

Stochastic Model of Waiting Time: A Case of Two Selected Banks in the Sekondi-Takoradi Metropolis

Samuel Mawuli Azumah, John Awuah Addor, Frank B. K. Twenefour, Emmanuel M. Baah

Department of Mathematics, Statistics and Actuarial Science, Takoradi Technical University, Takoradi, Ghana

Email: frank.twenefour@gmail.com

How to cite this paper: Azumah, S.M., Addor, J.A., Twenefour, F.B.K. and Baah, E.M. (2021) Stochastic Model of Waiting Time: A Case of Two Selected Banks in the Sekondi-Takoradi Metropolis. *Open Journal of Statistics*, 11, 906-924.

<https://doi.org/10.4236/ojs.2021.115053>

Received: August 27, 2021

Accepted: October 26, 2021

Published: October 29, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Queue is an act of joining a line to be served and it is part of our everyday human involvement. The objectives of the study focused on using a mathematical model to determine the waiting time of two selected banks as well as compare the average waiting time between the banks. The study uncovered the extent of usage of queuing models in achieving customer satisfaction as well as permitting to make better decisions relating to potential waiting times for customers. The study adopted a case study and observational research with the source of data being primary. Purposive sampling technique was used to select the two banks under study with the target population comprising of all the customers who intended to transact businesses with the banks within the period of 11 am to 12 pm. The sample sizes for the first, second and third day of the first bank are twenty-eight (28), seventeen (17) and twenty (20) respectively with three servers on each day whereas that for the first, second and third day of the second bank is twenty (20), nine (9) and seventeen (17) with two servers on each day. A multiple server (M/M/s) Model was adopted, and Tora Software was the statistical tool used for the analysis. Findings of the study revealed that the second bank had a higher utilization factor than the first bank. Also, the number of customers in the banking hall of the second bank was higher than that of the first bank during the entire period of observation. Finally, it takes customers of the first bank lesser minutes to complete their transaction than the second bank. In conclusion, the three days observations revealed different banking situations faced by customers in both banks which had effect on waiting time of customer service. The waiting time of customer service has effect on the number of customers in the queue and system, the probability associated with the emptiness of the system and the utilization factor. Based on the results, the study recommended, *inter alia*,

that the management of the second bank should adopt a three-server (M/M/3) model.

Keywords

Average Waiting Time, Multiple Server Model, Queuing Modeling

1. Introduction

Queue is an act of joining a line or waiting. Queues are mostly formed when customers or clients (arrivals) demanding services have to wait because their number surpasses the number of servers available or the facility does not work resourcefully or takes more than the time prescribed to service a customer. It is part of everyday human involvement at banks, filling stations, barber shops, salon shops, bus-stops, supermarkets, telephone booths, toll gates, food canteens etc. A queuing system consists of a server, customers who demand service and a queue of customers waiting to be attended to. Queuing models offer a valuable method for researchers to develop and analyze the efficiency of queuing systems [1]. A queuing system consists essentially of the arrival or inputs to the system, the discipline of queuing and the service facility. The models require a suitable balance to be found between the cost of service and the amount of waiting. Queuing analysis attempts to understand the behavior of the system using real-life data to project a current system into a future system, build an empirical model for use in system design and construct simulations that model a system. The study of queuing offers a conceptual context and insights into how queuing processes operate [2]. The waiting line is one of the key areas of concern for any company or service provider, such as banks. The theory of queuing is basically a mathematical method, used to evaluate waiting lines. The concerns that require waiting are resolved (queuing). The principle of queuing is used to test the blockages and delays of standing in queues. It also helps to build a more effective queuing mechanism that decreases the waiting time for customers and increases the number of customers served by the server.

A heavy and ever-growing rivalry within and through markets, which also includes financial services, characterizes today's age of globalization. Banks continue to face wrenching problems that they need to keep up with in order to survive. Banks, which are the key component of the financial system and mediate between the economy's surplus and deficit sectors, are often the center of interest for many consumers who wish to carry out one trade or the other through the services offered by these banks. Competition in the banking sector is becoming more serious partly due to uniform banking regulatory imperatives and due to consumer awareness of their entitlements. Bank customers have become highly demanding, as high quality, low pricing and prompt delivery of service are expected. Customers expect their preferred banks to further boost their value [3].

Overcrowded banking halls are a typical feature of banks. This also contributes to inefficient customer loyalty levels and the transfer of clients or customers from one bank to the other, without any hesitation, finding better banking facilities. In order to minimize the queue dilemma, the new automation of bank facilities (such as Online Real Time, Automatic Teller Machine (ATM), etc.) has not yielded the necessary results due to numerous network failures and poor server maintenance by high-level specialists. Because of this, the banks tended to have lengthy queues. Since there are alternative uses for the time available to either economic agent, for recreation or for work, ceaseless queues are created in banks that also contribute to low productivity. The number of hours devoted to work decides the income of a person. Economic agents or consumers often devote significant sums of their time as they come to gas stations, away from work or recreation conditions.

