THE EFFECT OF THE DEPENDENCY IN THE MARKOV RENEWAL ARRIVAL PROCESS ON THE VARIOUS PERFORMANCE MEASURES OF AN EXPONENTIAL SERVER QUEUE

by

Butje Eddy Patuwo

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Industrial Engineering and Operations Research

APPROVED:

Ralph L. Disney, Co-Chairman	Joel A. Nachlas, Co-chairman
Osman Balci	Subhash C. Sarin

Ching-Cheng Wang

January 28, 1989

Blacksburg, Virginia

The Effect of the Dependency in the Markov Renewal Arrival Process on the Various Performance Measures of an Exponential Server Queue

(Abstract)

The thesis of this paper is to investigate how the dependency in the arrival process affects the queueing performance measures. The Markov renewal arrival process (MRAP) was chosen as the arrival process. This choice was made because many of the typical arrival processes can be obtained as special cases of the MRAP. But the main reason behind this choice is that the interarrival times of the MRAP are dependent. We assume that the queue is a single server queue with exponential service time and the investigation was carried out numerically because no analytical solution was available.

There are 5 parameters of the arrival process used in this investigation: the traffic intensity (ρ), the squared coefficient of variation (*scv*), the serial correlation defined by the lag-1 correlation (*corr*) plus the rate ξ and the coefficient of skewness (γ). Here are the performance measures of the MR/M/1 queue we investigate: the expected queue length at arbitrary times (L^t), the standard deviation (σ) of the queue length at arbitrary times and the caudal characteristic η. The other performance measures such as: the expected queue length at arrival time, the waiting time, the sojourn time, etc. can be easily obtained from L^t. We compare these performance measures against those of the corresponding GI/M/1 queue.

When the lag-1 correlation of the arrival process is negative (this means that the lags of the serial correlation alternate in signs), the L^t of the MR/M/1 queue is smaller (but not by much) than the L^t of the GI/M/1 queue. Therefore, we focus our

attention to the MR/M/1 queue with positive serial correlation. The results are presented using graphs.

We find that the coefficient of skewness of the arrival process (γ) plays an important role. The L^t curve decreases rapidly as γ increases and after certain values of γ called *the turning region*, the L^t curves flatten. This important observation indicates that to the left of the turning region, the L^t is *almost insensitive* to the dependency in the arrival process. However, to the right of the turning region, the L^t is sensitive to the positive serial correlation in the arrival process. Highly correlated arrival process (large *corr* and ξ) can cause the L^t to be significantly larger than the L^t for the uncorrelated queue.

For the MR/M/1 queue, the magnitude of the standard deviation σ is larger than the corresponding L^t. However, the shapes of the σ curves are similar to those of the L^t curves. So, all of the conclusions drawn for the L^t also apply to the standard deviation σ.

For the M/M/1 queue, the caudal characteristic η equals to the traffic intensity ρ ($\eta=\rho$). For the uncorrelated GI/M/1 queue, one would expect that when $scv<1.0$, $\eta<\rho$ (i.e., the GI/M/1 queue would behave like an E_k/M/1 queue) and when $scv>1.0$, $\eta>\rho$ (i.e., the queue would behave like a H/M/1 queue). Our results indicates that this is not necessarily true. We found again that the coefficient of skewness (γ) plays an important role. For the uncorrelated GI/M/1 queue with $scv>1.0$, η can be smaller than ρ when γ is large enough. For the correlated MR/M/1 queue, even for $scv<1.0$, a low γ value combined with the positive serial correlation can cause η to be larger than ρ. On the other hand, $scv>1.0$ does not necessarily results in $\eta>\rho$. A large value of γ can cause η to be smaller than ρ, even for the queue with highly correlated interarrival times.

ACKNOWLEDGEMENTS

I would like to thank Dr. Joel Nachlas, the Co-chairman of my dissertation committee, and Dr. Osman Balci, Dr. Subhash Sarin, Dr. Ching-Cheng Wang, the committee members, for their support and interest in my work. At the beginning of this dissertation work, I was helped by Dr. Don McNickle from University of Canterbury, New Zealand, who was visiting the IEOR department at the time. I thank him.

I was fortunate to be a student of Dr. Ralph Disney, an outstanding researcher and educator, who introduced me to queueing theory and stochastic processes, and taught, through his own example, the importance of professionalism and generosity. To him, I express my deepest gratitude. I also would like to thank for opening her home to make the completion of this dissertation possible.

, and have shared their home and love with me while I work on this dissertation. I thank them for their support. Also, I would like to thank my parents, my sisters and my brother for their continuing emotional support and financial sacrifices throughout my tenure as a student. I also acknowledge the financial assistance from the IEOR department at Virginia Tech.

Lastly, to the many friends at Dayspring, thank you for the constant encouragement and friendship, especially to and , and .

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

Chapter 1

INTRODUCTION AND BACKGROUND

1.1 The Problem.

Nearly all queueing theory assumes that the arrival process to a queue is a renewal process (i.e. the interarrival times are independent and identically distributed (i.i.d.) random variables). This assumption is often chosen for the sake of mathematical tractability even when the physical characteristic of the arrivals to the queueing system being modeled exhibits dependencies. There are only a few research articles which deal with non-renewal arrivals (excluding the works coming from the eastern European school , for example, Franken, König, Arndt, and Schmidt (1982)). These are the works of Runnenburg (1961, 1962), Loynes (1962a, 1962b), Finch (1963), Finch and Pearce (1965), Pearce (1966, 1968), Çinlar (1967), Gopinath and Morrison (1977), Jacobs (1979), Latouche (1981, 1985) and Tin (1985).

This dissertation addresses the problem of assuming that the arrival process to a queue is a renewal process when in fact the interarrival times are serially dependent by investigating the effects of the dependency in the arrival process on the queueing performance measures such as: the expected queue length, the expected sojourn time and the expected waiting time. In other words, if the interarrival times to a queue are dependent, how do the queueing performance measures of the queue differ from those with renewal arrivals? The articles cited above either show specific examples or are concerned only with developing analytical solutions to the specific arrival process chosen. To the best of our knowledge, the results presented in this dissertation are the first to systematically try to answer the question above.

Several authors have shown that the departure process from the M/GI/l queue can be weakly correlated (e.g. King (1971), Burke (1972), Disney and de Morais (1976)). This result may have led some researchers to conclude that since the departure serial correlations are small, the performance measures of any queueing system taking this departure process as an input will not be affected significantly by these small serial correlations. Our results will show cases where the dependency in the arrival process has significant effect on the queueing performance measures (even for small serial correlations), and also cases where the effect of the dependency is insignificant.

Disney and Kiessler (1987) showed that in the class of Markovian queueing networks, the traffic processes on the arc of the networks are Markov renewal processes on a countable state space. Even in networks as simple as those of Jackson (1957), if there are loops (direct or indirect feedbacks) so that a customer can travel from one node and back to that same node again, then the traffic processes on the arcs of the networks are Markov renewal. In only special cases (e.g. networks with tree structure) are the Markov renewal process equivalent to the renewal process. (The concept of equivalence between a Markov renewal process and a renewal process will be explained later in section 2.2.) Therefore, each node in a Markovian queueing network may be treated as a queue with a superposition of Markov renewal arrivals.

It is then natural to use a Markov renewal process as the arrival process in our investigation because there is a built in dependency; more importantly, because of the results of Disney and Kiessler (1987) cited above. Also, most of the common arrival process to a queue such as: Poisson, renewal, compound Poisson, batched/platooned arrivals are special cases of the Markov renewal arrival process. This will be discussed in the next section .

Therefore, as the first step, we decided to investigate the MR/M/1 queue (queue with Markov renewal arrival process and exponential service time). The main results of this research are presented in chapter 4. There, we show numerically how the mean, the coefficient of variation, the coefficient of skewness and the serial correlation of the interarrival times affect the expected queue length and other performance measures.

The remainder of this chapter is organized as follows. Section 1.2 briefly introduces the concepts of a point process and a marked point process to describe the renewal process and the Markov renewal process. The literature review is done in section 1.3. The organization of this dissertation and the numbering system used are given in section 1.4.

1.2 Background.

In this section, we informally introduce the concepts of *a point process* and *a marked point process* to describe a renewal and a Markov renewal process. The material in this section is an excerpt from Disney and Kiessler (1987, chapter 1).

Let $\{ t_n : n=0, 1, 2, \dots , t_n \in \Re_+ \}$ be a sequence of points on the non-negative real line such that $0=t_0<t_1<t_2<\dots$. Also, let $\{ z_n : n = 0, 1, 2, \dots \}$ be a sequence that takes values in some countable set E. For $n = 0, 1, 2, \dots$, define

$$\varphi_n = (z_n , t_n) \ , \tag{2.1}$$

and let

$$\varphi = \{ \varphi_n : n = 0, 1, 2, \ldots \} \ . \tag{2.2}$$

We think of φ_n as the nth epoch of a sample path (realization) φ. Define

$$\Phi = \{ \varphi : \varphi \textit{ is a sample path} \} \ . \tag{2.3}$$

Then, Φ is called a sample space.

Let $\sigma(\Phi)$ be a σ-algebra on Φ and let P be a probability measure on $\sigma(\Phi)$. Then, $(\Phi, \sigma(\Phi), P)$ is a probability space. $\sigma(\Phi)$ is also called *the event set.* Now define functions T_n and Z_n as follows:

$$T_n(\Phi) = t_n \tag{2.4}$$

and

$$Z_n(\Phi) = z_n \ . \tag{2.5}$$

Then, T_n and Z_n are random variables on $\Re_+$ and E, respectively. T_n may denote the nth epoch and Z_n is a descriptor for some quantity or quality of the event occuring at T_n. For example, T_n may be *the arrival time of the nth customer* with Z_n as *the type, the priority,* or *the service requirement* of this customer.

Let

$$\mathfrak{Z} = \{ Z_n : n = 0, 1, 2, \ldots \} \tag{2.6}$$

and

$$\mathfrak{T} = \{ T_n : n = 0, 1, 2, \ldots \} \ . \tag{2.7}$$

Then, $(\mathfrak{Z},\mathfrak{T}) = \{ (Z_n, T_n) : n = 0, 1, 2, \ldots \}$ *is a random (stochastic) process on* $E\times\Re_+$. $\mathfrak{Z}$ is called *a mark process* with state space E ; $\mathfrak{T}$ is called *a point process on* $\Re_+$; and $(\mathfrak{Z},\mathfrak{T})$ is called *a marked point process* on $E\times\Re_+$.

Define a new random variable

$$D_{n+1} = T_{n+1} - T_n \quad , \quad n = 0, 1, 2, \dots , \tag{2.8}$$

with $D_0 = T_0 = 0$, which is always non-negative because $T_{n+1} > T_n$. Let

$$\mathfrak{D} = \{ D_n : n = 0, 1, 2, \dots \} \quad . \tag{2.9}$$

Then the stochastic process $\mathfrak{D}$ is called *the incremental process of* $\mathfrak{T}$. Notice that (2.8) defines a one-one mapping between D_{n+1} and $T_{n+1} - T_n$. Consequently, one can use whichever process, $\mathfrak{T}$ or $\mathfrak{D}$, is appropriate within the given context.

Each random variable D_n in the $\mathfrak{D}$ process represents the time between two consecutive epochs (e.g., the interarrival time). In most queueing theory, a probability structure is imposed on the $\mathfrak{D}$ process. For exposition purposes, we will refer $\mathfrak{D}$ as the interarrival process (i.e. the time between two consecutive arrivals).

The simpliest structure we can impose on the interarrival process $\mathfrak{D}$ is to assume that the random variables D_n in the $\mathfrak{D}$ process are mutually independent and identically distributed (i.i.d.). Such a process is called *a renewal process.* In queueing theory, it is called *a GI process* (GI stands for general and independent). The most common assumption put on the D_n 's is that these random variables are i.i.d. and exponentially distributed. In this case, considerable use is made of the remarkable *forgetfulness* or *memorylessness* property of the exponentially distributed random variables, namely,

$$P[\, D_n > t + s \mid D_n > t \,] = P[\, D_n > s \,] = e^{-\lambda s} \quad , \tag{2.10}$$

Notice that $P[D_n > t+s \mid D_n > t]$ is independent of t. In this case, $\mathfrak{T}$ is a sequence of (Markov) dependent random variables with Erlang marginal distribution:

$$P[\, T_n \le t \,] = 1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda t}(\lambda t)^k}{k\,!} \,, \; t \ge 0 \,, \tag{2.11}$$

which is differentiable and its derivative with respect to t (the probability density function of T_n) is given by

$$\frac{d}{dt}P[\,T_n \leq t\,] = \frac{e^{-\lambda t}\,\lambda\,(\lambda t)^{n-1}}{(n-1)\,!}\,,\ t \geq 0\,. \tag{2.12}$$

Studies of queues with GI arrival processes or GI service time processes occupy the bulk of queueing theory literature. In many cases, special assumptions are placed on the distribution of the $\mathfrak{T}$ process.

The compound Poisson process is one of the most commonly used marked point processes in queueing theory. Here, the interarrival process $\mathfrak{T}$ consists of a sequence of i.i.d. exponentially distributed random variables. The mark process $\mathfrak{Z}$ is assumed to be a sequence of i.i.d. discrete positive random variables. We also assume that the $\mathfrak{T}$ process and the $\mathfrak{Z}$ process are independent of each other.

These compound Poisson processes are used in the $M^{[z]}$–queues: for example, queues with batch arrivals or batch services. Here, the interarrival process $\mathfrak{T}$ is a sequence of i.i.d. exponential random variables. For each T_n, the arrival time of the nth customer, there is a corresponding random variable Z_n representing, for example, the number of customers arriving at T_n . Then the positive random variable Z_n is called the batched size of the nth arrival. The mark process $\mathfrak{Z}$ and the interarrival process $\mathfrak{T}$ are assumed to be independent. The $\mathfrak{Z}$ process may also represent *the types of the arriving customers* or *the priority tags assigned to the arriving customers.*

A more complicated structure than that of the compound Poisson processes occurs when one allows the $\mathfrak{Z}$ process to have the Markov property (i.e., $\mathfrak{Z}$ is a Markov chain on a countable state space E). Assume also that the process $\mathfrak{Z}$ and the process $\mathfrak{T}$ are independent and the $\mathfrak{T}$ process consists of i.i.d. exponential random variables. The

Erlang processes and some generalization of these exemplify such processes.

