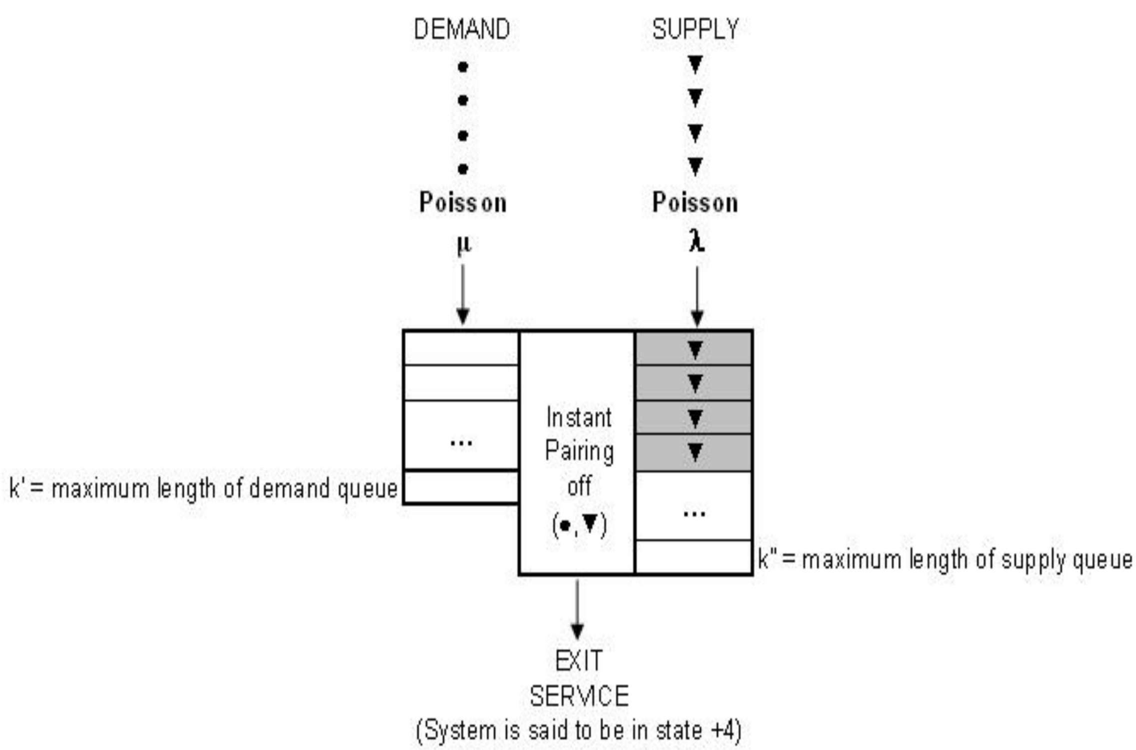

Figure 1: Model and Notation of Double-Ended Queue

ANALYTICAL FORMULATION AND SOLUTION

Bollapragada and Rao (2006) and Leeman (1964) considered the effects of managing supply. In this paper, we focus our attention to situations where supply is higher than demand, that is when $\boldsymbol{\lambda} > \boldsymbol{\mu}$. In these situations, either of two policies may be chosen; i.e.reducing the supply or increasing the demand. These two cases and their mathematical formulations are considered next.

Case I: Balance By Reducing Supply

Suppose that **c'** is the cost per time unit of one unit of excess demand in the queue, **c"** that of a unit of excess supply in the queue, $\boldsymbol{\alpha}$ is supply reduction factor ($\boldsymbol{\alpha}$ is applied to the given rate $\boldsymbol{\lambda}$ of supply and the new supply rate becomes $\boldsymbol{\alpha\lambda}$, $0 < \boldsymbol{\alpha} < 1$), and $\mathbf{c}_{\boldsymbol{\alpha}}$ is the cost incurred per time unit in reducing the supply rate by one unit. Let P_m be the probability that the system is in state m, where m = -k',…, 0,…, k". If $\boldsymbol{\rho = \alpha\lambda/\mu}$ denotes the utilization factor, the expected total cost when a supply reduction factor is applied is

$$C(\alpha) = c' \sum_{m=-k'}^{-1} (-m)\, P_m + c'' \sum_{m=1}^{k''} m\, P_m + c_\alpha\, \lambda(1 - \alpha) \qquad \text{(I)}$$

where the first summation is the expected undersupply cost, the second is the expected oversupply cost, and the third is the expected cost of reducing the supply rate from $\boldsymbol{\lambda}$ to $\boldsymbol{\alpha\lambda}$..

Our proposed model can be used to find stationary and transient probability distributions. In this paper we focus our attention on the stationary probabilities of being in a given state after the system is in operation long enough that all influences of the initial states have become negligible. Let P_m now denote the steady-state probability that the system is in state m, where m = -k',…, 0,…, k", it can be found using balance arguments that the balance equations for steady-state are

$$\begin{aligned} \alpha\lambda\, P_{-k'} &= \mu\, P_{-k'+1} \\ (\alpha\lambda + \mu)\, P_m &= \alpha\lambda\, P_{m-1} + \mu\, P_{m+1} \\ \alpha\lambda\, P_{k''-1} &= \mu\, P_{k''} \end{aligned} \qquad \text{(I-1)}$$

The analytical solution to the above system of linear equations can be obtained based on balance arguments in Markov chains. **S**ee Stewart (1991, 1994). If $\boldsymbol{\rho = (\alpha\lambda)/\mu}$**,** the solution is given by

$$\begin{aligned} P_m &= \rho^{m+k'} (1 - \rho) / (1 - \rho^{k'+k''+1}) && \text{if } \rho \neq 1 && \text{(I-2)} \\ &= 1 / (k' + k'' + 1) && \text{if } \rho = 1. && \text{(I-3)} \end{aligned}$$