Bank service delivery is personal, customers are served directly, or if the system is complete, they enter a queue. A queue exists where conveniences are scarce and at a given time will not meet demand made against them. Most clients, however, are not relaxed waiting or queuing [3]. The risk of having consumers in a line is that it could cost them their waiting time. Time, essentially, is both life and wealth. Waiting to be served in long queues is an uncomfortable practice for both consumers and internal company systems, as it turns into lost lives, wasted working hours and wasted opportunities [4]. As reported by Nafees [5], often there are insufficiencies in service due to an undue delay in service that could be due to new workers. Service employment shortages past their proper time result in the lack of potential job prospects. The pursuit of a desired equilibrium between operating performance and service quality is a key challenge in planning and maintaining a service facility. Many banks have made great efforts in Ghana to boost service quality and customer loyalty, but most of them face a severe customer waiting line dilemma. The waiting line of customers occurs in banks due to poor queuing system performance and low system operation rate. The customer's waiting queues grow when the customer service may not be provided directly at the point of the customers' entry to the service centre [6]. Time is money and waiting is a non-value-added practice that can encourage people to come into the system with long pain and economic expense. Therefore, a client's wish to receive an appropriate and timely service delivery from a service system is often a desire.

Today, banks are one of the most significant units of the public and have fundamental roles that cannot override by private or corporate sectors of the general economy. The development of a growing economy and the improvement in a developed economy primarily holds its nodes on the banking sector and most especially, commercial banks. The existent problem of long queues in the banking halls for hours cause customers who are not served before their time limit to leave the system because of impatience. This limits productivity and makes patronage more tedious. Expected long waiting lines have often resulted in frequent service balking or abandonment among customers of today [7].

Many consumers continue to question why banks do not hire additional personnel or expand their branch network by opening more branches to offer faster and more productive service and increase customer loyalty levels [8]. Time wasted is never recoverable. In the view of Gupta and Garg [9], clients prefer to enter the queue only when a limited waiting period is required and tend to stay in the queue when the client has verified that the waiting time is short, and that they exit before being served, beyond tolerable patience. When compared to others, some banks have longer queues. Longer waits are due to higher arrival costs, bank service delivery deficiencies, among others. Sometimes, at those periods when the lines are lengthy, there are also questions unanswered. Customers mostly bar waiting lines because they favor shorter queues to longer queues, recognizing that shorter queues travel quicker without taking into account the banking hall's distribution service time pace. Some banks may have longer queue but their time-of-service delivery will be quicker than banking halls with shorter waiting queues owing to their number of servers or advanced technology they operate with.

There are, undeniably, various variables that influence the understanding of the waiting experience by a client, some of which include: physical, social, and emotional. If there was no queue at all, it would generate the perception that the appeal of the attraction had declined to some degree [10]. It should, however, be noticed that short queues appear to draw less media interest. It is therefore necessary, in theory, not to aim at removing queues, but instead to focus on offering individuals an opportunity to join the queue or miss part of the queue and spend time elsewhere. In view of the robust role that banks play in a country's economy, a minor downturn in output may have a largely opposite impact on the economy of the country. In addition to creating confusion and lack of human hours a day, queuing in banking halls has high adverse costs.

In view of these grievances, it became wise to use a statistical model to examine the condition of the queuing and also to compare the average waiting time between the banks chosen in the metropolis of Sekondi-Takoradi.

2. Materials and Methods

Purposive sampling technique was used to select the banks under study. This sampling technique was used because, the two selected banks have almost the same number of customers as well as queue length. It is generally known that a sample is a subset of the population. The observation was done for three days to check whether customers face the same situation at any time, they enter the banking hall to do their transactions in the two banks. The sample size was based on the number of customers that visited the banking hall from 11 am to 12 pm each day during the period of data collection. The sample size for each day is shown in **Table 1** below.

The observation was made during the working hours (11 am - 12 pm) each of the three days. A customer was considered to have arrived when he or she joins

Table 1. Sample size of the two selected banks.

Bank	Bank A			Bank B		
	Date	13/07/20	14/07/20	15/07/20	20/07/20	21/07/20
Total time (hours)	1	1	1	1	1	1
Mean arrival rate	28	17	20	20	9	17
Mean service rate	24	16	18	17	6	15
Number of tellers	3	3	3	2	2	2
Sample size	28	17	20	20	9	17

Source: Field survey, 2020.

the queue at the pay point center. The service time was recorded from the time that the customer gets to the counter for service and the time that the customer departs.

The method of analysis for this study is the multi-server queuing modeling system which follows $(M/M/S): (\infty/FCFS)$ specification. In that case, the performance measure analysis including, the arrival time, waiting time service time, average customers, and the number of servers available were computed using the appropriate tools. The queuing system consists essentially of three major components that is, the source population and the way customers arrive at the system, the servicing system, and the condition of the customers exiting the system. The system consists of more servers, an arrival pattern of customer, service pattern, queue discipline, the order in which service is provided and customer behavior. There are several everyday examples that can be described as queuing systems, such as bank-teller service, computer systems, manufacturing systems, maintenance systems, communication systems and so on. The following sections discuss each of these areas.