Many types of dependencies can be imposed on these $(\mathfrak{Z},\mathfrak{T})$ processes to obtain the more general arrival process or service time process. For queueing uses, the most general models we are aware of are those called *the complete arrival process (CAP).* (See Franken, König, Arndt, and Schmidt (1982).) Here, $\mathfrak{Z}$ need not be Markov, and the $\mathfrak{Z}$ process may depend on the $\mathfrak{T}$ process. Furthermore, the state space E does not have to be the positive integers. The use of these models allows general types of dependencies in the arrival process.

Somewhere between the CAP and other $(\mathfrak{Z},\mathfrak{T})$ processes described earlier is *the Markov renewal process.* Essentially, this Markov renewal process $(\mathfrak{Z},\mathfrak{T})$ is a two-tuple random process, though $\mathfrak{Z}$ may be vector-valued. The $\mathfrak{Z}$ process is a Markov chain on a countable state space E and the $\mathfrak{T}$ process is a sequence of non-negative real-valued random variables. Unlike the previous $(\mathfrak{Z},\mathfrak{T})$ processes (excluding the CAP's), the $\mathfrak{Z}$ process and the $\mathfrak{T}$ process are not necessarily independent. For each pair of random variables (Z_{n+1}, D_{n+1}) we impose the following condition: for $t \in \Re_+$, $j \in E$, $n = 1, 2, \dots$

$$P[Z_{n+1}=j, D_{n+1} \le t \mid Z_0, \dots, Z_n; T_0, \dots, T_n] = P[Z_{n+1}=j, D_{n+1} \le t \mid Z_n], \quad (2.13)$$

that is, the pair (Z_{n+1}, D_{n+1}) depends only on the previous Z_n .

Notice that one can write (2.13) as

$$P[Z_{n+1}=j, D_{n+1} \le t \mid Z_n = i]$$

$$= P[Z_{n+1}=j \mid Z_n = i] \,.\, P[D_{n+1} \le t \mid Z_n = i \,,\, Z_{n+1}=j] \,. \quad (2.14)$$

Thus, (2.14) says that the Markov renewal process $(\mathfrak{Z},\mathfrak{T})$ is made up of two processes.

The $\mathcal{Z}$ process is a Markov chain with state space E, and the $\mathcal{T}$ process has increments that depend on the current state Z_n and the state Z_{n+1} to be visited next. Notice that there is no specific distributional assumption placed on $\mathcal{Z}$ and $\mathcal{T}$.

Let $\mathbf{A}(t)$ be an m by m matrix whose $(i,j)\,th$ element is defined by the left hand side of equation (2.14). We call $\mathbf{A}(t)$ *the semi-Markov kernel* of the Markov renewal process $(\mathcal{Z},\mathcal{T})$. Define also an m by m stochastic matrix $\mathbf{A} = [a_{ij}]$ as the transition matrix of the Markov chain $\mathcal{Z}$. Then by virtue of (2.14) we can write

$$\mathbf{A}(t) = [\, a_{ij} \cdot F_{ij}(t)] \;,\;\; i, j = 1, 2, \ldots, m \;, \tag{2.15}$$

where

$$F_{ij}(t) = P[\, D_{n+1} \leq t \mid Z_n = i \;,\; Z_{n+1} = j] \tag{2.16}$$

is any proper distribution function.

There are many special cases of this Markov renewal process which are useful in queueing theory. The followings are a few of the examples.

EXAMPLE 2.17. Suppose $\mathcal{Z}$ is a Markov chain and D_{n+1} depends only on Z_n (but not Z_{n+1}). Furthermore, let

$$P[\, D_{n+1} \leq t \mid Z_n = i\,] = 1 - e^{-\lambda_i t} \;. \tag{2.18}$$

Then, $(\mathcal{Z},\mathcal{T})$ is *a Markov process* with $\mathcal{T}$ as the stopping times (see Çinlar (1975, p. 247)). Thus, the kernel (the matrix of transition functions) of a Markov process can be written as

$$\mathbf{A}(t) = [\; a_{ij} \cdot (1 - e^{-\lambda_i t}) \,] \;. \tag{2.19}$$

Latouche (1981) uses this arrival process to study the effects of weakly correlated arrivals on the queue length process.

EXAMPLE 2.20 . Suppose the state space E of $\mathfrak{Z}$ contains only one element. Then $\mathfrak{T}$ is a renewal process. This simplification occurs, for example, in the study of the overflow process from the GI/M/1/L queue (see Çinlar and Disney (1967).

EXAMPLE 2.21 . In *compound Poisson processes*, the process $\mathfrak{T}$ does not depend on the process $\mathfrak{Z}$. Therefore,

$$P[D_{n+1} \le t \mid Z_n, Z_{n+1}] = P[D_{n+1} \le t] = 1 - e^{-\lambda t}, \tag{2.22}$$

i.e. the interarrival times are exponentially distributed. Furthermore, since the $\mathfrak{Z}$ process is a sequence of i.i.d. random variables where

$$P[Z_{n+1} = j \mid Z_n] = P[Z_{n+1} = j], \tag{2.23}$$

then,

$$P[Z_{n+1} = j, D_{n+1} \le t \mid Z_n = i] = P[Z_{n+1} = j]\,(1 - e^{-\lambda t}). \tag{2.24}$$

EXAMPLE 2.25 . Example 2.21 can be generalized by removing the exponential assumption (2.22) of the interarrival times: i.e. by letting $P[D_{n+1} \le t]$ to have some general distribution function.

EXAMPLE 2.26 . If $E = \{1, 2\}$ and

$$P[Z_{n+1} = j, D_{n+1} \le t \mid Z_n = i] = \begin{cases} F_j(t) & \text{if } i \ne j \\ 0 & \text{if } i = j \end{cases}. \tag{2.27}$$

Then, after a type-1 customer arrives at the queueing system, the arrival is of type 2 and the one after that is of type 1, and so on. Here, the types of the arriving customers alternates. This arrival process is called *the alternating renewal* arrival process.

In summary, under the Markov renewal assumption, a more complex structure is encountered. In particular, we do not require the independence between the $\mathfrak{Z}$ process and the $\mathfrak{T}$ process. This implies that $\mathfrak{T}$ is not a renewal process. However, $\mathfrak{Z}$ is always

a Markov chain but is not independent of $\mathfrak{T}$.

1.3 Literature Review.

The stationary Poisson process is perhaps the most commonly used random process to model the arrival process to a queue. There are three assumptions implicitly implied here:

1. the interarrival times D_n , $n = 0, 1, 2, \ldots$, are identically distributed with distribution function $P[\ D_n \le t\ ] = A(t)$;
2. the D_n's are mutually independent random variables;
3. $A(t) = 1 - e^{-\lambda t}$, $t \ge 0$, $\lambda > 0$.

The heavy usage of the Poisson arrivals in queueing theory is attributed to *the memoryless property* of the exponential distribution in assumption 3. This property reduces considerably the complexity of the mathematical analysis of queues. By dropping assumption 3, we have the renewal (GI) arrival process. A substantial amount of literatures also exist for this type of queues (the GI/./. queues).

To further generalize the arrival process, we ask the following two questions:

1. What happens when assumption 2 is dropped (the interarrival times D_n's are no longer mutually independent) ?
2. What type of *dependency structure* should be imposed on these dependent interarrival times ?

Runnenburg (1961 and 1962) started by assuming the random process $\mathfrak{D}$ to be a Markov chain on a finite state space ; i.e., D_n can only take values in $E=\{1, 2, ..., r\}$ and

$$P[D_{n+1}= d \mid D_0, D_1, ..., D_n] = P[D_{n+1}= d \mid D_n] . \tag{3.1}$$

Choosing $E=\{1,2\}$, (i.e. the interarrival time is either equal to 1 time unit or is equal to 2 time units), $E[D_n] = \frac{3}{2}$, and choosing the service times to be exponential with rate $\mu=1$, Runnenburg (1962) computed the expected waiting times, and the standard deviations of these waiting times with respect to the various lag-1 arrival correlations. Table 3.2 summarizes his findings. Entries on the row where $Corr[D_n, D_{n+1}] = 0$ correspond to the renewal arrival queue.

These results show that we simply cannot ignore the presence of the correlations in the arrival process; particularly when the correlations are positive. Moreover, the expected waiting times may not be an accurate performance measure for such queues because the standard deviations of the waiting times are larger than their corresponding expected values.

This queue of Runnenburg's is a special case of the MR/M/1 queue with the Markov renewal arrival kernel $\mathbf{A}(t) = [a_{ij} F_{ij}(t)]$ where $F_{ij}(t)=1$ for $t \in \{1, 2\}$ and $F_{ij}(t)=0$ for $t \notin \{1, 2\}$.

Another approach to modeling the correlated arrival process is to represent it as a time series (an autoregressive-moving-average process). (See Loynes (1962a, 1962b), Finch (1963), Finch and Pearce (1965), Pearce (1966, 1968), Gopinath and Morrison (1977), and Jacobs (1979).) Solution to this type of queue is usually given in terms of the Laplace-Stieltjes transform and the generating function, and, hence they are difficult to compute and to interpret.

Table 3.2 . Runnenburg's example: correlated arrivals and the expected waiting times.

$Corr[D_n, D_{n+1}]$	$E[W_n]$	$\sigma(W_n)$
−1.0	0.75	1.4
−0.5	0.79	1.5
0.0	0.86	1.6
0.5	1.03	1.8
0.928	2.11	3.7

Table 3.3 . Tin's example : correlated arrivals and the expected queue lengths.

arrival	*traffic intensity*				
correlation	ρ=0.1	ρ=0.2	ρ=0.5	ρ=0.8	ρ=0.9
0.0	0.1111	0.2500	1.0000	4.0000	9.0000
0.1	0.1133	0.2592	1.0736	4.3940	9.9461
0.2	0.1161	0.2705	1.1645	4.8839	11.1256
0.3	0.1196	0.2849	1.2799	5.5104	12.6379
0.4	0.1242	0.3038	1.4315	6.3405	14.6481
0.5	0.1306	0.3296	1.6403	7.4950	17.4333
0.6	0.1400	0.3674	1.9478	9.2142	21.7457
0.7	0.1552	0.4285	2.4491	12.0555	28.6045
0.8	0.1843	0.5454	3.4253	17.6820	42.4547
0.9	0.2657	0.8739	6.2500	34.3441	83.7453

Tin (1985) assumed the $\mathfrak{D}$ process to be Markov dependent and marginally, the D_n's are gamma distributed. The Laplace-Stieltjes transforms and the generating functions are applied to obtain the solutions. Table 3.3 on the previous page is an excerpt from Tin's (1985) results. In this table, the numerical values for the expected queue lengths for various traffic intensities ρ and various arrival correlations are given.

The first row entries correspond to the ordinary M/M/1 queues. We notice the larger expected queue lengths caused by the positive lag-1 correlations in the arrivals, particularly in heavy traffic situation. Unfortunately, there is no result given for the negative interarrival correlation. For traffic intensity $\rho=0.9$, we see the expected queue length grows from 9.0000 for the renewal arrival case to 83.7453 for the dependent arrival with lag-1 correlation of 0.9 (more than nine times larger).

The examples of Runnenburg and Tin (Table 3.2 and Table 3.3) show us that one needs to exercise caution when dealing with queues whose arrivals are suspected to be correlated. To model them as renewal queues (GI/./.) could yield misleading results even when the arrival correlations are small. The result of Tin (1985) is the first to bring to our attention the problem wherein the interarrival times are dependent and to give us the effect of this dependency on queueing performance measures

Latouche (1981) used a special case of a Markov renewal process $(\mathfrak{Z},\mathfrak{T})$ with the semi-Markov kernel given by (2.19). In Example 2.17 we show that $(\mathfrak{Z},\mathfrak{T})$ is a Markov process. Notice that the interarrival times depend only on the current state and not on the state to be visited next. The arrival rates λ_i are chosen to be equal to λ plus a small factor ε. The service times are assumed to be exponential distributed with rate μ. The continuous time stationary queue length probabilities are found as analytic functions of ε, for sufficiently small ε. Hence the resultant arrival correlations are also small. Therefore, the negligible effect of the arrival correlation on the queue length is expected. In spite of this, he found that in this queue, the variability in the arrival process influenced the queue lengths more than the arrival correlation did.

In 1985, Latouche constructed another type of Markov renewal arrival process similar to that in Latouche (1981), except that $F_{ij}(t)=F_i(t)$, $i=1,\ 2$ were the generalized Erlang distributions. He called this new process *the semi-Poisson process* because, macroscopically, the only difference between this process and a Poisson process is that the D_n's were correlated. He compared this special MR/M/1 queue to the ordinary M/M/1 queue, realizing that the only difference between the two queues was in the correlations of the interarrival times. Five different patterns of correlations of this semi-Poisson process are constructed. The comparisons are made on *the caudal characteristic curves* of this queue (see Neuts (1986) or Remark 5 at the end of section 3.1 for the discussion on caudal characteristic curves).

In two of the five examples, the sample paths for these semi-Poisson arrival processes look almost identical to that of a Poisson process. However, the caudal characteristic curves of these MR/M/1 queues are very different from that of an M/M/1 queue. This aberration is due to the presence of dependency in the semi-Poisson arrival process.