If above expressions for P_m are substituted in **(I)**, the expected total cost to the system when a supply reduction factor is applied is given by:

$$\begin{aligned} C(\alpha) &= [c'\, k' (k' + 1) + c''\, k'' (k'' + 1)] / [2(k' + k'' + 1)] + c_\alpha\, \lambda (1 - \alpha) && \text{if } \rho = 1 && \text{(I-4)} \\ &= f(\rho) / g(\rho) + c_\alpha\, \lambda (1 - \alpha) && \text{if } \rho \neq 1 && \text{(I-5)} \end{aligned}$$

Where

$$f(\rho) = - c' (-k' + \rho + k' \rho - \rho^{k'+1}) + c''[\rho^{k'+1} - (1 + k'')\, \rho^{k'+k''+1} + k''\, \rho^{k'+k''+2}]$$

and

$$g(\rho) = (1 - \rho) (1 - \rho^{k'+k''+1}) .$$

Case II: Balance By Increasing Demand

Strategies to manage demand that control costs due to imbalance of supply and demand have been considered by DeCroix and Arriola-Risa (1998). In our proposed model, when policy is the expansion of demand, λ remains unchanged, μ is replaced by $\beta\mu$ and the last term of (I) is replaced by $\mathbf{c_\beta \mu(\beta-1)}$, where β is the demand expansion factor (β>1). If $\mathbf{c_\beta}$ denotes the cost incurred per time unit to increase the demand by one unit, the utilization factor is now $\boldsymbol{\rho = \lambda/(\beta\mu)}$ and the expected cost when a demand expansion factor is applied becomes

$$C(\beta) = c' \sum_{m=-k'}^{-1} (-m)\, P_m + c'' \sum_{m=1}^{k''} m\, P_m + c_\beta\, \mu\, (\beta - 1) \qquad \text{(II)}$$

The first summation is the expected undersupply cost, the second is the expected oversupply cost, and the third is the expected cost of increasing the demand rate from μ to $\beta\mu$.

If P_m now denotes the steady-state probability that the system is in state m, where m = -k',…, 0,…, k", the balance equations for steady-state are

$$\begin{aligned} \lambda\, P_{-k'} &= \beta\mu\, P_{-k'+1} \\ (\lambda + \beta\,\mu)\, P_m &= \lambda\, P_{m-1} + \beta\,\mu\, P_{m+1} \\ \lambda\, P_{k''-1} &= \beta\,\mu\, P_{k''} \end{aligned} \qquad \text{(II-1)}$$

The solution to equations **(II-1)** with $\boldsymbol{\rho = \lambda/(\beta\mu)}$ is given by

$$\begin{aligned} P_m &= \rho^{\,m+k'}(1 - \rho) / (1 - \rho^{\,k'+k''+1}) && \text{if } \rho \neq 1 && \text{(II-2)} \\ &= 1 / (k' + k'' + 1) && \text{if } \rho = 1. && \text{(II-3)} \end{aligned}$$

If above expressions for $\mathbf{P_m}$ are substituted in **(II)**, the expected total cost to the system when a demand expansion factor is applied is given by:

$$\begin{aligned} C(\beta) &= [c'\, k'\, (k' + 1) + c''\, k''\, (k'' + 1)] / [2(k' + k'' + 1)] + c_\beta\, \mu\, (\beta - 1) && \text{if } \rho = 1 && \text{(II-4)} \\ &= f(\rho) / g(\rho) + c_\beta\, \mu\, (\beta - 1) && \text{if } \rho \neq 1 && \text{(II-5)} \end{aligned}$$

Where

$$f(\rho) = - c'\,(-k' + \rho + k'\rho - \rho^{\,k'+1}) + c''[\rho^{\,k'+1} - (1 + k'')\,\rho^{\,k'+k''+1} + k''\,\rho^{\,k'+k''+2}]$$

and

$$g(\rho) = (1 - \rho)\,(1 - \rho^{\,k'+k''+1}).$$

EXACT OPTIMAL POLICIES FOR A SET OF SCENARIOS

The analytical expressions for **C(α)** in **(I-4, I-5)** and **C(β)** in **(II-4, II-5)** in the previous section can be used by numerically-savvy users to find the policy value, $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$, that minimizes the corresponding expected total cost.

To determine the model parameters that most affect the optimal α and β and to find easy-to-use formulas to get approximate optimal values for α and β and corresponding expected total costs, exact optimal values of α and β were found for a number of combinations of the model parameters.

Table 1: Parameter Values Used to Generate 1,440 Scenarios

	Initial	Increment	Final
k’	5	10	25
k”	5	10	25
λ	1	0.5	4
μ	1	1	2
c’	1	1	1
c”	1	1	4
c_α	0.5	0.5	2

We use the term “scenario” to describe a fixed set of environmental parameters in the model under consideration. We generated the scenarios using the parameter values obtained from the initial, increment and final values in Table 1. From the 3 x 3 x 7 x 2 x 1 x 4 x 4 = 2016 potential scenarios, only 1440 scenarios correspond to excess supply ($\lambda > \mu$.) Unitary costs **c”** and c_α (c_β in Case II below) are given as multiples of **c’**. Consequently, total costs are all expressed in **c’** units. The optimal policy depends on the relative costs **c”/c’** and c_α/**c’** but not on **c’**. However, to remind readers that all costs are in **c’** units, we keep in the table **c’** = 1 for all scenarios.

As mentioned in the Illustration above, the exact values of α and β that minimize such expected losses were found by iteration using procedure NLP in SAS. A discussion and summary of such exact numerical results for Cases I and II are presented below.