If A_i is the inter-arrival time between the arrivals of $(i - 1)$ th and i th customers, we shall denote the mean (or expected) inter-arrival time by $E(A)$ and $\lambda = 1/E(A)$ is the arrival frequency.

Arrivals may originate from one or several sources referred to as the calling population. The calling population can be limited or 'unlimited'.

An example of a limited calling population may be that of a fixed number of machines that fail randomly. The arrival process consists of describing how customers arrive to the system. If A_i is the inter-arrival time between the arrivals of $(i - 1)$ th and i th customers, we shall denote the mean (or expected) inter-arrival time by $E(A)$ and $\lambda = 1/E(A)$ is the arrival frequency. The service mechanism of a queuing system is specified by the number of servers (denoted by s), each server having its own queue or a common queue and the probability distribution of customer's service time (see **Figure 1**).

Let S_i be the service time of the i th customer, we shall denote the mean service time of a customer by $E(S)$ and $\mu = 1/E(S)$ the service rate of a server.

According to Little [11], the long-term average number of customers in a

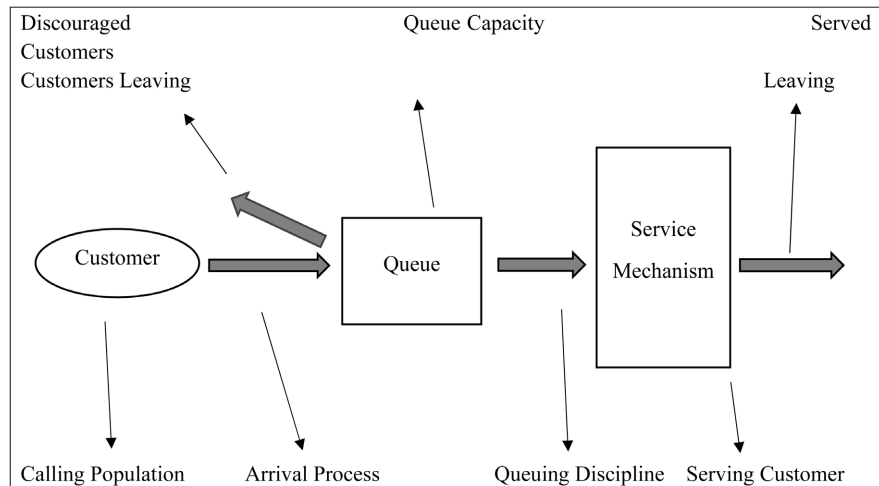


Figure 1. A queuing system in the banks. Source: Researchers construct.

stable system L , is equal to the long-term average arrival rate, λ , multiplied by the long-term average time a customer spends in the system, W ; i.e. $L = \lambda W$.

The relation between L and W is given by Little’s Law. Let L be the average number of customers in the system at any moment of time assuming that the steady – state has been reached.

Consider a system from $t = 0$ to the indefinite future and the values of the various quantities of interest as time progresses which has a connection between average and typical with the number of customers in the system, the customers delay and so on [12] [13].

Let

$L(t)$ = Number of customers in the system at time t

$\lambda(t)$ = Number of customers who arrived in the interval $[0, t]$

W_j = Time spent in the system by the j th arriving customer our intuitive notion of the “typical” number of customers in the system observed up to time t is

$$L_t = \frac{1}{t} \int_0^t L(T) dT$$

where time average of $L(T)$ up to time t . Naturally, L_t changes with the time t , but in many systems of interest, L_t tends to a steady-state L as t increases, that is, $L = \lim_{t \rightarrow \infty} L_t$.

In this case, L the steady-state time average (or simply time average) of $L(T)$.

It is also natural to view

$$\lambda_t = \frac{\lambda(t)}{t}$$

as the time average arrival rate over the interval $[0, t]$ The steady-state arrival rate is defined as

$$\lambda = \lim_{t \rightarrow \infty} \lambda_t$$

(Assuming that the limit exists) The time average of the customer delays up to time t is similarly defined as

$$W_t = \frac{\sum_{j=0}^{\infty(t)} W_j}{\infty(t)}$$

that is, the average time spent in the system per customer up to time t . The steady-state time average customer delay is defined as

$$W = \lim_{t \rightarrow \infty} W_t$$

It turns out that the quantities L , λ , and W above are related by a simple formula that makes it possible to determine one given the other. This result, known as Little's Theorem, has the form

$$L = \lambda W$$

Little's Theorem [11] expresses the natural idea that crowded systems (large W) are associated with long customer delays (large W) and reversely. For example, on a rainy day, traffic on a rush hour moves slower than average (large W), while the streets are more crowded (large L). Similarly, a fast-food restaurant (small W) needs a smaller waiting room (small L) than a regular restaurant for the same customer arrival rate. In addition, for the average number of customers in the queue at time t is given by

$$L_q = \lambda W_q$$

and the average number of customers in the service at time t

$$L_s = \lambda W_s$$

Since all queues are characterized by arrival, service and queue and its discipline, the queue system is usually described in shorten form by using these characteristics. The general notation is:

$$[A/B/s]:\{d/e/f\}$$

where,

A = Probability distribution of the arrivals

B = Probability distribution of the departures

s = Number of servers (channels)

d = The capacity of the queue(s)

e = The size of the calling population

f = Queue ranking rule (Ordering of the queue)