Other special cases of queues with Markov renewal arrivals are the $M^{[z]}$–queues: i.e., the queues with compound Poisson, batched, or platooned arrivals (e.g., Gross and Harris (1985), Neuts and Chakravarthy (1981), and Baily and Neuts (1981)). These types of arrival processes were discussed in Example 2.21 and Example 2.25. Notice that the interarrival times in the compound Poisson process are i.i.d. Thus, queues with these types of queue can be treated macroscopically as the GI/./. queues.

The first to use Markov renewal process as the arrival process to a queues is Çinlar (1965, 1967). He gives solutions in terms of the Laplace-Stieltjes transforms and the generating functions. Computability of his results, however, are technically impossible.

Later in 1978, Neuts gave the stationary queue length and the waiting time probabilities of the MR/M/1 queue in the matrix-geometric form. The same results are documented in his book (Neuts (1981)). His solution procedure is algorithmic in nature and is easy to compute. Therefore, we adopt Neuts' technique as the tool to study the MR/M/1 queue numerically.

The busy period for the MR/M/1 queue is given in Ramaswami (1980). McNickle (1974) gives the number of departures from the MR/M/1 queue. Other than Çinlar (1965) we are not aware of any published literature that deals with the finite capacity MR/M/1/L queue.

Lastly, based on the work of Arjas (1972b), de Smit (1985) studies a broader class of Markov renewal queues. A system of Wiener-Hopt integral equations is obtained. The solution depends on the ability to obtain explicit matrix factorization to this Wiener-Hopt system. Even when the explicit matrix factorization is found, the solution is still in the transformed space. Somewhere in the solution procedure, the roots of some nonlinear equation are to be found in algebraic forms. Although a larger class of problems could be solved this way, this approach does not lend itself readily to a computable solution procedure.

1.4 Organization.

The remainder of this dissertation is organized as follows. Chapter 2 gives detailed treatment of the Markov renewal arrival process. The matrix-geometric results for the MR/M/1 queue are given in chapter 3. In chapter 4, we show how the mean, the coefficient of variation, the coefficient of skewness and the serial correlation of the interarrival times affect the mean and standard deviation of the queue length and hence the expected waiting time and the expected sojourn time. Chapter 5 summarizes the results of this research and gives the direction for the future research.

The numbering system used in this dissertation is as follows. The chapters are numbered with Arabic numbers. Each chapter is divided into sections and a section may be divided into subsections. The sections are labeled using two Arabic numbers, one for the chapter and the other for the section; similarly, the subsections are labeled using 3 Arabic numbers. Thus, the fourth section of chapter three would be labeled section 3.4. Theorems, lemmas, corollaries, conjectures, examples, equations, figures and tables are numbered consecutively as n.m where n and m are Arabic numbers. Inside a section, the chapter number is suppressed. Thus one would refer to Theorem 1.47 from within chapter 3 but Theorem 3.1.47 from outside chapter 3.

Chapter 2

MARKOV RENEWAL ARRIVAL PROCESS

2.0 Introduction.

This chapter describes the Markov renewal arrival process (MRAP) used as an arrival process to an exponential server queue. Here, the emphasis is on the *interval process* (the interarrival times) of the process rather than the associated *counting process.*

Section 2.1 derives the marginal interarrival probability distribution of the MRAP and its moments. The autocovariance function and the autocorrelation function of the interarrival times are also given in this section. This section is an excerpt from Çinlar (1969, 1975a, 1975b chapter 10), and Disney and Kiessler (1987).

Section 2.2 gives conditions under which the Markov renewal process is equivalent to a renewal process. The theorems in this section are due to Simon and Disney (1984), and the presentation follows that of Disney and Kiessler (1987).

Section 2.3 particularizes the results in section 2.1 to the 2-type (2-state) MRAP. This 2-state MRAP is used in the numerical investigation later in chapter 4.

Finally, section 2.3 gives a heuristic algorithm to find the parameters in the 2-state MRAP for given values of mean, variance and lag-1 covariance of the interarrival times. These parameters are needed for computing the queue lengths and other performance measures (see chapter 3). This heuristic algorithm allows us systematically to vary the mean, variance and lag-1 covariance of the interarrival times in order to investigate the effect of the moments and the auto-covariance function on the queueing properties.

2.1 Moments of the m-state Markov Renewal Arrival Process.

Let $\mathfrak{T}^a = \{T_n^a : n = 0, 1, 2, \ldots\}$ be a sequence of arrival times of customers at a queueing system and $0 = T_n^a < T_1^a < T_2^a < T_3^a < T_4^a < \ldots$. The superscript a in all of the subsequent symbols means *arrival*. For example, T_4^a is the arrival time of the 4th customer. Define the interarrival time between the nth and (n+1)st customers as

$$D_{n+1}^a = T_{n+1}^a - T_n^a \ , \quad n = 0, 1, 2, 3, \ldots . \tag{1.1}$$

Then the random process $\mathfrak{D}^a = \{ D_n^a : n = 1, 2, \ldots\}$ is the sequence of interarrival times. Notice that (1.1) defines a one-one mapping between D_{n+1}^a and $T_{n+1}^a - T_n^a$. Consequently, one can use whichever process, $\mathfrak{T}^a$ or $\mathfrak{D}^a$, is appropriate in a given context.

With each T_n^a , the arrival of the nth customer, define a random variable Z_n, $n = 0, 1, 2, 3, \ldots$, which takes values on a finite set $E = \{1, 2, \ldots, m\}$. Z_n represents the *type* of the nth customer. Let $\mathfrak{Z} = \{ Z_n : n = 0, 1, 2, \ldots \}$. Furthermore, assume that

$$P[Z_{n+1}=j,\ D_{n+1}\le t \mid Z_0,\ldots,Z_n;\ T_0^a,\ldots,T_n^a] = P[Z_{n+1}=j, D_{n+1}\le t \mid Z_n] \tag{1.2}$$

for all $n = 0, 1, 2, \ldots, j \in E$, and $t \in \Re_+$. Then the random process $(\mathfrak{Z},\mathfrak{T}^a)$ is a Markov renewal process (see section 1.2). Also assume $(\mathfrak{Z},\mathfrak{T}^a)$ to be homogeneous, for all $i,j \in E$, $t \in \Re_+$,

$$A_{ij}(t) = P[Z_{n+1}=j \ , D_{n+1}\le t \mid Z_n = i] \tag{1.3}$$

is independent of n. The m by m matrix $\mathbf{A}(t)$ with elements $A_{ij}(t)$ is called *the semi-Markov kernel* of the Markov renewal process $(\mathfrak{Z},\mathfrak{T}^a)$.

Let

$$a_{ij} = \lim_{t\to\infty} A_{ij}(t) \tag{1.4}$$

It is easy to see from (2.3) that

$$a_{ij} \geq 0 \quad , \quad \text{and} \quad \sum_{j=1}^{m} a_{ij} = 1 \; . \tag{1.5}$$

Here, the a_{ij}'s are the one-step transition probabilities for *the underlying Markov chain* $\mathfrak{Z}$. The one-step transition matrix for $\mathfrak{Z}$ is then given by the m by m matrix $\mathbf{A}$:

$$\mathbf{A} = [\; a_{ij} : \; i, j \in E] \; . \tag{1.6}$$

It follows from (1.5) that matrix $\mathbf{A}$ is stochastic. If a_{ij}=0 for some pair (i,j), then $A_{ij}(t)$=0 for all $t \in \Re_+$ and define $\frac{A_{ij}(t)}{a_{ij}} = 1$. With this convention, define

$$F_{ij}(t) = P\,[\; D_{n+1} \leq t \mid Z_n = i, \; Z_{n+1} = j \;] = \frac{P\,[Z_{n+1} = j \; , \; D_{n+1} \leq t \mid Z_n = i\,]}{P\,[Z_{n+1} = j \mid Z_n = i\,]} \tag{1.7}$$

or

$$F_{ij}(t) = \frac{A_{ij}(t)}{a_{ij}} \quad , \quad i, j \in E, \;\; t \geq 0. \tag{1.8}$$

$F_{ij}(t)$ is the conditional interarrival distribution between two consecutive arrivals, one of type i and the other of type j. Consequently, the semi-Markov kernel $\mathbf{A}(t)$ of $(\mathfrak{Z}, \mathfrak{T}^a)$ can be written as

$$\mathbf{A}(t) = \begin{bmatrix} a_{11}F_{11}(t) & a_{12}F_{12}(t) & \cdots & a_{1m}F_{1m}(t) \\ a_{21}F_{21}(t) & a_{22}F_{22}(t) & \cdots & a_{2m}F_{2m}(t) \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}F_{m1}(t) & a_{m2}F_{m2}(t) & \cdots & a_{mm}F_{mm}(t) \end{bmatrix} \tag{1.9}$$

and the stochastic matrix $\mathbf{A}$ of the underlying Markov chain $\mathfrak{Z}$ becomes

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} . \tag{1.10}$$

Assume $\mathcal{Z}$ is irreducible, recurrent and let $\pi = [\pi_1, \pi_2, \ldots, \pi_m]$ be the stationary probability vector for the Markov chain $\mathcal{Z}$:

$$\pi . \mathbf{A} = \pi \quad , \quad \pi . \mathbf{e} = 1 \, , \quad 0 \leq \pi_j \leq 1 \, . \tag{1.11}$$

where $\mathbf{e}$ is the column vector of ones. Then, the stationary interarrival time distribution is given by

$$P[\, D_n^a \leq t \,] = \sum_{i=1}^{m} \sum_{j=1}^{m} P[\, D_{n+1}^a \leq t \, ; \, Z_n = i, \, Z_{n+1} = j \,]$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} P[\, Z_{n+1} = j, \, D_{n+1}^a \leq t \mid Z_n = i \,] \, . \, P[Z_n = i \,]$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} \pi_i \cdot A_{ij}(t) \, , \tag{1.12}$$

or

$$P[\, D_n^a \leq t \,] = \pi \, . \, \mathbf{A}(t) \, . \, \mathbf{e} \, . \tag{1.13}$$

To find the p-th moment of the interarrival time D_n^a where $p = 1, 2, 3, \ldots$, let

$$F_{ij}^p = \int_0^\infty t^p . \, dA_{ij}(t) \quad , \; i, j = 1, 2, \ldots, m \, , \tag{1.14}$$

and

$$\mathbf{A}^{(p)} = \begin{bmatrix} a_{11} F_{11}^p & a_{12} F_{12}^p & \cdots & a_{1m} F_{1m}^p \\ a_{21} F_{21}^p & a_{22} F_{22}^p & \cdots & a_{2m} F_{2m}^p \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} F_{m1}^p & a_{m2} F_{m2}^p & \cdots & a_{mm} F_{mm}^p \end{bmatrix} . \tag{1.15}$$

LEMMA 1.16. The mean and the variance of the interarrival time D_n^a are given by

$$E[D_n^a] = \pi \, . \, \mathbf{A}^{(1)} . \, \mathbf{e} \quad , \tag{1.17}$$

and

$$Var\,[D_n^a] = \pi \, . \, \mathbf{A}^{(2)} . \, \mathbf{e} \; - \; E\,[D_n^a]^2 . \tag{1.18}$$

Proof.

$$E\,[D_n^a] = \int_0^\infty t \,.\, d\,P[\,D_n^a \le t\,]$$

$$= \pi \int_0^\infty t \,.\, d\mathbf{A}(t) \,.\, \mathbf{e}$$

$$= \pi \,.\, \mathbf{A}^{(1)} . \,\mathbf{e} \quad .$$

$$E\,[\,(D_n^a)^2\,] = \int_0^\infty t^2 \,.\, d\,P[\,D_n^a \le t\,]$$

$$= \pi \int_0^\infty t^2 \,.\, d\mathbf{A}(t) \,.\, \mathbf{e}$$

$$= \pi \,.\, \mathbf{A}^{(2)} . \,\mathbf{e} \quad . \tag{1.19}$$

$$Var\,[D_n^a] = \pi \,.\, \mathbf{A}^{(2)} . \,\mathbf{e} - E\,[D_n^a]^2. \tag{1.20}$$

□

For $r = 1, 2, 3, \ldots,\ t \in \Re_+,\ u \in \Re_+$, the joint probability distribution of D_n^a and D_{n+r}^a is

$$P[\,D_n^a \le t\,,\, D_{n+r}^a \le u]$$

$$= \sum_{i,j,k,l\,\in\,E}^{m} P[\,D_n^a \le t,\ D_{n+r}^a \le u\,,\, Z_{n-1} = i\,,\, Z_n = j,\, Z_{n+r-1} = k\,,\, Z_{n+r} = l\,]$$

$$= \sum_{i,j,k,l\,\in\,E}^{m} P\,[Z_{n-1} = i].\ P[\,Z_n = j,\, D_n^a \le t \mid Z_{n-1} = i\,]\,.\ P\,[Z_{n+r-1} = k,\, Z_n = j].$$

$$P\,[Z_{n+r} = l,\, D_{n+r}^a \le u \mid\ Z_{n+r-1} = k\,]$$

$$= \sum_{i,j,k,l\,\in\,E}^{m} \pi_i \,.\, A_{ij}(t) \,.\, A_{jk}^{r-1} . \,A_{kl}(u)\,, \tag{1.21}$$

or

$$P[\,D_n^a \le t\,,\, D_{n+r}^a \le u] = \pi \,.\, \mathbf{A}(t)\,.\,\mathbf{A}^{r-1} . \,\mathbf{A}(u)\,.\,\mathbf{e}\,. \tag{1.22}$$

Hence,

$$E[\, D_n^a,\ D_{n+r}^a] = \pi \,.\int_0^\infty t\,.\,d\mathbf{A}(t)\,.\ \mathbf{A}^{r-1}\,.\int_0^\infty u\,.\,d\mathbf{A}(u)\,.\,\mathbf{e}$$

$$= \pi\,.\,\mathbf{A}^{(1)}.\,\mathbf{A}^{r-1}.\,\mathbf{A}^{(1)}.\ \mathbf{e}\ ,\quad r = 1, 2, 3, \dots . \tag{1.23}$$

The auto-covariance function and the auto-correlation function are then given by

$$Cov\,(r) = Cov\,[\, D_n^a,\ D_{n+r}^a] = E[\, D_n^a,\ D_{n+r}^a] - E[\, D_n^a\,]^2\ , \tag{1.24}$$

$$Corr\,(r) = Corr\,[\, D_n^a,\ D_{n+r}^a] = \frac{Cov\,[\, D_n^a,\ D_{n+r}^a\,]}{Var\,[\, D_n^a\,]}\ , \tag{1.25}$$

where $E\,[D_n^a]$, $Var\,[D_n^a]$, $E\,[\, D_n^a,\ D_{n+r}^a]$ are given by (1.17), (1.18), and (1.23), respectively.