Case I: Balance By Reducing Supply

For each scenario, the value of α that minimizes $C(\alpha)$ was found. We focus our discussion on the 1354 “feasible” scenarios where the value of α that minimizes $C(\alpha)$ lies between 0 and 1. Figures (I-1), (I-2) and (I-3) illustrate typical shapes of the relationship between the minimum expected total cost $C(\alpha)$ and the supply reduction factor α. Figure (I-1) illustrates the approximately 93% of scenarios where $C(\alpha)$ has two inflection points and its minimum occurs at a value of α between 0 and 1 where the derivative of $C(\alpha)$ is zero. Figure (I-2) illustrates the approximately 1% of scenarios where $C(\alpha)$ has one inflection point and its minimum occurs at a value of α close to 1, say greater than .90, where the derivative of $C(\alpha)$ is zero and where $C(1)$ is slightly higher than the minimum. Finally, Figure (I-3) illustrates the approximately 6% of scenarios where c_α is large relative to c’ so that $C(\alpha)$ is monotonically decreasing between 0 and 1 and its minimum occurs at α = 1.

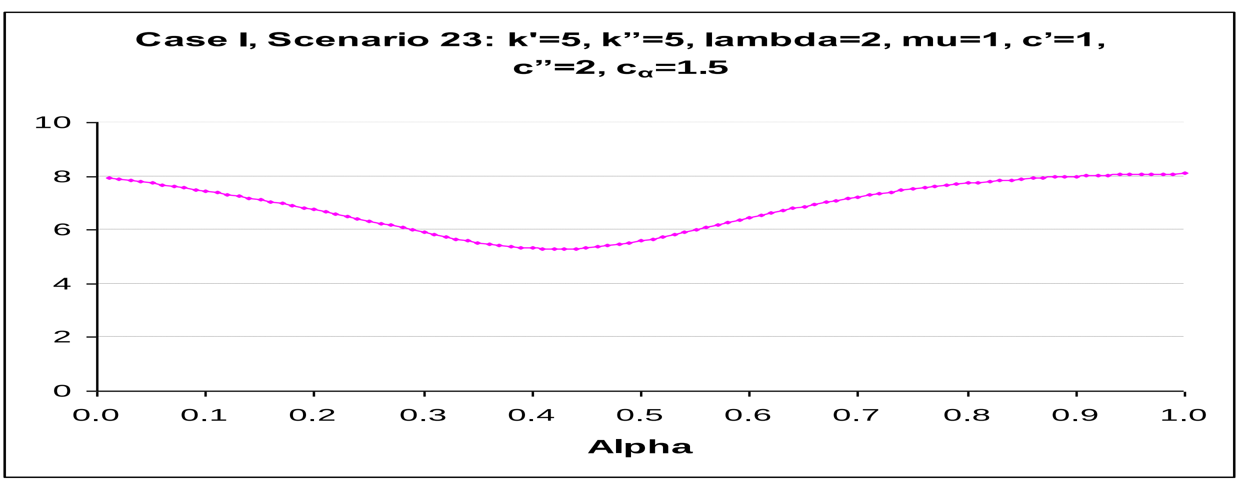

Figure I-1
Minimum at (Alpha, Total Cost) = (0.42126, 5.29315)

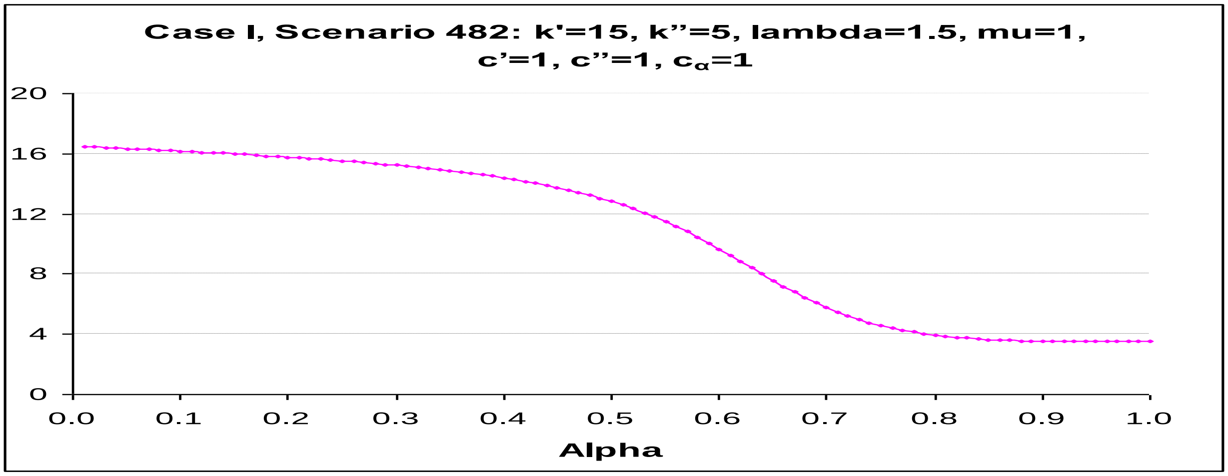

Figure I-2
Minimum at (Alpha, Total Cost) = (0.95299, 3.51410)