Mathematical Notations used in Connection with M/M/s Queuing Model:

s = number of servers in the system at time t

n = number of customers in the system (waiting + service facility) at time t

λ = mean arrival rate (number of arrivals per unit of time)

μ = mean service rate per busy server (number of customers served per unit of time)

ρ = server utilization factor

L_q = expected number of customers in the queue

L_s = expected number of customers in the system (waiting and being served)

L = average number of customers in the system (waiting and being served)

W_q = expected waiting time per customer in the queue (expected time a customer keeps waiting in line)

W_s = expected time a customer spends in the system. (In waiting + being served).

W = the average time customer spends in the system (In waiting + being served).

This model treats the condition in which there are several service stations in parallel and each customer in the waiting queue can be served by more than one station channel. Consider an M/M/s queue with arrival rate λ , service rate μ and s servers.

The condition for stability is $\rho = \frac{\lambda}{s\mu} < 1$ where ρ is called the service utilization factor, the proportion of time on average that each server is busy. The total service rate must be greater than the arrival rate, that is $s\mu > \lambda$, and if $s\mu \leq \lambda$ the queue would eventually grow infinitely large.

According to Xiao *et al.* [14], the steady distribution of queuing system follows

$$P_n = p(N = n), \quad (n = 0, 1, 2, \dots)$$

which is the probability distribution of the queue length N , as the system is in steady state, when the number of system servers is s , then we have $\lambda_n = \lambda$, $n = 0, 1, 2, \dots$

If there are n customers in the queuing system at any point in time, then the following two cases may arise [14] [15]:

- If $n < s$, (number of customers in the system is less than the number of servers), then there will be no queue. However, $(s - n)$ number of servers will not be busy. The combined service rate will then be $\mu n = n\mu$; $n < s$.
- If $n \geq s$, (number of customers in the system is more than or equal to the number of servers) then all servers will be busy and the maximum number of customers in the queue will be $(n - s)$. The combined service rate will be $\mu n = s\mu$; $n \geq s$.

From the model, the probability of any given number of customers being in the system is expressed in P_0 [16], which follows from normalization, as

$$P_0 = \left(\sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!} * \frac{1}{1-\rho} \right)^{-1}$$

The expression for the average waiting time and queue lengths are fairly complicated and depend on the probability of the average number of customers waiting in queue to be served L_q

$$L_q = P_0 \frac{S^s \rho^{S+1}}{S!(1-\rho)^2}, \quad \text{where } \rho = \frac{\lambda}{s\mu}$$

The average number of customers in service L_s ,

$$L_s = \sum_{j=1}^{s-1} jP_j + \sum_{j=s}^{\infty} sP_j = s\rho$$

Now, the average number of customers in the system becomes

$$L = L_q + L_s = L_q + s\rho = L_q + \frac{\lambda}{\mu}$$

The average time customer spends in waiting in queue before service starts W_q is

$$W_q = \frac{L_q}{\lambda}$$

The average time customer spends in the system, waiting plus being served W is

$$W = \frac{L}{\lambda} = \frac{L_q + \frac{\lambda}{\mu}}{\lambda} = \frac{L_q}{\lambda} + \frac{1}{\mu} = W_q + \frac{1}{\mu}$$

The average time customers are served W_s is

$$W_s = W - W_q = \left(W_s + \frac{1}{\mu} \right) - \frac{1}{\mu} = \frac{1}{\mu}$$

3. Results and Discussion

3.1. Summary Results of Bank A

Results for sample computation of the waiting time of Bank A on day 1 s shown below. Similarly, the results in which the data was collected, inputs, intermediate calculation, and performance measures are shown as follows in **Table 2** below using Tora.

1) Utilization factor for day 1 is given by:

$$\rho = \frac{\lambda}{s\mu} = \frac{28}{3 \times 24} = 0.3888$$

2) The probability that at any given time the system will be idle (there are no customers waiting).

$$P_0 = \left(\sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!} \times \frac{1}{1-\rho} \right)^{-1}$$

$$P_0 = \left(\sum_{j=0}^{3-1} \frac{(3 \times 0.3888)^j}{j!} + \frac{(3 \times 0.3888)^3}{3!} \times \frac{1}{1-0.3888} \right)^{-1}$$

$$P_0 = \left(\sum_{j=0}^{3-1} \frac{(3 \times 0.3888)^0}{0!} + \frac{(3 \times 0.3888)^1}{1!} + \frac{(3 \times 0.3888)^2}{2!} + \frac{(3 \times 0.3888)^3}{3!} \times \frac{1}{1-0.3888} \right)^{-1}$$

$$P_0 = (1 + 1.1664 + 0.6802 + 0.2644 \times 1.6361)^{-1}$$

Table 2. Summary results of Bank A.