2.2 Equivalence Between Markov Renewal Process and Renewal Process.

The two theorems in this section give conditions for which the underlying point process $\mathfrak{T}^a$ or, equivalently, the interval process $\mathfrak{D}^a$ is a renewal process. These theorems are due to Simon and Disney (1984). This section is an excerpt from Disney and Kiessler (1987) section 2.11.

First, the joint interarrival distribution is found using similar arguments to those in (1.21). It is

$$P[\, D_{n+1}^a \le t_1,\ D_{n+2}^a \le t_2, \cdots, D_{n+m}^a \le t_m] = \pi.\mathbf{A}(t_1).\mathbf{A}(t_2)\cdots\mathbf{A}(t_m).\mathbf{e}\ , \tag{2.1}$$

where π is some initial probability vector for the Markov chain $\mathfrak{Z}$.

The next lemma is apparent from (1.13) and (2.1). It simply means

$$P[\ D^a_{n+1} \le t_1,\ D^a_{n+2} \le t_2,\ \cdots,\ D^a_{n+m} \le t_m\ ]$$
$$= P[\ D^a_{n+1} \le t_1\ ] \cdot P[\ D^a_{n+2} \le t_2\ ]\ \cdots\ P[\ D^a_{n+m} \le t_m\ ] \tag{2.2}$$

LEMMA 2.3 *The sequence of interarrival times* $\mathfrak{D}^a_n = \{\ D^a_n\ ;\ n \in \mathbb{N}\ \}$ *has i.i.d. non-negative random variable (i.e.* $\mathfrak{T}^a$ *is a renewal process) if and only if for all* $n \in \mathbb{N}$ *and* $t \in \Re_+$,

$$\pi.\mathbf{A}(t_1).\mathbf{A}(t_2)\cdots\mathbf{A}(t_n).\mathbf{e} = (\pi.\mathbf{A}(t_1).e).(\pi.\mathbf{A}(t_2).\mathbf{e})\cdots(\pi.\mathbf{A}(t_n).\mathbf{e}). \tag{2.4}$$

Then we define *equivalence* between a Markov renewal process and a renewal process.

DEFINITION. Let $(\mathfrak{Z},\mathfrak{T}^a)$ *be a Markov renewal process and let* $\mathfrak{R}=\{T_n\ :\ n \in \mathbb{N}\}$ *be a renewal process with interarrival distribution* $P[\,D_{n+1} = T_{n+1} - T_n \le t] = F(t)$. *The process* $(\mathfrak{Z},\mathfrak{T}^a)$ *is equivalent to the process* $\mathfrak{R}$ *(written* $(\mathfrak{Z},\mathfrak{T}^a) \simeq \mathfrak{R}$*) if*

$$\pi\ .\ \mathbf{A}(t_1)\ .\ \mathbf{A}(t_2)\ \cdots\ \mathbf{A}(t_n)\ .\ \mathbf{e}\ =\ F(t_1)\ .\ F(t_2)\ \cdots\ F(t_n)\ , \tag{2.5}$$

for all $n \in \mathbb{N}$ *and* $t_1,\ t_2,\ \ldots,\ t_n \in \Re_+$*, and for some initial probability vector* π *of* $\mathfrak{Z}$.

Clearly, if $(\mathfrak{Z},\mathfrak{T}^a) \simeq \mathfrak{R})$, then $P\ [\ D_n\ \le\ t\ ]\ =\ \pi.\mathbf{A}(t).\mathrm{e}$ for all $n \in \mathbb{N}$ and $t \in \Re_+$. Below is a simple example of equivalence.

EXAMPLE 2.6 *Suppose* $(\mathfrak{Z},\mathfrak{T}^a)$ is a Markov renewal process and the processes $\mathfrak{Z}$ and $\mathfrak{T}^a$ are independent. Then from (1.7), $F_{ij}(t) = F(t)$ for all $i,\ j \in E$ and $t \in \Re_+$. Therefore,

$$P[\, D^a_{n+1} \le t_1,\ D^a_{n+2} \le t_2,\ \cdots,\ D^a_{n+m} \le t_m \,]$$

$$= \; P[\, D^a_{n+1} \le t_1,\ D^a_{n+2} \le t_2,\ \cdots,\ D^a_{n+m} \le t_m \mid Z_n = i_n, \ldots, Z_{n+m} = i_m \,]$$

$$= \; F(t_1) \,.\, F(t_2) \cdots F(t_m) \,. \qquad (2.7)$$

Thus, $\mathfrak{T}^a$ is a renewal process. Compound Poisson processes discussed in section 1.2 have this property.

The following two theorems give conditions for $(\mathfrak{Z},\mathfrak{T}^a) \simeq \mathfrak{R}$.

THEOREM 2.8. *Let $(\mathfrak{Z},\mathfrak{T}^a)$ be a Markov renewal process; $\mathfrak{R}$ be a renewal process with distribution $F(t)$; and π be an initial probability vector of $\mathfrak{Z}$. If*

$$\pi \,.\, \mathbf{A}(t) = F(t) \,.\, \pi \;, \qquad (2.9)$$

Then $(\mathfrak{Z},\mathfrak{T}^a) \simeq \mathfrak{R}$.

Proof. The proof is clear from repeated substitution of (2.9) into (2.5).

□

Suppose the process $\mathfrak{T}^a$ is stationary and D^a_{n+1} *and* Z_n are independent. The following theorem shows that $\mathfrak{T}^a$ is a renewal process.

THEOREM 2.10. If for all $t \in \Re_+$ and $m \in \mathbf{N}$,

$$\mathbf{A}(t) \,.\, \mathbf{e} = \mathbf{e} \,.\, F(t) \;, \qquad (2.11)$$

Then $(\mathfrak{Z},\mathfrak{T}^a) \simeq \mathfrak{R}$.

Proof. The proof is accomplished by repeated substitution of (2.11) into (2.5).

□

Theorem 2.8 says that if π is a left eigenvector of $\mathbf{A}(t)$ for all $t \in \Re_+$, then $(\mathfrak{Z},\mathfrak{T}^a) \simeq \mathfrak{R}$ and $F(t)$ is the eigenfunction of $\mathbf{A}(t)$ associated with eigenvector π. But this implies that D^a_{n+1} *and* Z_{n+1} are independent. However, T^a_{n+1} may be dependent on

Z_n. Such a case occurs in the celebrated Burke's theorem on departure processes of M/M/s queues.

According to Theorem 2.10, if the row-sum of the semi-Markov kernel $\mathbf{A}(t)$ are the same for every $t \in \Re_+$, then $(\mathfrak{Z},\mathfrak{T}^a) \simeq \mathfrak{R}$, and $F(t)$ is the common value of the row-sums. On the other hand, if the row-sums are all the same, then the time between the nth and (n+1)st transitions does not depend on the state of the process at T_n^a . Notice that Theorem 2.10 is not true for Burke's theorem.

In contrast to example 2.6, Theorem 2.8 and Theorem 2.10 do not imply that it is necessary for $\mathfrak{Z}$ to be independent of $\mathfrak{T}^a$ for the Markov renewal process $(\mathfrak{Z},\mathfrak{T}^a)$ to be equivalent to the renewal process $\mathfrak{T}^a$. They only require that the semi-Markov kernel $\mathbf{A}(t)$ possess certain structures given by (2.9) and (2.11).

Theorems 2.8 and 2.10 are related to the quasi-reversibility of Markovian queueing system. See Disney and Kiessler (1987) for further discussion on this topic.

2.3 Two-State Markov Renewal Arrival Process.

If there are only two types of customers, say, *type 1* and *type 2* ($E = \{1, 2\}$), then the semi-Markov kernel of MRAP $(\mathfrak{Z},\mathfrak{T}^a)$ can be written as

$$\mathbf{A}(t) = \begin{bmatrix} a\ F_{11}(t) & (1\text{-}a)\ F_{12}(t) \\ (1\text{-}b)\ F_{21}(t) & b\ F_{22}(t) \end{bmatrix}, \tag{3.1}$$

and the one-step transition probability matrix of the underlying Markov chain $\mathfrak{Z}$ is

$$\mathbf{A} = \begin{bmatrix} a & (1-a) \\ (1-b) & b \end{bmatrix}, \tag{3.2}$$

where $0 \le a \le 1$, $0 \le b \le 1$ and $F_{ij}(t)$, $i, j = 1, 2$, are any proper distribution functions.

If $a = 0$ and $b \neq 0$ *or* 1, a type 2 customer always follows a type 1 customer. If $a \neq 0$ *or* 1 and $b=0$, a type 1 customer always follows a type 2 customer. In both of these cases, the states of the Markov chain $\mathfrak{Z}$ are aperiodic, positive recurrent. This follows from equation (3.17) and Çinlar (1975b, Theorem 3.2, p.126).

If $a=0$ and $b=0$, the MRAP $(\mathfrak{Z},\mathfrak{T}^a)$ becomes an *alternating renewal arrival process.* The states of the underlying Markov chain $\mathfrak{Z}$ are recurrent and periodic with period 2. Although the stationary distribution of $\mathfrak{Z}$ exists, the limiting distribution of $\mathfrak{Z}$ does not exist; there are 2 possible limiting distributions (see Çinlar (1975b, pp.160-166) for the procedure to compute these limiting distributions).

If $a=1$ or $b=1$ (not both), then in the limit (i.e. $n \to \infty$), either the Markov chain $\mathfrak{Z}$ starts in the absorbing state and stays there or the Markov chain is eventually absorbed. In either cases, the arrival process is ultimately a renewal process. If $a=1$ and $b=1$, the arrival process is always a renewal process. Which one depends on where the process starts. For these reasons, we will assume $a \neq 1$ and $b \neq 1$ which guarantee the states of MRAP $(\mathfrak{Z},\mathfrak{T}^a)$ to be positive recurrent.

Let $\pi = [\ \pi_1\ ,\ \pi_2\ ]$ be the stationary probability vector of the Markov chain $\mathfrak{Z}$ ($\pi.\mathbf{A}=\pi$, $\pi.\mathrm{e}=1$). Then

$$P\,[\ D_n^a \le t\ ] = \pi_1[\ a\ F_{11}(t) + (1\text{-}a)\ F_{12}(t)] + \pi_2[b\ F_{21}(t) + (1\text{-}b)\ F_{22}(t)]\ , \tag{3.3}$$

where

$$\pi_1 = \frac{1-b}{1-\xi}\ , \tag{3.4}$$

$$\pi_2 = \frac{1-a}{1-\xi}\ , \tag{3.5}$$

and

$$\xi = a + b - 1 \; . \qquad (3.6)$$

ξ is the subdominant eigenvalue of the stochastic matrix $\mathbf{A}$ (the dominant eigenvalue equals 1) and $-1 \le \xi < 1$. The marginal interarrival distribution (3.3) is a mixture (convex combination) of four distribution functions $F_{ij}(t)$, $i, j = 1, 2$.

Barbour (1976) and Kelly (1979, pp.77, 80) showed that mixtures of Erlang distributions can approximate the distribution function of any positive random variable. Consequently, we will take $F_{ij}(t)$ to be an Erlang distribution. A further attempt to reduce the number of parameters in the marginal interarrival distribution leads to choosing

$$F_{11}(t) = F_{21}(t) = F_1(t) = \mathit{Erlang}\ (k, \lambda_1) \; ,$$

and

$$F_{12}(t) = F_{22}(t) = F_2(t) = \mathit{Erlang}\ (k, \lambda_2) \; .$$

The semi-Markov kernel (3.1) becomes

$$\mathbf{A}(t) = \begin{bmatrix} a\, F_1(t) & (1\text{-}a)\, F_2(t) \\ (1\text{-}b)\, F_1(t) & b\, F_2(t) \end{bmatrix} , \qquad (3.7)$$

where

$$\frac{dF_i(t)}{dt} = \frac{(\lambda_i k)\,(\lambda_i kt)^{k-1}\, e^{-\lambda_i kt}}{(k-1)\,!} \;, \; t \ge 0 \;, \;\; i = 1, 2 \;, \qquad (3.8)$$

$0 < \lambda_i < \infty$, *k is a positive integer.*

Then

$$P\,[\, D_n^a \le t\,] = \pi_1\, F_1(t) + \pi_2\, F_2(t) \;, \qquad (3.9)$$

since

$$a\, \pi_1 + (1\text{-}b)\, \pi_2 = \pi_1 \; ,$$

$$(1\text{-}a)\ \pi_1 + b\ \pi_2 = \pi_2 \quad . \tag{3.10}$$

Let

$$m_1 = \int_0^\infty t \ \ dF_1(t) = \frac{1}{\lambda_1} \ , \tag{3.11}$$

$$m_2 = \int_0^\infty t \ \ dF_2(t) = \frac{1}{\lambda_2} \ , \tag{3.12}$$

and hence

$$\mathbf{A}^{(1)} = \begin{bmatrix} a\ m_1 & (1-a)\ m_2 \\ (1-b)\ m_1 & b\ m_2 \end{bmatrix} . \tag{3.13}$$