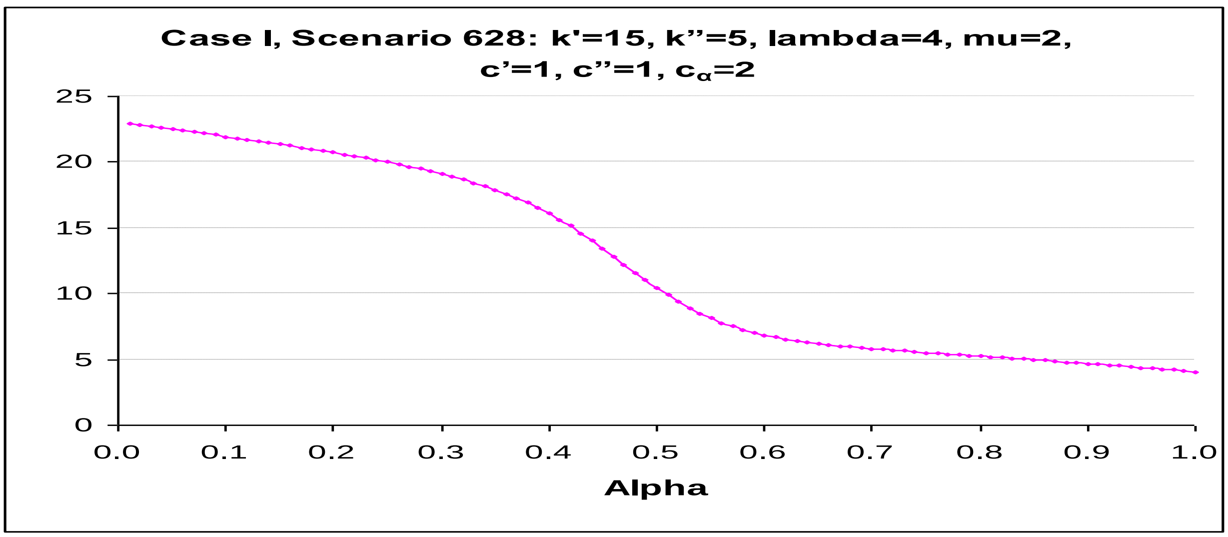

Figure I-3
Minimum at (Alpha, Total Cost) = (1.00, 4.06249)

Case II: Balance By Increasing Demand

For each scenario, the value of β that minimizes $\mathbf{C(\beta)}$ was found. We focus our discussion on the 1360 "feasible" scenarios where the value of β that minimizes $\mathbf{C(\beta)}$ is greater than 1. Figures (II-1), (II-2) and (II-3) correspond to the same three scenarios displayed before for Case I (with c_β instead of c_α,) and illustrate typical shapes of the relationship between the minimum Expected Total Cost $\mathbf{C(\beta)}$ and the demand expansion factor β. Figure (II-1) illustrates the approximately 93% of scenarios where $\mathbf{C(\beta)}$ has two inflection points and its minimum occurs at a β value greater than 1 where the derivative of $\mathbf{C(\beta)}$ is zero. Figure (II-2) illustrates the approximately 1% of scenarios where $\mathbf{C(\beta)}$ has one inflection point and its minimum occurs at a value of β close to 1, say less than 1.1, where $\mathbf{C'(\beta)}$ is zero and where $\mathbf{C(1)}$ is slightly higher than the minimum. Finally, Figure (II-3) illustrates the approximately 6% of scenarios where $\mathbf{c_\beta}$, the cost incurred per time unit to increase the demand by one unit, is large relative to c' so that $\mathbf{C(\beta)}$ is increasing for $\beta > 1$ and its minimum occurs at $\beta = 1$.

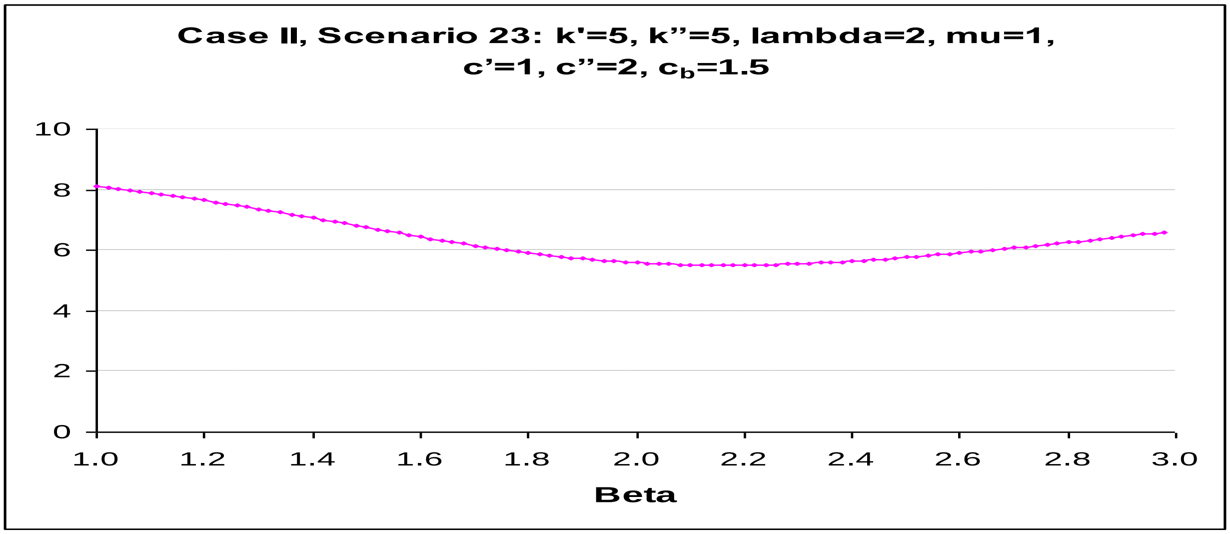

Figure II-1
Minimum at (Beta, Total Cost) = (2.16545, 5.51008)