BANK	BANK A			UNITS
	DATE	Day 1	Day 2	
INPUTS				
Total Time Involved (t)	1.00	1.00	1.00	Hour
Number of Customers Arrived	28	17	20	Customers
Number of Customers Served	24	16	18	Customers
Number of Servers	3	3	3	Servers
Model Type	M/M/3	M/M/3	M/M/3	
Intermediate Calculations				
Average Arrival Rate	0.0357	0.0588	0.05	
Average Serve Rate	0.0416	0.0625	0.0555	
Performance Measures				
Rho (average server utilization)	0.3888	0.3541	0.3703	
Probability of System empty	0.3048	0.3405	0.3234	
Average Customers in the system (L)	1.2506	1.1203	1.1802	
Average Customers waiting in a queue (L_q)	0.0840	0.0578	0.0690	
Average customer's wait in the System (L_s)	1.1666	1.0625	1.1112	
Average time in the system (W)	0.0446	0.0659	0.0590	
Average time in the queue (W_q)	0.0030	0.0034	0.0034	
Average time a Customer is served (W_s)	0.0416	0.0652	0.0556	

Source: Field work, (2020).

$$P_0 = 0.3049$$

3) Average queue length is given by

$$L_q = P_0 \frac{S^s \rho^{S+1}}{S!(1-\rho)^2}$$

$$L_q = 0.3049 \frac{3^3 \times 0.3888^{3+1}}{3!(1-0.3888)^2}$$

$$L_q = 0.0839$$

4) The average number of customers in service

$$L_s = s\rho$$

$$L_s = 0.3888 \times 3$$

$$L_s = 1.1664$$

5) Average number of customers in the system

$$L = L_q + L_s$$

$$L = 0.0839 + 1.1664$$

$$L = 1.2503$$

6) Average time a customer spends in waiting in queue before service starts

$$W_q = \frac{L_q}{\lambda}$$

$$W_q = \frac{0.0839}{28}$$

$$W_q = 0.0029$$

7) The average time customer spends in the system, waiting plus being served W is

$$W = W_q + \frac{1}{\mu}$$

$$W = 0.0029 + \frac{1}{24}$$

$$W = 0.0445$$

8) The average time customers are served W_s is

$$W_s = W - W_q$$

$$W_s = 0.0445 - 0.0029$$

$$W_s = 0.0416$$

Utilization factor helps to determine how busy the tellers or servers were as shown in **Table 3**.

Results on **Table 3** reveals the busiest day to be day 1 with its utilization factor of 38.88% followed by day 3, with a utilization factor of 37.03%. Day 2 happened to be busy with a utilization factor of 35.41%. The days with higher utilization factor recorded a high turnout of customers and vice versa.

The probability of the system being empty implies that no customer will be in the system to be served. The greater the probability, the faster the system will become empty. On the 2nd day, the system was the idlest with the probability of 34.05% followed by Day 3 with a probability of 32.34%. Day 1 recorded the lowest probability (30.48) of the system been empty (see **Table 4**).

As presented in **Table 5**, day 1 recorded the highest number of people with 1.2506 customers in the banking hall as compared to the other two days (day 2 and day 3) in which the least number of customers (1.1203 and 1.15802 customers respectively) were recorded.

Table 3. Utilization factor for each day of bank A.

Date	Day 1	Day 2	Day 3
Rho (average server utilization)	0.3888	0.3541	0.3703

Table 4. Probability of the system being empty or idle for each day of bank A.

Date	Day 1	Day 2	Day 3
Probability of System empty	0.3048	0.3405	0.3234

Table 5. Customers in the queue and system for each day of bank A.

Date	Day 1	Day 2	Day 3
Customers in the system (L)	1.2506	1.1203	1.1802

Waiting time in the system indicates how much time a customer is supposed to spend in the banking hall. Day 1 recorded the least waiting time (0.0446 hours) in the system which is equivalent to three (3) minutes followed by day 3 with a waiting time of 0.0590 hours which is equivalent to three and half (3:30) minutes unlike day 2 that recorded the highest waiting time (0.0659) hours which is equivalent to 4 minutes (**Table 6** referred).

Waiting time in the queue indicates how much time a customer is supposed to spend in the queue at the banking hall. The waiting time a customer spent in the queue is almost the same for all the days that is 0.0034 hours which is less than a minute (**Table 7**).

Results on **Table 8** shows that day 2 recorded the highest waiting time it takes a customer to be served with a time of 0.0652 hours which equivalents to three minutes forty seconds (3:40) followed by day 3 recording 0.0556 hours which counterparts to three minutes twenty seconds (3:20). Day 1 recorded the least waiting time of 0.0416 hours which counterparts to two and half minutes (2:30).