It follows from (1.17) and (1.18),

$$E\,[D_n^a\,] = \pi_1\ m_1 + \pi_2\ m_2 \ , \tag{3.14}$$

$$E\,[\,(D_n^a)^2\,] = (\,\frac{k+1}{k}\,)\,(\,\pi_1\ m_1^2 + \pi_2\ m_2^2\,) \ , \tag{3.15}$$

$$Var\,[D_n^a\,] = \frac{1}{k}\,(\,\pi_1\ m_1^2 + \pi_2\ m_2^2\,) + \pi_1\ \pi_2\ (m_1 - m_2)^2 \ . \tag{3.16}$$

To find the autocovariance function $Cov\ [\ D_n^a,\ D_{n+r}^a\ ]$, we first find

$$\mathbf{A}^{r-1} = \begin{bmatrix} \pi_1+\pi_2\ \xi^{\ r-1} & \pi_2-\pi_2\ \xi^{\ r-1} \\ \pi_1-\pi_1\ \xi^{\ r-1} & \pi_2+\pi_1\ \xi^{\ r-1} \end{bmatrix} . \tag{3.17}$$

$\mathbf{A}^{r-1}$ can be decomposed into 2 parts:

$$\mathbf{A}^{r-1} = \begin{bmatrix} \pi_1 & \pi_2 \\ \pi_1 & \pi_2 \end{bmatrix} + \begin{bmatrix} \pi_2 & -\pi_2 \\ -\pi_1 & \pi_1 \end{bmatrix} \xi^{\ r-1} \tag{3.18}$$

$$= \mathbf{e}\ .\ \pi \ + \begin{bmatrix} \pi_2 \\ -\pi_1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \xi^{\ r-1} \ . \tag{3.19}$$

From (1.23),

$$E[\,D_n^a,\ D_{n+r}^a] = \pi\,.\,\mathbf{A}^{(1)}.\ \mathbf{A}^{r-1}.\ \mathbf{A}^{(1)}.\ \mathbf{e}$$

$$= \pi.\mathbf{A}^{(1)}.\mathbf{e}\ .\ \pi.\mathbf{A}^{(1)}.\mathbf{e} + \pi.\mathbf{A}^{(1)}\begin{bmatrix}\pi_2\\ -\pi_1\end{bmatrix}\begin{bmatrix}1 & -1\end{bmatrix}\mathbf{A}^{(1)}.\mathbf{e}\ \ \xi^{\,r-1}$$

$$= E[\,D_n^a\,]^2 + \pi_1\,\pi_2\,(m_1 - m_2)^2\,\xi^r\,. \tag{3.20}$$

or

$$Cov(r) = Cov[\,D_n^a,\,D_{n+r}^a\,] = E[\,D_n^a,\,D_{n+r}^a\,] - E\,[D_n^a]^2$$

$$= \pi_1\,\pi_2\,(m_1 - m_2)^2\,\xi^r\,, \tag{3.21}$$

and

$$Corr\,(r) = Corr[\,D_n^a,\,D_{n+r}^a\,] = \frac{Cov\,(r)}{Var[\,D_n^a\,]}\,. \tag{3.22}$$

Equation (3.20) shows that the auto-covariance function $Cov(r)$ (and hence the auto-correlation function $Corr(r)$) decreases geometrically with rate ξ ($|\xi|<1$) as the lag (r) increases. If $\xi > 0$, $Cov(r)$ and $Corr(r)$ are strictly positive ; if $\xi < 0$, they are alternating in sign. Notice also that for large k, $corr(r)\simeq\xi^r$ so that when the interarrival times are nearly constant, all of the correlation is given by the second eigenvalue of the matrix $\mathbf{A}$. Other moments play no role.

If $\xi = 0$ (i.e. $a + b = 1$), the interarrival times $\mathfrak{D}^a=\{\ D_n^a : n=1, 2, \ldots\}$ are i.i.d. . ($\mathfrak{T}^a=\{\ T_n^a : n=1, 2, \ldots\}$ is a renewal arrival process). This follows directly from Theorem 2.10. This is a way to model renewal arrival process with 2 different types customers. The interarrival times between any two customers are i.i.d. with $P[\,D_n^a \le t\,] = \pi_1 F_1(t)+\pi_2 F_2(t)$.

If $F_1(t) = F_2(t) = F(t)$ (and hence $m_1=m_2$), not only is the process $\mathfrak{D}^a$ uncorrelated but also the process $\mathfrak{T}^a=\{\ T_n^a : n=1, 2, \ldots\}$ is a renewal process by

Theorem 2.10. Here, D_n^a's are i.i.d. with P [$D_n^a \leq t$] = $F(t)$.

The *coefficient of skewness* (normalized third moment measure) of the interarrival times γ is defined to be

$$\gamma = \frac{E[\,(D_n^a - E[D_n^a])^3]}{\{\,Var\,[D_n^a]^{3/2}\}}\,, \tag{3.23}$$

where

$$E[\,(D_n^a - E[D_n^a])^3\,] = E[\,(D_n^a)^3] - 3\,E\,[D_n^a]\,.Var\,[D_n^a] - E\,[D_n^a]^3\,, \tag{3.24}$$

$$E[\,(D_n^a)^3\,] = \frac{(k+1)(k+2)}{k^2}\,(\,\pi_1\,m_1^3 - \pi_2\,m_2^3\,)\,. \tag{3.25}$$

For ease of notations we will use the following symbols in subsequent sections :

$$z_1 = E\,[D_n^a\,] = \pi_1\,m_1 + \pi_2\,m_2\,, \tag{3.26}$$

$$z_2 = Var\,[D_n^a\,] = \frac{1}{k}\,(\,\pi_1\,m_1^2 + \pi_2\,m_2^2\,) + \pi_1\,\pi_2\,(m_1 - m_2)^2\,, \tag{3.27}$$

$$z_3 = Cov(1) = Cov\,[\,D_n^a,\,D_{n+1}^a\,] = \pi_1\,\pi_2\,(m_1 - m_2)^2\,\xi\,, \tag{3.28}$$

$$\gamma = \frac{(k+1)(k+2)}{k^2}\,(\,\pi_1\,m_1^3 - \pi_2\,m_2^3\,) - 3\,z_1 z_2 - z_1^3\,, \tag{3.29}$$

There are 5 parameters in this 2-type MRAP; they are a, b, m_1, m_2, and k. The numerical investigation in chapter 4 uses these 5 parameters to investigate systematically the effects of the moments and the auto-correlation function of the MRAP on the queueing properties. There, we systematically prescribe values for z_1, z_2 and z_3. Then the sets of parameters { a, b, m_1, m_2, k} that satisfy equations (3.26)−(3.28) are computed using the algorithm developed in the next section. (γ is computed using (3.29) after the 5 parameters are found.) These 5 parameters are used as input to the algorithm for solving MR/M/1 queues (see chapter 4).

Since there are 5 unknowns and only 3 nonlinear equations given, the set of solutions $\{ a, b, m_1, m_2, k\}$ has 2 degrees of freedom. Section 2.4 explains the heuristic algorithm used to find a, b, m_1, m_2, and k.

2.4 A Heuristic Nonlinear Algorithm.

Let us restate the problem at the end of section 2.3 . Given

$$z_1 = E[D_n^a] = \pi_1 m_1 + \pi_2 m_2 , \tag{4.1}$$

$$z_2 = Var[D_n^a] = \frac{1}{k} (\pi_1 m_1^2 + \pi_2 m_2^2) + \pi_1 \pi_2 (m_1 - m_2)^2 , \tag{4.2}$$

$$z_3 = Cov(1) = Cov[D_n^a, D_{n+1}^a] = \pi_1 \pi_2 (m_1 - m_2)^2 \xi , \tag{4.3}$$

find all 5-parameter solutions $\{ a, b, m_1, m_2, k\}$ that satisfy equations (4.1) - (4.3). Here,

$$\pi_1 = \frac{1-b}{1-\xi} , \tag{4.4}$$

$$\pi_2 = \frac{1-a}{1-\xi} , \tag{4.5}$$

and

$$\xi = a + b - 1 . \tag{4.6}$$

Realizing that there are 2 degrees of freedom, we choose to fix 2 of the 5 parameters: k and m_2, for the following reason:

i. k is a positive integer variable for which it will be difficult to solve from the mixed

nonlinear system of equations (4.1)−(4.3).

ii. Since z_1 is a convex combination of m_1 and m_2 (equation (4.1)), then either $0 < m_2 < z_1 < m_1$ or $0 < m_1 < z_1 < m_2$. It turns out either choice will yield indentical solutions. It is a matter of interchanging (renaming) the two types of arrivals. So we arbitrarily choose

$$0 < m_2 < z_1 < m_1 . \tag{4.7}$$

Two cases will be considered separately according to whether $z_3 \geq 0$ or $z_3 < 0$. The case of negative lag-1 covariance ($z_3 < 0$) requires additional constraints on m_2.

CASE $z_3 \geq 0$.

Fix m_2 such that $0 < m_2 < z_1 < m_1$. Next, we find an upper bound and a lower bound for the positive integer k. From (4.1) and (4.5),

$$\pi_1 (m_1 - m_2) = z_1 - m_2 . \tag{4.8}$$

Hence,

$$\pi_1 = \frac{z_1 - m_2}{m_1 - m_2} , \tag{4.9}$$

and

$$\pi_2 = (1 - \pi_1) = \frac{m_1 - z_1}{m_1 - m_2} . \tag{4.10}$$

Substituting (4.3) into (4.2) gives

$$k (z_2 - \frac{z_3}{\xi}) = \pi_1 m_1^2 + \pi_2 m_2^2 , \tag{4.11}$$

or

$$m_1 (z_1 - m_2) = k (z_2 - \frac{z_3}{\xi}) - m_2 z_1 . \tag{4.12}$$

Substituting (4.9) and (4.10) into (4.3) gives

$$(z_1 - m_2) (m_1 - z_1) = \frac{z_3}{\xi} , \tag{4.13}$$

or

$$m_1 (z_1 - m_2) = \frac{z_3}{\xi} + z_1^2 - m_2 z_1 . \tag{4.14}$$

Now equate (4.12) and (4.13) :

$$\frac{z_3}{\xi} + z_1^2 = k (z_2 - \frac{z_3}{\xi}) . \tag{4.15}$$

Letting

$$y = \frac{z_3}{\xi} , \tag{4.16}$$

we have

$$y = \frac{k z_2 - z_1^2}{k + 1} . \tag{4.17}$$

An upper bound for y ,

$$y < z_2 \tag{4.18}$$

is obtained from (4.11) and a lower bound for y,

$$|z_3| < y \tag{4.19}$$

is obtained from (4.16) since $|\xi| < 1$. Notice y is always positive because z_3 *and* ξ have the same sign.

Using (4.17), we found that inequality (4.18) is trivially true and inequality (2.19) yields

$$k > \frac{|z_3| + z_1^2}{z_2 - |z_3|} . \tag{4.20}$$

The denominator in (4.20) is positive because $\left|\frac{z_3}{z_2}\right| < 1$. This can be verified directly using (4.2) and (4.3). Since k is a positive integer we choose $\left\lfloor \frac{|z_3| + z_1^2}{z_2 - |z_3|} + 1 \right\rfloor$ as the lower bound for k. Notice that this lower bound uses only z_1 , z_2 and z_3. There is no upper bound for k because inequality (4.18) is trivially true.

Once k and m_2 are fixed, the other parameters are found using

$$m_1 = \frac{1}{(z_1 - m_2)} \left[k\,(z_2 - y) - m_2\, z_1 \right] , \tag{4.21}$$

$$\pi_1 = \frac{z_1 - m_2}{m_1 - m_2} , \tag{4.22}$$

$$\pi_2 = (1 - \pi_1) = \frac{m_1 - z_1}{m_1 - m_2} , \tag{4.23}$$

$$a = \pi_1 + \pi_2\, \xi \quad , \tag{4.24}$$

$$b = (1-a) = \pi_2 + \pi_1\, \xi \ . \tag{4.25}$$

(4.21) is obtained from (4.12) ; (4.24) and (4.25) are obvious from (3.17).

CASE $z_3 < 0$.

All the formulas derived for the case $z_3 \geq 0$ are also valid for the case $z_3 < 0$ except the condition $0 \leq a < 1$ and $0 \leq b < 1$ may be violated. Extra conditions must be imposed on m_2 to ensure the above two conditions are satisfied. Since

$$0 \leq a = \pi_1 + \pi_2\, \xi \ < 1 \ , \tag{4.26}$$

$$0 \leq b = \pi_2 + \pi_1\, \xi \ < 1 \ , \tag{4.27}$$

and

$$-1 < \xi < 0 \ , \tag{4.28}$$

$$-\xi \leq \frac{\pi_1}{\pi_2} < -\frac{1}{\xi} \ . \tag{4.29}$$

From (4.22) and (4.23),

$$-\xi \leq \frac{z_1 - m_2}{m_1 - z_1} < -\frac{1}{\xi} \ . \tag{4.30}$$

Substituting (4.13) into (4.30) we get

$$-z_3 \leq (z_1 - m_2)^2 < -\frac{z_3}{\xi^2} \ . \tag{4.31}$$

Hence,

$$z_1 - \frac{\sqrt{-z_3}}{|\xi|} < m_2 \leq z_1 - \sqrt{-z_3} \ . \tag{4.32}$$

Since $m_2 > 0$, the refined bound for m_2 becomes

$$max \{ 0 , (z_1 - \frac{\sqrt{-z_3}}{|\xi|}) \} < m_2 \leq z_1 - \sqrt{-z_3} \tag{4.33}$$

instead of $0 < m_2 < z_1$.

We summarize all the results of this section in the following algorithm :

ALGORITHM 4.34 .