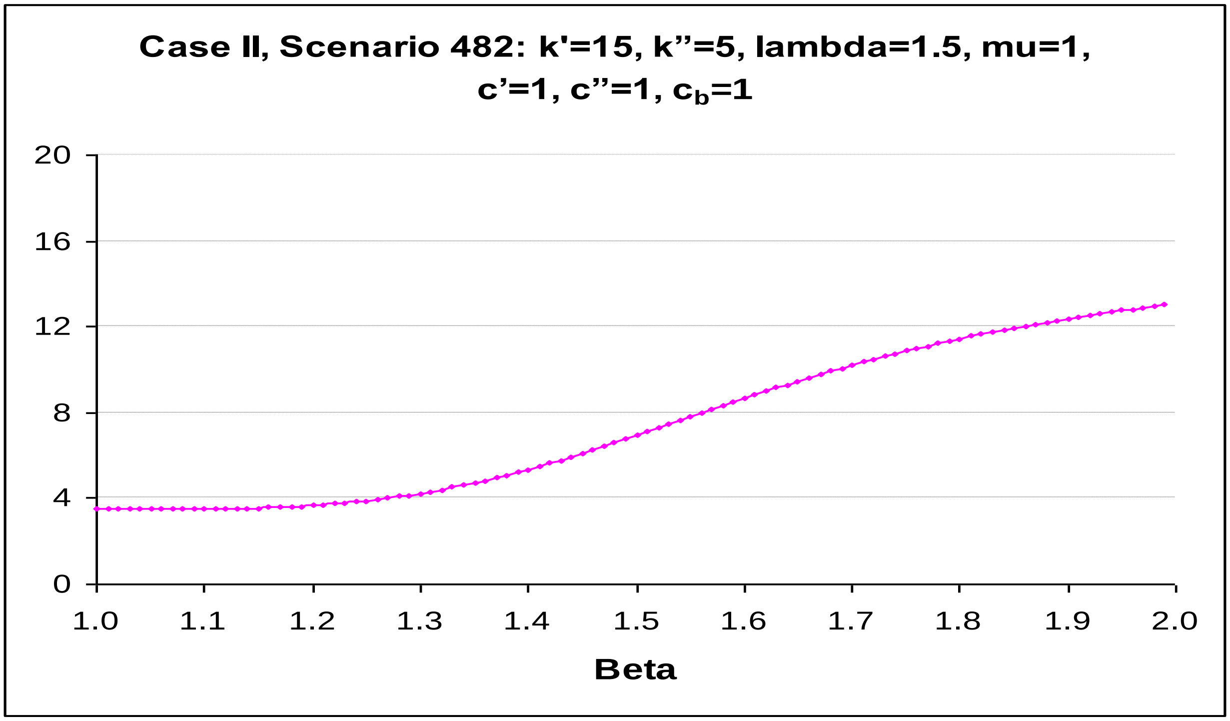

Figure II-2
Minimum at (Beta, Total Cost) = (1.07469, 3.48814)

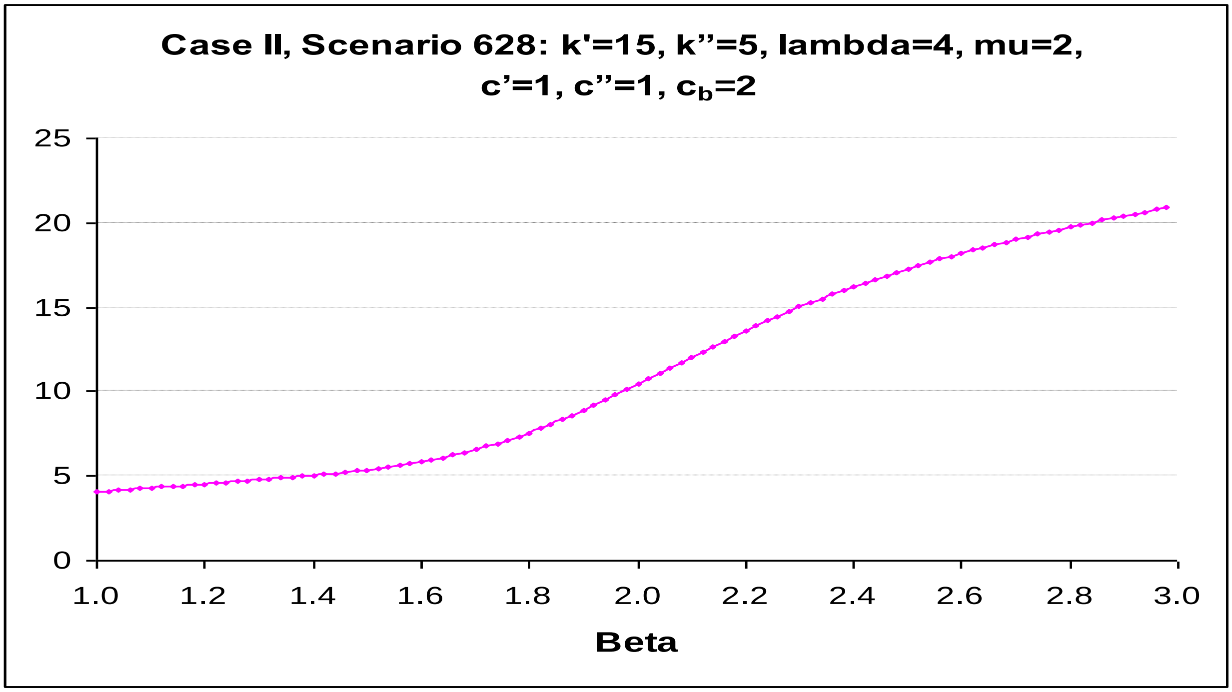

Figure II-3
Minimum at (Beta, Total Cost) = (1.00000, 4.06249)

REGRESSION APPROXIMATIONS TO OPTIMAL POLICIES

As indicated before, to determine the model parameters that most affect the optimal α and β, and to find easy-to-use formulas to get approximate optimal values for α and β and the corresponding expected total costs, the exact optimal values found for α and β were regressed on the environmental model parameters.