The differences in the time a customer waits to be served are an indication of the fact that banking situations within the facility A differ for each of the three days. This could probably explain in part the differences in customers' waiting time in the queue for each of the days aside the length of the queue itself. This represents a true reflection of the personal observations of the researchers over the three days period. The differences in the banking situations faced by customers in each day ultimately affected the results over the three days in terms of the number of customers in the queue and system, the probability associated with the system emptiness and the utilization factor. For instance, the utilization factor for day 1 in Bank A revealed it to be the busiest day which means there was a higher turnout, however, the waiting time for service was the least indicating that the servicing time is not affected by the length of the queue. Thus, servicing time depends on so many banking situations which confront customers at different times based on transaction type, efficiency of tellers and customers at the counter, reliability of internet service among other reasonable factors.

3.2. Results Summary of Bank B

Results for sample computation of the waiting time of Bank B on day 1. Similarly, the results in which the data was collected, inputs, intermediate calculation, and performance measures are shown as follows in **Table 9** below using Tora.

1) Utilization factor for 13th July is given by:

$$\rho = \frac{\lambda}{s\mu} = \frac{20}{2 \times 17} = 0.5882$$

Table 6. Waiting time in the system for each day of bank A.

Date	Day 1	Day 2	Day 3
Waiting time in the system (W)	0.0446	0.0659	0.0590

Table 7. Waiting time in the queue for each day of bank A.

Date	Day 1	Day 2	Day 3
Waiting time in the queue (W_q)	0.0030	0.0034	0.0034

Table 8. Waiting time a customer is served for each day of bank A.

Date	Day 1	Day 2	Day 3
Waiting time a customer is served (W_s)	0.0416	0.0652	0.0556

Table 9. Summary results of bank B.

BANK	BANK B			
DATE	Day 1	Day 2	Day 3	UNITS
INPUTS				
Total Time Involved (t)	1.00	1.00	1.00	Hour
Number of Customers Arrived	20	9	17	Customers
Number of Customers Served	17	6	15	Customers
Number of Servers	2	2	2	Servers
Model Type	M/M/2	M/M/2	M/M/2	
Intermediate Calculations				
Average Arrival Rate	0.05	0.1111	0.0588	
Average Serve Rate	0.0588	0.1666	0.0666	
Performance Measures				
Rho (average server utilization)	0.5882	0.75	0.5666	
Probability of System empty	0.2592	0.1428	0.2766	
Average Customers in the system (L)	1.7989	3.4285	1.6693	
Average Customers waiting in a queue (L_q)	0.6224	1.9285	0.5360	
Average customer's wait in the System (L_s)	1.1765	1.5	1.1333	
Average time in the system (W)	0.0899	0.3809	0.0982	
Average time in the queue (W_q)	0.0311	0.2142	0.0315	
Average time a Customer is served (W_s)	0.0588	0.1667	0.0667	

2) The probability that at any given time the system will be idle (there are no customers waiting).

$$P_0 = \left(\sum_{j=0}^{s-1} \frac{(s\rho)^j}{j!} + \frac{(s\rho)^s}{s!} \times \frac{1}{1-\rho} \right)^{-1}$$

$$P_0 = \left(\sum_{j=0}^{3-1} \frac{(2 \times 0.5882)^j}{j!} + \frac{(2 \times 0.5882)^2}{2!} \times \frac{1}{1-0.5882} \right)^{-1}$$

$$P_0 = \left(\sum_{j=0}^{2-1} \frac{(2 \times 0.5882)^0}{0!} + \frac{(2 \times 0.5882)^1}{1!} + \frac{(2 \times 0.5882)^2}{2!} \times \frac{1}{1-0.5882} \right)^{-1}$$

$$P_0 = (1 + 1.1764 + 0.6919 \times 2.4283)^{-1}$$

$$P_0 = 0.2592$$

3) Average queue length is given by

$$L_q = P_0 \frac{S^s \rho^{S+1}}{S!(1-\rho)^2}$$

$$L_q = 0.2592 \frac{2^2 \times 0.5882^{2+1}}{2!(1-0.5882)^2}$$

$$L_q = 0.6221$$

4) The average number of customers in service $L_s = s\rho$

$$L_s = 0.5882 \times 2$$

$$L_s = 1.1764$$

5) Average number of customers in the system

$$L = L_q + L_s$$

$$L = 0.6221 + 1.1764$$

$$L = 1.7985$$

6) Average time a customer spends in waiting in queue before service starts

$$W_q = \frac{L_q}{\lambda}$$

$$W_q = \frac{0.6221}{20}$$

$$W_q = 0.0311$$

7) The average time customer spends in the system, waiting plus being served W is

$$W = W_q + \frac{1}{\mu}$$

$$W = 0.0311 + \frac{1}{17}$$

$$W = 0.0899$$

8) The average time customers are served W_s is

$$W_s = W - W_q$$

$$W_s = 0.0899 - 0.0311$$

$$W_s = 0.0588$$

Utilization factor helps to determine how busy the tellers or servers were as shown in **Table 10**.

It can be observed that the busiest is day 2 with a utilization factor of 75% followed by day 1 and day 3, with a utilization factors of 58.82% and 56.66% in that order (**Table 10**).

The probability of the system being empty implies that no customer will be in the system to be served. The greater the probability the faster the system will become empty. On day 3, the system was the idlest with the probability of 27.66% followed by day 1 with a probability of 25.92%. Day 1 recorded the lowest probability (14.28%) of the system been empty (**Table 11**).