Given z_1, z_2, z_3 *(the mean, variance, lag-1 covariance of the interarrival times), and* $z_2 > |z_3|$, $z_1 > 0$, $z_2 > 0$. *The 5 parameters* a, b, m_1, m_2, k *are computed as follows :*

Choose an integer k such that $\left\lfloor \frac{|z_3| + z_1^2}{z_2 - |z_3|} + 1 \right\rfloor \leq k < \infty$. *(4.35)*

Compute: $$y = \frac{k\, z_2 - z_1^2}{k + 1} \tag{4.36}$$

$$\xi = \frac{z_3}{y} \tag{4.37}$$

If $z_3 \geq 0$, *choose* m_2 *such that*

$$0 < m_2 < z_1 \tag{4.38}$$

else choose m_2 *such that*

$$max \{ 0 , (z_1 - \frac{\sqrt{-z_3}}{| \xi |}) \} < m_2 \leq z_1 - \sqrt{-z_3} \tag{4.39}$$

endif

Compute : $$m_1 = \frac{1}{(z_1 - m_2)} [k (z_2 - y) - m_2\, z_1] \tag{4.40}$$

$$\pi_1 = \frac{z_1 - m_2}{m_1 - m_2} \tag{4.41}$$

$$\pi_2 = (1 - \pi_1) = \frac{m_1 - z_1}{m_1 - m_2} \tag{4.42}$$

$$a = \pi_1 + \pi_2\, \xi \tag{4.43}$$

$$b = (1-a) = \pi_2 + \pi_1\, \xi \tag{4.44}$$

End.

This algorithm is implemented as the FORTRAN main program listed in the Appendix A.

Usually it is easier to work with unitless entities such as the squared coefficient of variation (*scv*) and the coefficient of correlation (*corr. coef.*) where

$$scv = \frac{z_2}{z_1^2} , \tag{4.45}$$

and

$$corr.\ coef. = \frac{z_3}{z_2} . \tag{4.46}$$

The computer programs in Appendix A use *scv*, *corr. coef.* and ρ, the traffic intensity, as input data. We choose to normalize the mean interarrival time such that $z_1 = 1$.

In chapter 4 we will vary z_1, z_2 and z_3 systematically in order to investigate the effects of the moments and the auto-correlation function of the interarrival times on the mean queue length and other performance measures. For every prescribed values of z_1, z_2, and z_3, Algorithm 4.35 will be used to find a, b, m_1, m_2 and k . These 5 parameters are then used as input to the MR/M/1 queue algorithm developed in chapter 3.

Chapter 3

MR/M/1 QUEUE

3.0 Introduction.

In this chapter we consider a single server queue where the arrival process is the general m-state Markov renewal process (see section 2.1) and the service time distribution is exponential with rate μ. The steady-state queue length distribution embedded at arrival times and its moments are derived in section 3.1. The distributions and the expected values of the waiting time and the sojourn time are given in section 3.2.

By considering the arrival times as stopping times, the queue length process observed at arbitrary times is a semi-regenerative process. Consequently, the queue length distribution at arbitrary times (the continuous time queue length distribution) and its moment are found using the results on semi-regenerative processes given in Çinlar (1975b, chapter 10 section 6). The results in section 3.1–3.2 are known as the *customer's properties;* where as the results in section 3.3 are known as the *time properties.*

Section 3.4 gives the relationships between customer's properties and time properties of the MR/M/1 queues. Some of these results are known as the Little's formulas. Finally, section 3.5 outlines the computational procedure to compute the steady-state queue length distribution of the MR/M/1 queues.

With the exception of section 3.4, the content of this chapter follows the works of Neuts (1978a, 1981). The first published results on MR/M/1 queues is found in Çinlar (1967). We choose to use the matrix-geometric approach of Neuts simply because his results are computable.

3.1 The Queue Length Process Embedded at Arrival Times.

Consider a single server queueing system. Assume the service times are exponentially distributed with rate μ. Let the arrival process be the m-state Markov renewal process $(\mathfrak{Z},\mathfrak{T}^a)$ with semi-Markov kernel $\mathbf{A}(t)$ given by (2.1.9). This Markov renewal arrival process (MRAP) was discussed in section 2.1. We will assume that the states of MRAP $(\mathfrak{Z},\mathfrak{T}^a)$ are recurrent and aperiodic (see Çinlar 1975b, pp.321-323). Furthermore, let

$$\mathbf{A}^*(s) = \int_0^\infty e^{-st}\, d\mathbf{A}(t) \ , \tag{1.1}$$

be the Laplace-Stieltjes transform of the semi-Markov kernel $\mathbf{A}(t)$. Then, the transition matrix $\mathbf{A}$ (see (2.1.10)) of the underlying Markov chain $\mathfrak{Z}$ is given by

$$\mathbf{A} = \lim_{t\to\infty} \mathbf{A}(t) = \mathbf{A}^*(0) \ . \tag{1.2}$$

Let $\pi = [\ \pi_1, \pi_2, \ldots, \pi_m\ ]$ be the stationary probability vector of the Markov chain $\mathfrak{Z}$ (see (2.1.11)).

Let $N(t)$ be the number of the customers in the queueing system at time t. We write $N(t)=j$ if there are j customers in the system at time t (including the one in the service center); $j = 1, 2, 3, \ldots$. The limiting probability distribution for $N(t)$ will be derived later in section 3.3.

Let

$$N_n^a = N(T_n^a - 0)\ , \ n=0, 1, 2, \ldots, \tag{1.3}$$

be the number of customers in the system *just before the nth customer arrives.* In other words, N_n^a is the number of customers in the system seen by the nth customer when it arrives. Notice that *the queue length process embedded at arrival times* $\mathcal{N}^a=\{N_n^a : n=0, 1, 2, \ldots\}$ is not Markov because the arrival process $(\mathcal{Z},\mathcal{T}^a)$ is Markov renewal. The two-tuple process $(\mathcal{N}^a,\mathcal{T}^a)$ is not Markov renewal. But $(\mathcal{N}^a,\mathcal{Z};\mathcal{T}^a)$ is a Markov renewal process.

THEOREM 1.4. The stochastic process $(\mathcal{N}^a,\mathcal{Z};\mathcal{T}^a)$ is a (two-dimensional) Markov renewal process with the state space $\{0,1,2,\ldots\}\times\{1,2,\ldots,m\}$.

Proof.

$$P[\ N_{n+1}^a=j,\ Z_{n+1}=l\ ;\ D_{n+1}^a\le t\ |\ N_o^a,\ldots,N_n^a;\ Z_0,\ldots,Z_n;\ D_0^a,\ldots,D_n^a\ ]$$

$$= \int_0^t P[\ N_{n+1}^a=j\ |\ N_o^a,\ldots,N_n^a;\ Z_0,\ldots,Z_n,\ Z_{n+1}=l\ ;\ D_0^a,\ldots,D_n^a,\ D_{n+1}^a=y\ ]$$

$$\cdot \frac{d}{dy}P[\ Z_{n+1}=l\ ;\ D_{n+1}^a\le y\ |\ N_o^a,\ldots,N_n^a;\ Z_0,\ldots,Z_n;\ D_0^a,\ldots,D_n^a\ ]. \tag{1.5}$$

But N_{n+1}^a depends only on N_n^a and D_{n+1}^a; and D_{n+1}^a depends only on Z_n and Z_{n+1} because $(\mathcal{Z},\mathcal{T}^a)$ is Markov renewal. Thus

$$P[\ N_{n+1}^a=j,\ Z_{n+1}=l\ ;\ D_{n+1}^a\le t\ |\ N_o^a,\ldots,N_n^a;\ Z_0,\ldots,Z_n;\ D_0^a,\ldots,D_n^a\ ]$$

$$= \int_0^t P[\ N_{n+1}^a=j\ |\ N_n^a,\ D_{n+1}^a=y\ ]\cdot\frac{d}{dy}P[\ Z_{n+1}=l\ ,\ D_{n+1}^a\le y\ |\ Z_n\ ]$$

$$= P[\ N_{n+1}^a=j,\ Z_{n+1}=l\ ;\ D_{n+1}^a\le t\ |\ N_n^a,\ Z_n]\ . \tag{1.6}$$

This completes the proof.

□

The following Corollary is a direct consequence of Theorem 1.4.

COROLLARY 1.7. The stochastic process $(\mathcal{N}^a, \mathcal{Z})$ *is a (two-dimensional) Markov chain with state space* $\{0,1,2,\ldots\}\times\{1,2,\ldots,m\}$. *Its one-step transition probabilities are given by*

$$P[\, N^a_{n+1}=j,\ Z_{n+1}=l \mid N^a_n=i,\ Z_n=k]$$

$$= \lim_{t\to\infty} P[\, N^a_{n+1}=j,\ Z_{n+1}=l\ ;\ D^a_{n+1}\le t \mid N^a_n=i,\ Z_n=k]\ . \tag{1.8}$$

Once the transition matrix $\mathbf{P}$ *of the Markov chain* $(\mathcal{N}^a,\mathcal{Z})$ is found, the joint stationary probabilities of $(\mathcal{N}^a,\mathcal{Z})$ can be computed. The marginal stationary queue length probabilities of N^a_n are obtained by summing (1.18) over Z_n.

Equation (1.6) provides a formula for computing the semi-Markov kernel of $(\mathcal{N}^a,\mathcal{Z};\ \mathcal{T}^a)$. For $i, j = 0, 1, 2, \ldots$, and $t\in\Re_+$, let

$$G_{ij}(t) = P[\, N^a_{n+1}= j \mid N^a_n=i\ ;\ D^a_{n+1}=t\,]\ . \tag{1.9}$$

Then

$$G_{ij}(t) = \begin{cases} \sum_{n=i+1}^{\infty} q_n(t) & \text{if } j=0\ ,\quad i=0, 1, 2, \ldots, \\ q_{i-j+1}(t) & \text{if } j\ge 1\ ,\quad i+1\ge j\ , \\ 0 & \text{otherwise,} \end{cases} \tag{1.10}$$

where

$$q_n(t) = \frac{(\mu t)^n\, e^{-\mu t}}{n\,!}\ ,\quad n = 0, 1, 2, \ldots, \tag{1.11}$$

is the probability that n customers are served during an interarrival time of length t.

Define the matrix $\mathbf{G}(t)$ as the matrix with elements $G_{ij}(t)$:

$$\mathbf{G}(t) = \begin{bmatrix} \sum_{k=1}^{\infty} q_k(t) & q_0(t) & \mathbf{0} & \mathbf{0} & \cdots \\ \sum_{k=2}^{\infty} q_k(t) & q_1(t) & q_0(t) & \mathbf{0} & \cdots \\ \sum_{k=3}^{\infty} q_k(t) & q_2(t) & q_1(t) & q_0(t) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{1.12}$$

Let

$$P_{i,k\,;\,j,l}(t) = P\,[\,N^a_{n+1}=j,\ Z_{n+1}=l\,;\ D^a_{n+1}\le t \mid N^a_n=i\,,\ Z_n=k\,]\,. \tag{1.13}$$

Then from (1.6), (1.9) and (2.1.3),

$$P_{i,k\,;\,j,l}(t) = \int_0^t G_{ij}(y)\ d\,A_{kl}(y)\,. \tag{1.14}$$

Now, order the state space of $(\mathcal{N}^a,\mathcal{Z})$ lexicographically and let $\mathbf{P}(t)$ be the semi-Markov kernel of $(\mathcal{N}^a,\mathcal{Z};\ \mathcal{T}^a)$. Then

$$\mathbf{P}(t) = \int_0^t \mathbf{G}(y) \otimes d\mathbf{A}(t)\,. \tag{1.15}$$

The matrices $\mathbf{G}(y)$ and $\mathbf{A}(y)$ are given by (1.12) and (2.1.9), respectively, and $\otimes$ denotes the *Kronecker product* of matrices.

The infinite dimensional matrix $\mathbf{P}(t)$ can be partitioned into *m by m* blocks as

$$\mathbf{P}(t) = \begin{bmatrix} \mathbf{B}_0(t) & \mathbf{A}_0(t) & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{B}_1(t) & \mathbf{A}_1(t) & \mathbf{A}_0(t) & \mathbf{0} & \cdots \\ \mathbf{B}_2(t) & \mathbf{A}_2(t) & \mathbf{A}_1(t) & \mathbf{A}_0(t) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{1.16}$$

The m by m block matrices $\mathbf{A}_n(t)$ and $\mathbf{B}_n(t)$, n=0, 1, 2, ..., are given by

$$\mathbf{A}_n(t) = \int_0^t q_n(y).\ d\mathbf{A}(y)\ , \tag{1.17}$$

$$\mathbf{B}_n(t) = \int_0^t \sum_{k=n+1}^{\infty} q_k(y).\ d\mathbf{A}(y)\ . \tag{1.18}$$

Consequently, the transition matrix $\mathbf{P}$ of the Markov chain $(\mathcal{N}^a, \mathcal{Z})$ is

$$\mathbf{P} = \lim_{t \to \infty} \mathbf{P}(t)\ , \tag{1.19}$$

or

$$\mathbf{P} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{A}_0 & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{A}_0 & \mathbf{0} & \cdots \\ \mathbf{B}_2 & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{1.20}$$

where

$$\mathbf{A}_n = \lim_{t \to \infty} \mathbf{A}_n(t) = \lim_{t \to \infty} \int_0^t q_n(y)\ .\ d\mathbf{A}(y)$$

$$= \int_0^\infty \frac{e^{-\mu t}(\mu t)^n}{n\,!}.\ d\mathbf{A}(t) \tag{1.21}$$

and

$$\mathbf{B}_n = \lim_{t\to\infty} \mathbf{B}_n(t) = \sum_{k=n+1}^{\infty} \mathbf{A}_k \quad . \tag{1.22}$$

If $E[D_n^a] \neq 0$ and $\mu \neq 0$, the Markov chain $(\mathcal{N}^a, \mathcal{Z})$ is irreducible. Theorem 1.47 will give the conditions for which this Markov chain is positive recurrent.