Case I: Balance By Reducing Supply

In 1,354 of the original 1,440 scenarios, the value of α that minimizes $\mathbf{C(\alpha)}$ was between 0 and 1. Regressions based on these 1,354 scenarios lead to the following best estimate for the optimal α in terms of the model parameters:

$$\begin{aligned}\alpha\text{-hat} = {} & \underset{[0.01409]}{1.289746} + \underset{[0.01111]}{(-0.534105)} * (\lambda/\mu) + \underset{[0.00017]}{0.008112} * k' + \underset{[0.00212]}{0.070196} * (\lambda/\mu)^2 + \underset{[0.00017]}{(-0.005705)} * k'' \\ & + \underset{[0.00124]}{(-0.026132)} * c''\end{aligned} \tag{5.1}$$

with coefficient of determination $R^2 = 0.9251$ and standard error of estimate $s_e = 0.0498$. The figures in square parentheses are the regression coefficients' standard errors. The equation identifies (λ/μ)**, k', $(\lambda/\mu)^2$, k",** and **c",** the cost per time unit of one unit of excess supply in the queue, as the most important regressors in determining the optimal α**.** It is interesting that c_α, the cost incurred per time unit in reducing the supply rate by one unit, is not a major predictor of the optimal α**.** The fitted regression equation allows a user to get approximate values of α and the corresponding expected total cost for the parameter values that he or she has, while the standard errors of the coefficients indicate how sensitive **α-hat** is to errors in estimating these regressors. If (5.1) leads to an estimated supply reduction factor greater than 1, use **α-hat** = 1. Remember that when using (5.1), c" should be expressed in c' units.

The absolute percentage error (APE) when estimating α using **(α-hat)** ranges from 0 to 58 and has a mean average percentage error (MAPE) of **9.7**. The corresponding APE of the estimated minimum total cost ranges from 0 to 110 with a MAPE of **8.7**. It should be noted that these errors are inflated by scenarios where **$\lambda/\mu > 3$.** These scenarios are unlikely in practice as they would indicate that management waited too long to control imbalance allowing the supply rate to become more than three times the demand rate. If these scenarios are omitted, APEs of the estimated minimum total cost range from 0 to 53 with a MAPE of just 7.2. Even a better MAPE would be obtained if the regression to obtain **α-hat** is based only on scenarios with **$\lambda/\mu \leq 3$.** For the sake of greater parameter coverage, we report the regression fitted using all 1,354 scenarios.

Case II: Balance By Increasing Demand

As in Case I, the exact optimal values found for β were regressed on the environmental model parameters. The fitted regression equation allows a user to get approximate values of the optimal β and the corresponding expected total cost for the parameter values that he or she has.

In 1,360 of the original 1,440 scenarios the value of β that minimizes **$C(\beta)$** was greater than 1. As in Case I, we report the regression fitted using all 1,360 feasible scenarios.

$$\begin{array}{l} \beta\text{-hat} = -0.22300 + 1.09590 * (\lambda/\mu) + (-0.03959) * k' + (0.03052) * k'' + 0.13979 * c'' \\ \qquad [0.03072] \quad [0.00766] \qquad\qquad [0.00082] \qquad\qquad [0.00084] \qquad\qquad [0.00613] \end{array} \tag{5.2}$$

with coefficient of determination $R^2 = 0.9476$ and standard error of estimate $s_e = 0.2461$. It identifies **(λ/μ), k', k”,** and **c”,** the cost per time unit of one unit of excess supply in the queue, as the most important regressors in determining the optimal **β.** However, **c_β** , the cost incurred per time unit to increase the demand by one unit, is not an important predictor of the optimal **β.** The fitted regression equation allows a user to get approximate values of the optimal β and the corresponding expected total cost for the parameter values that he or she has. The standard errors of the coefficients indicate how sensitive **β-hat** is to errors in estimating these regressors. If (5.2) leads to an estimated demand expansion factor less than 1, use **β-hat** = 1. Again, remember than when using (5.2), c” should be expressed as a multiple of c’.

The absolute percentage error (APE) when estimating β using **(β-hat)** ranges from 0 to 41 and has a mean average percentage error (MAPE) of **8.0**. The corresponding APE of the estimated minimum total cost ranges from 0 to 27 with a MAPE of **3.2**.

MODEL ILLUSTRATIONS

An Illustration Of Exact And Approximate Solutions

Consider an educational job training/hiring system that specializes in training/placing data managers. As soon as a trainee completes his/her training, he/she is entered in the supply side of a centralized database. If there is one opening and there is no other registrant ahead of him/her, he/she will be given the job. Otherwise, he/she will join the supply queue. We assume that the arrivals of those who go through the training program follow a Poisson process with supply rate λ per unit time. On the demand side, we assume that a company A which needs a data manager will be entered to the demand side of the database. Again, if there is at least one registrant waiting for a position and no other company had registered ahead of company A, then the registrant waiting in the line will be hired by company A. Otherwise, company A would have to join the queue for those companies waiting to hire trainees who have completed their training. We assume that the company requests coming to the registration office looking for available data managers follow a Poisson process with demand rate μ. When there are data managers waiting to be hired, the queue is of positive length. When there are companies waiting to hire data managers, the queue is of negative length.

Suppose that λ = 1080 trainees per year, μ = 720 positions per year, and k’ = k” = 15. Further, assume that the annual cost of a unit of excess demand in the queue (not having a trained data manager to fill an available position) is c' = \$5000 (company fee), the net annual cost of a unit of excess supply in the queue (a non-hired data manager) is **c"** = \$20000 (to cover advertising, screening, selection, and training costs), the annual cost of reducing the supply rate by one data manager is $\mathbf{c}_\alpha$ = \$5000 (marginal cost of unused training resources), and the annual cost of reducing the demand rate by one position is $\mathbf{c}_\beta$ = \$5000 (forgone fee). Assuming for ease 360 days in a year, on a per day basis $\boldsymbol{\lambda}$ = 1080/360 = 3 trainees/day, $\boldsymbol{\mu}$ = 720/360 = 2 positions/ day, c' = 5000/360, c" = 20000/360, $\mathbf{c}_\alpha$ = 5000/360, and $\mathbf{c}_\beta$ =20000/360.