As presented in **Table 12**, day 2 recorded the highest number of people with 3.4285 customers in the banking hall as compared to day 1 and day 3 recording the least number of customers with 1.7989 and 1.6693 customers, respectively.

Result in **Table 13** indicates that day 1 recorded the least waiting time (0.0899 hours) in the system which is equivalent to five (5) minutes followed by Day 3 with a waiting time of 0.0982 hours which is equivalent to six (6) minutes unlike Day 2 that recorded the highest waiting time (0.3809) hours which is equivalent to twenty-three (23) minutes.

Waiting time in the queue indicates how much time a customer is supposed to spend in the queue at the banking hall. With respect to **Table 14**, the waiting time a customer spent in the queue is almost the same for the day 1 and day 3 unlike day 2 with 0.2142 hours which is equivalent to thirteen (13) minutes.

Table 10. Utilization factor for each day of bank B.

Date	Day 1	Day 2	Day 3
Rho (average server utilization)	0.5882	0.75	0.5666

Table 11. Probability of the system being empty for each day of bank B.

Date	Day 1	Day 2	Day 3
Probability of System empty	0.2592	0.1428	0.2766

Table 12. Customers in the queue and system for each day of bank B.

Date	Day 1	Day 2	Day 3
Customers in the system (L)	1.7989	3.4285	1.6693

Table 13. Waiting time in the system for each day of bank B.

Date	Day 1	Day 2	Day 3
Waiting time in the system (W)	0.0899	0.3809	0.0982

Table 14. Waiting time in the queue for each day of bank B.

Date	Day 1	Day 2	Day 3
Waiting time in the queue (W_q)	0.0311	0.2142	0.0315

Results in **Table 15** shows that day 2 recorded the highest waiting time it takes a customer to be served with a time of 0.1667 hours which equivalent to ten (10) minutes followed by day 3 July recording 0.0667 hours which counterparts to four (4) minutes. Day 1 recorded the least waiting time of 0.0588 hours which counterparts to three and half minutes (3:30).

In Bank B too, it is important to relate the differences in the time a customer waits to be served to the different banking situations faced by customers for each of the three days. This is a true reflection of the personal observations of the researchers over the three days period. A critical look at the time a customer waits to be served in day 2 might show a close link with the highest utilization factor, which could be deceptive. Comparing the utilization factors in day 1 (0.5882) and day 3 (0.5668) with their corresponding waiting service times (0.0588 and 0.0667 respectively) is an indication that length of the queue does not affect the waiting time of customer service which depends on the different banking situations that confront customers at different times. Thus, the difference in waiting time to be served influenced by factors as mentioned above could partly affect the results in respect of number of customers in the queue, the probability of the system to be empty and the utilization factor.

Table 16 depicts the results summary of the two selected banks. Bank B had a higher average utilization factor (63.49%) than Bank A (37.10%), which means

Table 15. Waiting time a customer is served for each day of bank B.

Date	Day 1	Day 2	Day 3
Waiting time, a customer is served (W_s)	0.0588	0.1667	0.0667

Table 16. Summary results of the two selected banks.

BANKS	Bank A	Bank B
INPUTS		
Total Time Involved (t)	3	3
Number of Customers Arrived	65	46
Number of Customers Served	58	38
Number of Servers	3	2
Model Type	M/M/3	M/M/2
Intermediate Calculations		
Average Arrival Rate	0.0461	0.0733
Average Serve Rate	0.0517	0.9733
Performance Measures		
Rho (average server utilization)	0.3710	0.6349
Probability of System empty	0.3229	0.2262
Average Customers in the system (L)	1.1837	2.2989
Average Customers waiting in a queue (L_q)	0.0700	1.0289
Average customer's wait in the System (L_s)	1.1134	1.2699
Average time in the system (W)	0.0565	0.1896
Average time in the queue (W_q)	0.0032	0.0922
Average time a Customer is served (W_s)	0.0541	0.0974

that the banking hall of Bank B is busier than that of Bank A. With respect to the average number of customers in the system, the number of customers in Bank A is lesser (1.1837) than the average number of customers in Bank B (2.2989). That means, the total number of customers in the queue and in the system of Bank A are fewer than that of Bank B. However, Bank A spends lesser time when serving customers in the banking hall whereas Bank B spends more. That is 0.0565 hours (4 minutes) is lesser than 0.1896 hours (11 minutes) respectfully.

Customers waiting in the queue of Bank B turn to spend more time (0.0922 hours) in the queue than the customers Bank B (0.0032 hours). In the system, customers of Bank B wait a bit longer (0.0974 hours) than customers o Bank A (0.0541) when been served.

4. Conclusions

This study reviews the application of queuing models. The M/M/s queuing model was used to analyze the waiting lines of Bank A and Bank B. The data used in the analysis were collected at the banking halls of the two selected banks for three days within a period of 11 am to 12 pm each day. Bank A was busy at a rate of 37.10%. The average number of customers in the system within the system was 1.1837 customers. In Bank A, it takes a customer 4 minutes to complete a transaction in the banking hall or system. Customers in the waiting line of Bank A spend an average of 11 seconds in the queue. In the system of Bank A, customers wait for an average three minutes and thirty seconds when been served.