Let

$$\mathbf{A} = \sum_{n=0}^{\infty} \mathbf{A}_n \quad . \tag{1.23}$$

Then

$$\mathbf{B}_n = \mathbf{A} - \sum_{k=0}^{n} \mathbf{A}_k \quad . \tag{1.24}$$

LEMMA 1.25 . The m by m matrix $\mathbf{A}$ *defined by* (1.23) *is the same matrix as the transition matrix for the embedded Markov chain* $\mathcal{Z}$ *of the Markov renewal arrival process* $(\mathcal{Z}, \mathcal{T}^a)$ *in* (2.1.10). *Hence, the matrix* $\mathbf{A}$ *is stochastic.*

Proof. The proof is immediate from (1.23) and from the fact that

$$P[\, Z_n \,] = \sum_{i=0}^{\infty} P[\, N_n^a = i \,,\, Z_n \,] \quad . \tag{1.26}$$

Notice that the transition matrix $\mathbf{P}$ given in (1.20) has the same structure as the transition matrix of the GI/M/1 queue with the exception that the elements of $\mathbf{P}$ are m by m block matrices instead of scalars.

Stationary Queue Length Probabilities.

Let $\mathbf{x}$ be the stationary probability vector for the Markov chain $(\mathcal{N}^a, \mathcal{Z})$ and partition $\mathbf{x}$ as

$$\mathbf{x} = [\ \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots\ ]\ , \tag{1.27}$$

where

$$\mathbf{x}_n = [\ \mathbf{x}_{n1}, \mathbf{x}_{n2}, \dots, \mathbf{x}_{nm}\ ]\ , \quad n = 0, 1, 2, \dots\ . \tag{1.28}$$

Here, $\mathbf{x}_{nj}$ is the stationary probability that there are n customer in the system and the incoming customer is of type j. The stationary probability vector $\mathbf{x}$ is found by solving the infinite system of linear equations

$$\mathbf{x} \,.\, \mathbf{P} = \mathbf{x}\ \ ,\ \ \mathbf{x} \,.\, \mathbf{e} = 1\ , \tag{1.29}$$

which can be partitioned into

$$\mathbf{x}_0 = \sum_{k=0}^{\infty} \mathbf{x}_k . \mathbf{B}_k\ , \tag{1.30}$$

$$\mathbf{x}_n = \sum_{k=0}^{\infty} \mathbf{x}_{n+k-1} .\ \mathbf{A}_k\ , \quad n = 1, 2, 3, \dots\ , \tag{1.31}$$

$$\sum_{k=0}^{\infty} \mathbf{x}_k \,.\, \mathbf{e} = 1\ . \tag{1.32}$$

It is easy to verify that $\mathbf{x}_n$ (1.29) or, equivalently, in (1.30)-(1.32) has a *matrix-geometric solution :*

$$\mathbf{x}_{n+1} = \mathbf{x}_n . \mathbf{R}\ , \quad n = 0, 1, 2, \dots\ , \tag{1.33}$$

or

$$\mathbf{x}_n = \mathbf{x}_0 . \mathbf{R}^n\ , \quad n = 1, 2, \dots\ , \tag{1.34}$$

and the m by m matrix $\mathbf{R}$ is *the minimal nonnegative solution* to the matrix polynomial equation

$$\mathbf{R} = \sum_{n=0}^{\infty} \mathbf{R}^n .\ \mathbf{A}_n\ . \tag{1.35}$$

Below we restate some of the properties of the matrix-geometric solution from Neuts (1981).

LEMMA 1.36 . Let η be the largest (in modulus) eigenvalue of $\mathbf{R}$*. Then η is real, positive and $\eta<1$. Thus, sp(*$\mathbf{R}$*)<1 , where sp(*$\mathbf{R}$*) is the spectral radius of the matrix* $\mathbf{R}$ *. This implies that all eigenvalues of* $\mathbf{R}$ *are inside the unit disk. This eigenvalue η is also called the caudal characteristic of the MR/M/1 queue.*

Let

$$\mathbf{B}[\mathbf{R}] = \sum_{n=0}^{\infty} \mathbf{R}^n . \mathbf{B}_n . \tag{1.37}$$

LEMMA 1.38 . The initial probability vector $\mathbf{x}_0$ *is found by solving the linear equations*

$$\mathbf{x}_0 = \mathbf{x}_0 . \mathbf{B}[\mathbf{R}] \quad , \tag{1.39}$$

and is normalized by

$$\mathbf{x}_0 . (\mathbf{I}-\mathbf{R})^{-1} . \mathbf{e} = 1 . \tag{1.40}$$

Moreover, $\mathbf{B}[\mathbf{R}]$ *is stochastic.*

Proof. Equation (1.39) follows from (1.30) and (1.34). Equation (1.40) comes from the fact $\mathbf{x}.\mathbf{e}=1$, and in the partitioned form

$$\mathbf{x} . \mathbf{e} = \sum_{n=0}^{\infty} \mathbf{x}_n . \mathbf{e} = \sum_{n=0}^{\infty} \mathbf{x}_0 . \mathbf{R}^n . \mathbf{e} = \mathbf{x}_0 . (\mathbf{I}-\mathbf{R})^{-1} . \mathbf{e} = 1 . \tag{1.41}$$

The inverse matrix $(\mathbf{I}-\mathbf{R})^{-1}$ exists and is unique because *sp(*$\mathbf{R}$)<1 (see Lemma 1.36). The fact that $\mathbf{B}[\mathbf{R}]$ is stochastic can be easily established from (1.37), (1.24) and Lemma 1.25. □

The following lemma gives the explicit form of $\mathbf{x}_0$.

LEMMA 1.42 . The initial probability $\mathbf{x}_0$ is given by

$$\mathbf{x}_0 = \pi(\mathbf{I}-\mathbf{R}) \quad . \tag{1.43}$$

Proof. From (1.26)

$$\pi = \sum_{n=0}^{\infty} \mathbf{x}_n \quad . \tag{1.44}$$

Hence,

$$\pi = \sum_{n=0}^{\infty} \mathbf{x}_0 . \mathbf{R}^n . \mathbf{e} = \mathbf{x}_0 .(\mathbf{I}-\mathbf{R})^{-1} \ .$$

The conclusion is obvious.

□

We observe from (1.24) that

$$\sum_{k=0}^{\infty} \mathbf{B}_k = \sum_{k=0}^{\infty} k \, . \, \mathbf{A}_k \ . \tag{1.45}$$

Now define the column vector β as

$$\beta = \sum_{k=0}^{\infty} \mathbf{B}_k . \, \mathbf{e} = \sum_{k=0}^{\infty} k \, . \, \mathbf{A}_k . \, \mathbf{e} \quad . \tag{1.46}$$

In most cases β can be found explicitly from (1.46). The following theorem is the main theorem of Neuts (1981, p.19).

THEOREM 1.47 . If the matrix $\mathbf{A}$ is irreducible, the irreducible Markov chain $(\mathcal{N}^a, \mathcal{Z})$ with transition matrix $\mathbf{P}$ is positive recurrent if and only if

$$\pi . \beta > 1 \quad , \tag{1.48}$$

and the stochastic matrix $\mathbf{B}[\mathbf{R}]$ *has a strictly positive invariant left eigenvector* $\mathbf{x}_0$. *The partitioned stationary vector* $\mathbf{x}$ of $\mathbf{P}$ is given by

$$\mathbf{x}_n = \mathbf{x}_0.\mathbf{R}^n \quad , \quad n = 1, 2, \ldots \quad (1.49)$$

where

$$\mathbf{x}_0 \;=\; \pi\,(\mathbf{I}-\mathbf{R})\,. \quad (1.50)$$

By virtue of Lemma 1.42, we have avoided using (1.39) and (1.40) to compute $\mathbf{x}_0$ from $\mathbf{B}[\mathbf{R}]$.

LEMMA 1.51 .

$$\pi.\beta = \mu\;.\;E\,[\;D_n^a\;] = \frac{E\,[\;D_n^a\;]}{E\,[\;V_n\;]} = \mu\;.\;\pi\;.\;\mathbf{A}^{(1)}.\;\mathbf{e}\quad , \quad (1.52)$$

where D_n^a *is the interarrival time,* V_n *is the service time, and* $\mathbf{A}^{(1)}$ *is given by* (2.1.15). *Hence, the traffic intensity* ρ *is given by*

$$\rho = \frac{1}{\pi.\beta} = \frac{1}{\mu.\pi.\mathbf{A}^{(1)}.\;\mathbf{e}}\quad . \quad (1.53)$$

Proof.

$$\beta \;= \sum_{k=0}^{\infty} k\;.\;\mathbf{A}_k.\;\mathbf{e} = \sum_{k=0}^{\infty} k\;.\int_0^{\infty} \frac{(\mu t)^k\; e^{-\mu t}}{k\,!}\; d\mathbf{A}(t)\;.\;\mathbf{e}$$

$$= \int_0^{\infty} (\mu t) \sum_{k=1}^{\infty} \frac{(\mu t)^{k-1}\; e^{-\mu t}}{(k-1)\,!}\; d\mathbf{A}(t)\;.\;\mathbf{e}$$

$$= \int_0^{\infty} \mu\;.\;t\; d\mathbf{A}(t)\;.\;\mathbf{e}$$

$$= \mu \int_0^{\infty} t\; d\mathbf{A}(t)\;.\;\mathbf{e}\;.$$

The rest of the proof follows from (2.1.17).

□

Theorem 1.47 requires that the traffic intensity $\rho < 1$ for $(\mathcal{N}^a,\mathcal{Z})$ to be positive recurrent. Moreover, if $\mathbf{A}$ is aperiodic, the stationary probability vector $\mathbf{x}$ is also the limiting (steady state) probability vector of $(\mathcal{N}^a,\mathcal{Z})$. On the other hand, if $\mathbf{A}$ is periodic with period δ, then $\mathbf{P}$ is also periodic with period δ. The limiting distribution of $(\mathcal{N}^a,\mathcal{Z})$ does not exist; there are δ possible limiting distributions (see Çinlar (1975b, pp.160-166) for the procedure to compute these limiting distributions).

LEMMA 1.54 . The matrix polynomial equation $\mathbf{R}=\sum_{n=0}^{\infty}\mathbf{R}^n.\mathbf{A}_n$ *is identical to the matrix functional equation*

$$\mathbf{R} = \int_0^\infty e^{-\mu(\mathbf{I}-\mathbf{R})t}\, d\mathbf{A}(t) = \mathbf{A}^*(\ \mu(\mathbf{I}-\mathbf{R})\ )\ , \tag{1.55}$$

where $\mathbf{A}^*(t)$ *is the Laplace-Stieltjes transform of* $\mathbf{A}(t)$.

Proof.

$$\begin{aligned}
\mathbf{R} &= \sum_{n=0}^{\infty}\mathbf{R}^n\int_0^\infty \frac{(\mu t)^n\, e^{-\mu t}}{n\,!}\, d\mathbf{A}(t) \\
&= \int_0^\infty e^{-\mu t\mathbf{I}} \sum_{n=0}^{\infty}\frac{(\mu t\mathbf{R})^n}{n\,!}\, d\mathbf{A}(t) \\
&= \int_0^\infty e^{-\mu t\mathbf{I}}\, e^{\mu t\mathbf{R}}\, d\mathbf{A}(t) \\
&= \int_0^\infty e^{-\mu t(\mathbf{I}-\mathbf{R})}\, d\mathbf{A}(t) \\
&= \mathbf{A}^*(\ \mu(\mathbf{I}-\mathbf{R})\ )\ .
\end{aligned}$$

This completes the proof.

□

Expected Queue Length Seen By The Arriving Customers.

From (1.49) and (1.50), the (joint) limiting queue length probability vector at arrival times is given by

$$\mathbf{x}_n = \pi\,(\mathbf{I}-\mathbf{R})\,\mathbf{R}^n \quad , \quad n = 0, 1, 2, \ldots , \tag{1.56}$$

and the marginal queue length probabilities are given by

$$P[\, N_n^a = i \,] = \pi\,(\mathbf{I}-\mathbf{R})\,\mathbf{R}^i . \mathbf{e} \quad , \quad i = 0, 1, 2, \ldots . \tag{1.57}$$

Theorem 1.58 . Let L^a be the expected number of customers in the system seen by arriving customers and L_q^a be the corresponding expected number of customers waiting in the queue. Then

$$L^a = E[\, N_n^a \,] = \pi\,(\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{R} \,.\, \mathbf{e} \quad , \tag{1.59}$$

$$L_q^a = \pi\,(\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{R}^2 \,.\, \mathbf{e} \quad . \tag{1.60}$$

Also

$$E[\,(N_n^a)^2] = \pi\,(\mathbf{I}-\mathbf{R})^{-2}\,\mathbf{R}\,(\mathbf{I}+\mathbf{R}) \,.\, \mathbf{e} \quad , \tag{1.61}$$

$$Var[\, N_n^a \,] = E[\,(N_n^a)^2\,] - E[\, N_n^a \,]^2 \quad . \tag{1.62}$$

Proof. It can be shown using the spectral representation of the matrix $\mathbf{R}$ (see Çinlar (1975b, pp.364-370)), that the matrices $\mathbf{R}$, $(\mathbf{I}-\mathbf{R})^{-1}$ and $(\mathbf{I}+\mathbf{R})$ commute. For example:

- $\mathbf{R}(\mathbf{I}-\mathbf{R})^{-1} = (\mathbf{I}-\mathbf{R})^{-1}\mathbf{R}$,
- $\mathbf{R}(\mathbf{I}-\mathbf{R})^{-2} = (\mathbf{I}-\mathbf{R})^{-2}\mathbf{R} = (\mathbf{I}-\mathbf{R})^{-1}\mathbf{R}(\mathbf{I}-\mathbf{R})^{-1}$,

$$\bullet\ \mathbf{R}(\mathbf{I}-\mathbf{R})^{-1}(\mathbf{I}+\mathbf{R}) = (\mathbf{I}-\mathbf{R})^{-1}\mathbf{R}\,(\mathbf{I}+\mathbf{R}) = (\mathbf{I}+\mathbf{R})\mathbf{R}(\mathbf{I}-\mathbf{R})^{-1}.$$

Therefore,

$$\begin{aligned}
L^a &= \sum_{i=1}^{\infty} i\,\pi\,(\mathbf{I}-\mathbf{R})\,\mathbf{R}^i.\,\mathbf{e} \\
&= \pi\,(\mathbf{I}-\mathbf{R}) \sum_{i=1}^{\infty} i\,\mathbf{R}^i.\,\mathbf{e} \\
&= \pi\,(\mathbf{I}-\mathbf{R})\,(\mathbf{I}-\mathbf{R})^{-2}\mathbf{R}\,.\,\mathbf{e} \\
&= \pi\,(\mathbf{I}-\mathbf{R})^{-1}\mathbf{R}\,.\,\mathbf{e}\ . \\
L^a_q &= \sum_{i=2}^{\infty} (i-1)\,\pi\,(\mathbf{I}-\mathbf{R})\,\mathbf{R}^i.\,\mathbf{e} \\
&= \pi\,(\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{R}^2\,.\,\mathbf{e}\ .
\end{aligned}$$

$$\begin{aligned}
E[\,(N^a_n)^2] &= \sum_{i=1}^{\infty} i^2\,\pi\,(\mathbf{I}-\mathbf{R})\,\mathbf{R}^i.\,\mathbf{e} \\
&= \pi\,(\mathbf{I}-\mathbf{R}) \sum_{i=1}^{\infty} i^2\,\mathbf{R}^i.\,\mathbf{e} \\
&= \pi\,(\mathbf{I}-\mathbf{R})\,(\mathbf{I}-\mathbf{R})^{-3}\mathbf{R}\,(\mathbf{I}+\mathbf{R})\,.\,\mathbf{e} \\
&= \pi\,(\mathbf{I}-\mathbf{R})^{-2}\mathbf{R}\,(\mathbf{I}+\mathbf{R})\,.\,\mathbf{e}\ .
\end{aligned}$$

□

The following theorem gives the lag-1 covariance and lag-1 correlation of the number of customers in the system seen by two successive arrivals.