Solution By Reducing Supply

The exact value of $\boldsymbol{\alpha}$ that minimizes the total expected cost C($\boldsymbol{\alpha}$) in (I-5) can be found numerically by iteration techniques such as procedure NLP in SAS. For information about NLP visit the Website: http://support.sas.com/rnd/app/index.html. For k' = 15, k" = 15, $\boldsymbol{\lambda}$ = 3, $\boldsymbol{\mu}$ = 2, and relative unitary costs c"/c' = 4, and $\mathbf{c}_\alpha$/c' = 1, the supply reduction factor that minimizes the total expected is $\boldsymbol{\alpha}$ = 0.57089 with C($\boldsymbol{\alpha}$) **=** 12.6121 (in c' units). That is, we would reduce the annual number of trainees from 1080 to 1080*0.57089 = 617 leading to an annual expected total cost due to unbalance of 12.6121* c' = 12.6121*5000 = \$63,060.50.

Solution By Increasing Demand

As before, the exact value of $\boldsymbol{\beta}$ that minimizes the total expected cost C($\boldsymbol{\beta}$) in (II-5) can be found by iteration. For k' = 15, k" = 15, $\boldsymbol{\lambda}$ = 3, $\boldsymbol{\mu}$ = 2, c"/c' = 4, and $\mathbf{c}_\beta$/c' = 1, the demand expansion factor that minimizes the total expected cost is $\boldsymbol{\beta}$ = 1.74155 with C($\boldsymbol{\beta}$) **=** 12.8228 (in c' units). That is, we would increase the annual number of company requests from 720 to 720*1.74155 = 1254 leading to an annual expected total cost due to unbalance of 12.8228* c' = 12.8228*5000 = \$64,114.00.

Comparing the expected total costs of the two approaches, we conclude that the best policy is to reduce the annual number of trainees from 1080 to 617. It is interesting to compare this policy with the "naïve" approach of simply reducing the annual number of trainees from 1080 to 1018*(2/3) = 720 to exactly match the demand for data managers. It can be found that following such a policy would lead to an annual expected total cost of 20.176*5000 = \$100,880. Consequently, the trainee-reduction policy results in annual savings of \$37,819.50.

For completeness, we report the estimated policy values obtained using equations (5.1), (I-5), (5.2) and (II-5): α-hat = 0.578107, C(α-hat) = 12.636736, β-hat = 1.84396 and C(β-hat) =13.196226. Since C(α-hat) is less than C(β-hat), the approximate optimal solution would be to reduce the annual number of trainees from 1080 to 1080*.578107=624 which is very close to 617, the exact solution.

Numerical Comparisons Of Three Policies

We will compare for each of three selected scenarios the efficiency of the following three policies:

(a) Reduction [expansion] factor found exactly by minimizing C(α) **[C(β)]**
(b) Reduction [expansion] factor estimated using the regression equation (5-1) [5-2]
(c) Reduction [expansion] factor found ignoring the stochastic nature of the model and setting the utilization factor $\boldsymbol{\rho}$ equal to 1 (leading to $\boldsymbol{\alpha = \mu/\lambda}$ for **Case I,** and $\boldsymbol{\beta = \lambda/\mu}$ for **Case II**)

Table 2 tabulates factors and their minimum expected costs under the three policies under consideration. Comparing figures for scenarios 23 and 482 for Cases I and II reveals that when solutions occur at points with zero slopes Case I solutions are neither uniformly superior nor inferior to corresponding Case II solutions. However, the optimal minimum total costs are close. Comparing figures (I-3) and (II-3) for Scenario 628, where solutions occur at boundary values for α and β, Cases I and II both lead to exactly the same minimum expected total cost.

To simplify comparisons on whether to reduce supply (Case I) or expand demand (Case II), we will assume that $\mathbf{c}_\alpha = \mathbf{c}_\beta$ for each of the three scenarios being discussed. For Scenario 23, the overall best policy is to reduce supply by a factor $\boldsymbol{\alpha}^*$ **=** 0.42126 leading to a minimum expected total cost of **5.29315.** The next best is to use **α-hat** = 0.46209 with total cost 5.37769; using $\alpha = \mu/\lambda$ = 0.50000 has a total cost of 5.59091. For Scenario 482, the overall best policy is to increase demand by a factor β^* = 1.07469 leading to a minimum expected total cost of **3.48814.** The next best is to use **β-hat** = 1.11936 with total cost 3.53281**;** using $\beta = \lambda/\mu$ = 1.50000 has a total cost of 6.92857. For Scenario 482, the overall best policy is to increase demand by a factor β^* = 1.07469 leading to a minimum expected total cost of **3.48814.** The next best is to use **β-hat** = 1.11936 with total cost 3.53281**;** using $\beta = \lambda/\mu$ = 1.50000 has a total cost of 6.92857.

Table 2: Comparing Minimum Expected Costs for Three Policies

Reduction/Expansion Factor Expected Total Cost	**CASE**	**Scenario 23**	**Scenario 482**	**Scenario 628**
(I-a) Optimal α^*	I	0.42126	0.95299	1.0000
$C(\alpha^*)$	I	**5.29315**	3.51410	**4.06249**
(I-b) Regression Estimate: α-hat	I	0.46209	0.71355	0.56935
C(α-hat)	I	**5.37769**	5.39570	7.51630
(I-c) $\alpha = \mu/\lambda$	I	0.50000	0.66667	0.50000
$C(\mu/\lambda)$	I	**5.59091**	**6.92857**	**10.42857**
(II-a) Optimal β^*	II	2.16545	1.07469	1.0000
$C(\beta^*)$	II	5.51008	**3.48814**	**4.06249**
(II-b) Regression Estimate: β-hat	II	2.20302	1.11936	1.66734
C(β-hat)	II	5.56643	**3.53281**	**6.31315**
(II-c) $\beta = \lambda/\mu$	II	2.00000	1.50000	2.0000
$C(\lambda/\mu)$	II	5.59091	**6.92857**	**10.42857**

Since scenario 23 is representative of more than 90 percent of the scenarios investigated, it can be concluded that policies based on regression estimates are, in most situations, very close to policies based on exact values and much better than those found setting the utilization factor ρ equal to 1.