On the other hand, Bank B was averagely busy at a rate of 63.49%. Also, 2.2989 customers were revealed to be the average number of customers in the system. Further, customers of Bank B tend to spend 11 minutes to complete a transaction. Equally, 0.0922 was observed to be the average waiting hours in the queue whiles 0.0974 hours was the average time customers wait to be served.

Comparing the average waiting times of these banks, Bank B had a higher utilization factor which means, its facility is busier than Bank A. Averagely, the number of customers expected to be in the banking hall of Bank B tends to be higher than that of Bank A. Commonly, it takes customers of Bank A lesser minutes to complete their transaction relative to Bank B.

Conclusively, the three days observations revealed different banking situations faced by customers in both banks which had effect on waiting time of customer service. The waiting time of customers' service, which is influenced by transaction type, teller efficiency, customer's fastness at the counter, internet reliability, among others, has effect on the number of customers in the queue and system, the probability associated with the emptiness of the system and the utilization factor.

Based on the results, the paper recommends that management of Bank B should adopt a three-server (M/M/3) model to reduce waiting time at the facility. Also, management of the two banks should adopt a computerized system that

checks the capacity of the customers in the queue in addition to improving internet reliability. To say the least, marketers of the various banks should promote usage of most of their electronic products, such as mobile banking applications or the use of ATM among others.

Authors' Contribution

All authors contributed equally in relation to the study design, literature review, methodology, data collection, analysis and discussions and the initial draft and final review of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare that they have no competing interests.

References

- [1] Banks, J., Carson, J.S., Nelson, B.L. and Nicol, D.M. (2001) Discrete-Event System Simulation. 3rd Edition, Prentice Hall, Hoboken, 24-37.
- [2] Kozlowski, D. and Worthington, D. (2015) Use of Queue Modelling in the Analysis of Elective Patient Treatment Governed by a Maximum Waiting Time Policy. *European Journal of Operations Research*, **24**, 331-338. <https://doi.org/10.1016/j.ejor.2015.01.024>
- [3] Olaniyi, T.A. (2004) An Appraisal of Techniques for Minimizing Cost of Customers Waiting in First Bank of Nigeria, Plc. Ilorin. University of Ilorin, Ilorin.
- [4] Sherman, P.J. (2015) Queuing Theory and Practice: A Source of Competitive Advantage.
- [5] Nafees, A. (2007) Analysis of the Sales Checkout Operation in Ica Supermarket Using Queuing Simulation. Unpublished Master's Thesis, University of Dalarna, Falun.
- [6] Hillier, F.S. and Lieberman, G. J. (2007) Introduction to Operations Research. 8th Edition, Tata McGraw Hill, New York.
- [7] Maister, D.H. (2009) The Psychology of Waiting Lines. <https://www.davidmaister.com>
- [8] Ogunsakin, R.E., Babalola, B.T. and Adedara, M.T. (2013) Comparison of Service Delivery by ATM in Two Banks: Application of Queuing Theory. *IOSR Journal of Mathematics*, **9**, 52-56.
- [9] Zhang, F. and Liu, B. (2013) Equilibrium Balking Strategies in Markovian Queues with Working Vacations. *Applied Mathematical Modelling*, **37**, 8264-8282.
- [10] Seigha, G., Gordon, M.B. and Mobolaji, H.O. (2017) Application of Queuing Theory to a Fast-Food Outfit. *Independent Journal of Management & Production*, **8**, 441-458. <https://doi.org/10.14807/ijmp.v8i2.576>
- [11] Little, J.D.C. (1961) A Proof for the Queuing Formula: $L = \lambda W$. *Operations Research*, **9**, 383-387. <https://doi.org/10.1287/opre.9.3.383>
- [12] Little, J.D.C. and Graves, S.C. (2008) Little's Law. In: Chhajed, D. and Lowe, T.J., Eds., *Building Intuition*. Vol. 115, Springer, Boston, 81-100. https://doi.org/10.1007/978-0-387-73699-0_5
- [13] Martino, A., Guatteri, G. and Paganoni, A.M. (2020) Hidden Markov Models for

Multivariate Functional Data. *Statistics & Probability Letters*, **167**, Article ID: 108917.

<https://doi.org/10.1016/j.spl.2020.108917>

- [14] Lyu, X., Xiao, F. and Fan, X. (2021) Application of Queuing Model in Library Service. *Procedia Computer Science*, **188**, 69-77.
<https://doi.org/10.1016/j.procs.2021.05.054>
- [15] Sharma, A.K., Kumar, R. and Sharma, G.K. (2013) Queuing Theory Approach with Queuing Model. *International Journal of Engineering Science Invention*, **2**, 1-11.
- [16] Hu, Q. (2012) *Stochastic Operations Research*. Tsinghua University Press, Beijing.