CONCLUSION

We present a queuing model for stochastic supply/demand systems with excess supply where inter-arrival time of units of demand and supply are assumed to be exponentially distributed and supply and demand queues have finite maximum lengths k' and k", respectively. We denote c' and c" the costs per time unit due to a unit of excess of supply and demand, respectively, while $\mathbf{c}_\alpha$ and $\mathbf{c}_\beta$ denote the costs incurred per time unit in reducing the supply rate by one unit or increasing the demand by one unit, respectively. Under these assumptions and notation, we derived formulas (I-4, 5) and (II-4, 5) for the long-run total cost due to imbalance of demand and supply as a function of either the supply reduction factor ($0 < \boldsymbol{\alpha} < 1$) or the demand expansion factor ($\boldsymbol{\beta}>1$). These formulas can be used to numerically find the optimal policy factors, i.e., those that minimize the expected total cost. By comparing the minimum expected total losses, the best policy, either reducing supply or increasing demand, is the one leading to the smaller expected total loss.

We found regression equations to estimate the optimal policy factors based on exact results found in a large number of scenarios. Overall, measures of goodness of fit and detailed performance comparisons of representative scenarios indicate that policies based on regression estimates are, in most situations, very close to policies based on exact values and much better than those found setting the utilization factor equal to 1.

The proposed model can be extended in several directions. One, already in progress, is to consider systems with excess demand instead of excess supply. Another possibility is to extend the proposed model to one dealing with a finite number of stochastic supply/demand systems. More challenging extensions are the use of more general distributions for the queues or the study of the transient behavior of the system before stationarity is achieved.

AUTHOR INFORMATION

Gaston Mendoza holds a Ph. D. in Statistics from Temple University and a MS in Mechanical Engineering from the National Engineering University in Lima, Peru. He is currently Associate Professor of Statistics at the Silberman College of Business Administration, Fairleigh Dickinson University. From 1976 to 1982 he headed the Biometrics Unit of the International Center for Tropical Agriculture in Colombia, S. A. His research areas include statistical forecasting, operations research and educational assessment.

Mohammad Sedaghat is a full time faculty of the Department of Information Systems and Decision Sciences at Fairleigh Dickinson University. He obtained his Ph.D. in Operations Research from Polytechnic University, New York. His research interests include Queuing theory and stochastic processes.

K. Paul Yoon is Chair of Information Systems and Decision Sciences Department at Fairleigh Dickinson University. His M.S. and Ph.D. degrees in Operations Research are from Kansas State University. His research areas include multiple criteria decision making (MCDM) and its applications to service and production systems. He is the coauthor of three books on MCDM.

ACKNOWLEDGMENTS

The authors are grateful to Ms. Sangeetha, V. N. for computer assistance in the tabulation of exact solutions for the scenarios covered in the paper.

REFERENCES

1. Bollapragada, R and Rao, U. S., Replenishment planning in discrete-time, capacitated, non-stationary, stochastic inventory systems, *IIE Transactions*, Vol. 38, No.7, pp. 605-, 2006.
2. Brandt, A. and Brandt, M., On the M(n)/M(n)/s queue with impatient calls, *Performance Evaluation*, Vol. 35, pp. 1–18, 1999.
3. Brandt, A. and Brandt, M., On the Two-Class M/M/1 System under Preemptive Resume and Impatience of the Prioritized Customers, *Queueing Systems*, Vol. 47, pp. 147-168, 2004.
4. Chu, J. T. and Sedaghat, M., Queuing Approaches to Stochastic Demand/Supply Systems, AIDS Proceedings, pp. 740-742, 1984.
5. Conolly, B. W., Parthasarathyb, P.R. and Selvarajub, N., Double-ended queues with impatience, *Computers & Operations Research*, Vol. 29, pp. 2053-2072, 2002.
6. DeCroix, G. A. and Arreola-Risa, A., On offering economic incentives to backorder, *IIE Transactions*, Vol. 30, No. 8, pp. 715-721, 1998.
7. De Vany, Arthur, Uncertainty, Waiting Time, and Capacity Utilization: A Stochastic Theory of Product Quality, *Journal of Political Economy*, pp. 523-541, 2001.
8. Gross, D and Harris, C. M., *Fundamentals of Queuing Theory*, 2nd edn, John Wiley and Sons, New York, 1985.
9. Kendall, D. G., Some Problems in The Theory of Queues, *Journal of the Royal Statistical Society*, Ser. B., Vol. 13, No. 2, pp. 151-185, 1951.
10. Leeman, W., The Reduction of Queues through the Use of Price, *Operations Research*, Vol. 12, pp. 783-785, 1964.
11. Mendoza, G. and Sedaghat, A Stochastic Model for Demand/Supply Systems, FDU Working Paper Series, 1999.
12. Parra, I. and Gallego, J. A., Multiproduct monopoly: A queueing approach, *Applied Economics*, Vol. 31, No. 5, pp. 565-, 1999.
13. Perry, D. and Stadje, W., Perishable inventory systems with impatient demands, *Mathematical Methods in Operational Research*, Vol. 50, pp. 77–90, 1999.
14. Prabhakar, B., Bambos, N. and Mountford, T.S., On synchronization of Poisson processes and queueing networks with service and synchronization nodes, *Advances in Applied Probability*, Vol. 32, pp. 824–843, 2000.

15. Sasieni, Maurice W., Double Queues and Impatient Customers with Applications to Inventory Theory, *Operations Research*, Vol. 9, pp. 771-781, 1961.
16. Stewart, W. J., *The Numerical Solution of Markov Chains*, Marcel Dekker, Inc., New York, 1991.
17. Stewart, W. J., *An Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1994.
18. Takahashi, M., Osawa, H. and Takahashi T., On a synchronization queue with two finite buffers, *Queueing Systems*, Vol. 36, pp. 107-123, 2000.
19. Zenios, S.A., Modeling the transplant waiting list: a queueing model with reneging, *Queueing Systems*, Vol. 31, pp. 239–251, 1999.

NOTES

NOTES