THEOREM 1.63 . At steady state,

$$E[\,N^a_n\,,\,N^a_{n+1}\,] = \sum_{i=1}^{\infty}\ \sum_{j=1}^{i+1} i\,.\,j\,.\,\mathbf{x}_i\,.\,\mathbf{A}_{i-j+1}.\,\mathbf{e}\ , \tag{1.64}$$

and hence

$$Cov\,[\,N_n^a\,,\,N_{n+1}^a\,] = E\,[\,N_n^a\,,\,N_{n+1}^a\,] - E\,[N_n^a]^2 \quad , \tag{1.65}$$

$$Corr\,[\,N_n^a\,,\,N_{n+1}^a\,] = \frac{Cov\,[\,N_n^a\,,\,N_{n+1}^a\,]}{Var\,[\,N_n^a\,]} \quad . \tag{1.66}$$

Proof.

$$P\,[\,N_n^a = i\,,\,N_{n+1}^a = j\,] = \sum_{h=1}^{m}\ \sum_{k=1}^{m} P\,[\,N_{n+1}^a = j,\ Z_{n+1} = h,\ N_n^a = i,\ Z_n = k\,]$$

$$= \sum_{h=1}^{m} \sum_{k=1}^{m} P[\,N_n^a = i, Z_n = k]\,.\,P[\,N_{n+1}^a = j, Z_{n+1} = h \mid N_n^a = i, Z_n = k]$$

$$= \sum_{h=1}^{m}\ \sum_{k=1}^{m} \mathbf{x}_{ik}\,.\,\mathbf{P}_{i,k\,;\,j,l}$$

$$= \begin{cases} \mathbf{x}_i\,.\,\mathbf{B}_i\,.\,\mathbf{e} & \text{if } j=0\ ,\ \ i=0,\,1,\,2,\,\ldots, \\ \mathbf{x}_i.\mathbf{A}_{i-j+1}.\,\mathbf{e} & \text{if } j \geq 1\ \ ,\ \ i+1 \geq j\ , \\ 0 & \text{otherwise.} \end{cases}$$

Consequently,

$$E\,[\,N_n^a\,,\,N_{n+1}^a\,] = \sum_{j=1}^{\infty}\ \sum_{i=j-1}^{\infty} i\,.\,j\,.\,\mathbf{x}_i\,.\,\mathbf{A}_{i-j+1}.\,\mathbf{e}$$

$$= \sum_{i=1}^{\infty}\ \sum_{j=1}^{i+1} i\,.\,j\,.\,\mathbf{x}_i\,.\,\mathbf{A}_{i-j+1}.\,\mathbf{e}\,.$$

□

THEOREM 1.67 . Given that the incoming customer is of type j, the expected number of customers seen by that customer is

$$E\,[\,N_n^a \mid Z_n = j\,] = \frac{1}{\pi_j}\,.\,[\pi\,(\mathbf{I}-\mathbf{R})^{-1}\mathbf{R}\,]_j \quad ,\ j = 1, 2, \ldots, m\,. \tag{1.68}$$

Proof.

$$P\,[\,N_n^a \mid Z_n = j\,] = \frac{P\,[\,N_n^a\,,\,Z_n = j\,]}{P\,[\,Z_n = j\,]} = \frac{1}{\pi_j}\,.\,[\pi\,(\mathbf{I}-\mathbf{R})\,\mathbf{R}^i\,]_j \quad . \tag{1.69}$$

Thus,

$$E[\, N_n^a \mid Z_n = j\,] = \sum_{i=1}^{\infty} i \cdot \frac{1}{\pi_j} \cdot [\pi\,(\mathbf{I}-\mathbf{R})\,\mathbf{R}^i\,]_j$$
$$= \frac{1}{\pi_j} \cdot [\pi\,(\mathbf{I}-\mathbf{R}) \sum_{i=1}^{\infty} i \cdot \mathbf{R}^i\,]_j$$
$$= \frac{1}{\pi_j} \cdot [\pi\,(\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{R}\,]_j \;\;,\;\; j = 1, 2, \ldots, m\,. \qquad (1.70)$$

□

The conditional probability (1.69) and the conditional expectation (1.68) show how the queue lengths vary with the types of incoming customers. This may sheds light on the fluctuation of the queue lengths.

Remarks.

REMARK 1. In the GI/M/1 queue (see for example Gross and Harris (1985)), the stationary queue length probabilities embedded at arrivals are geometrically distributed and are given by $x_n = (1-z)\,z^n$, $n = 0, 1, 2, \ldots$, where z is the unique solution in the unit interval (0,1) to the functional equation $z = \sum_{k=1}^{\infty} a_k z^k$, or $z = A^*(\;\mu(1-z)\;)$, where $A^*(s)$ is the Laplace-Stieltjes transform of the interarrival distribution $A(t)$ and $a_k = \int_0^{\infty} q_k \; dA(t)$.

It is apparent from Lemma 1.54 that the matrix-geometric form of the stationary queue length probability vector for the MR/M/1 queue is the matrix analog of the scalar geometric form of that for the GI/M/1 queue.

REMARK 2. In most cases, the matrix $\mathbf{R}$ cannot be found algebraically. We must resort to numerical methods. Fortunately, there are procedures that find $\mathbf{R}$ iteratively. These iterative procedures converge monotonically to $\mathbf{R}$ with good accuracy (see Neuts 1981). Section 3.5 will discuss these procedures.

REMARK 3. Let $\omega_1, \omega_2, \ldots, \omega_m$ be the eigenvalues of $\mathbf{R}$. *Then*

$$\mathbf{R} = \omega_1\mathbf{E}_1 + \omega_2\mathbf{E}_2 + \cdots + \omega_m\mathbf{E}_m \quad , \tag{1.71}$$

where $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_m$ are the rank-one matrices constructed from the outer product of the left and right eigenvectors of $\mathbf{R}$ such that

$$\mathbf{E}_i \cdot \mathbf{E}_i = \mathbf{E}_i \quad \textit{and} \quad \mathbf{E}_i \cdot \mathbf{E}_j = 0 \quad \textit{if} \ i \neq j \ . \tag{1.72}$$

Equation (1.71) is called *the spectral representation of* $\mathbf{R}$ (see Çinlar (1975b, pp.364-370)).

By substituting (1.71) into (1.55), the eigenvalues $\omega_1, \omega_2, \ldots, \omega_m$ of $\mathbf{R}$ are found after solving the determinantal equation

$$\mathit{det}\,[\ \omega\,\mathbf{I} - \mathbf{A}^*(\ \mu(1-\omega)\ )\ ] = 0 \ . \tag{1.73}$$

If we solve the infinite difference equation $\mathbf{x.P}=\mathbf{x}$, $\mathbf{x.e}=1$ in scalar form (see for example, Morse (1958)), the functional equation (1.73) turns out to be the characteristic (secular) equation of that difference equation. Rouche's theorem is then used to verify that there are exactly m roots inside the unit disk. In most cases, it is impossible to find these roots algebraically. A numerical root-finding routine must be used. The presence of multiple roots may complicate the root-finding procedure. After all the roots are found, it is still necessary to find the constants by fitting the intial (boundary) conditions. Therefore, we find that it is easier to compute $\omega_1, \omega_2, \ldots, \omega_m$ as the eigenvalues of $\mathbf{R}$ than to compute the roots of (1.73).

REMARK 4. In Remarks 2 and 3, we discuss the difficulty of finding $\mathbf{R}$ explicitly in the case of the MR/M/1 queue or of finding z in the case of the GI/m/1 queue. Numerical method must be used. The difficulty appears to be inherent in all the GI/M/s−type queues. Laplace-Stieltjes transform/generating function approach does not help in this type of problem.

Therefore, we found the method of Neuts (1981) to be better suited to our needs than the classical Laplace-Stieltjes transform/generating function approach. It is a computational procedure and it will be outlined in section 3.5.

REMARK 5. *CAUDAL CHARACTERISTIC.*

Notice that the marginal stationary distribution $P[N_n^a = i] = \mathbf{x}_i \cdot \mathbf{e}$ is not geometric. However, asymptotically, it is geometric because it can be shown using the spectral representation of $\mathbf{R}$ (1.71) that

$$\lim_{i \to \infty} \frac{P[\, N_n^a = i+1 \,]}{P[\, N_n^a = i \,]} = \frac{\mathbf{x}_{i+1} \cdot \mathbf{e}}{\mathbf{x}_i \cdot \mathbf{e}} = \eta \ , \tag{1.74}$$

where η is the largest eigenvalue of $\mathbf{R}$ (see Lemma 1.36). Thus, if η is large (close to 1), the tail distribution of N_n^a *is fat.* This means that when the queue becomes congested, it will remain so for a long period of time. Neuts (1986) called η *the caudal characteristic* because it sheds light on the behavior of the tail (in Latin, *cauda*) distribution of the queue length. A graph of η as a function of ρ , $\eta(\rho)$ for $0<\rho<1$, is called *the caudal characteristic curve.*

The traffic intensity ρ is only a gross indication of congestion. It is the ratio of the work requested to the service capacity for the whole history of the queue. It does not capture any transitory behavior. Thus, the caudal characteristic η is a better indicator of congestion than ρ, because it relates to the duration of excursions to the congested states. In-depth discussion of caudal characteristic is found in Neuts (1985) and Latouche (1985).

In the GI/M/1 queue (see Remark 1), the stationary queue length probabilities are geometrically distributed: $P[\ N_n^a = i\ ] = (1-z)\ z^i$. It is well known that $z=\rho$ if and only if GI=M, i.e. the M/M/1 queues. In the degenerate MR/M/1 queue where MR=GI (see section 2.2), all the eigenvalues of $\mathbf{R}$ are zero except one which is η, and $\eta=z$. This can be shown using (1.71).

For the class of Erlang arrivals (E_k/M/1), it can be shown that $z \leq \rho$ with equality holding only when k=1 (i.e. E_1=M). This can be intuitively explained as follows: since the coefficient of variation of Erlang distributions is less than 1 , the Erlang arrival process is more *regular* (less variability) than the Poisson (M) arrival process ($c.v.$=1). Hence, it is expected that the queue length of the E_k/M/1 queue will be shorter *(better behaved)* than that of the M/M/1 queue. As k increases, z decreases to a lower limit, z_∞, which corresponds to the z of the E_∞/M/1 or D/M/1 queue. On the other hand, the class of H/M/1 queues (hyperexponential arrivals with $c.v. \geq 1$) yields $z > \rho$.

The caudal characteristic curve of the M/M/1 queue is the straight line $\eta(\rho)=\rho$. All of the caudal characteristic curves of the E_k/M/1 queues lie below that of M/M/1 queue (i.e. $\eta(\rho) \leq \rho$), with the lowest curve being that of the D/M/1 queue;

while the caudal characteristic curves of the H/M/1 queues lie above that of M/M/1 queue.

Using the M/M/1 queue ($\eta(\rho)=\rho$) as the principle reference, whenever $\eta(\rho)>\rho$ for some value of ρ, $0<\rho<1$, we conclude that when the queue becomes congested, it will stay congested longer than if the queue were an M/M/1 queue. In this case, we say that the M/M/1 queue is *better behaved* than that queue. A symmetric conclusion is drawn when $\eta(\rho)<\rho$. Thus, the E_k/M/1 queues are better behaved than the M/M/1 queue which in turn is better behaved than the H/M/1 queues.

Note that the mixture of hyperexponential queues and Erlang queues can give caudal characteristic curves such that for some $\eta(\rho)$, the system is better than the M/M/1 queue while for other values of ρ, the system is worse than the M/M/1 queue. That is, the caudal characteristic curves crosses the M/M/1 caudal characteristic curve. Such crossing occurs in cases that we later examine (see section 4.7).

3.2 The Waiting Time and the Sojourn Time.

Let W_n be *the waiting time* of the nth customer (i.e. the time it spent waiting in the queue before being served) and let S_n be the *sojourn time* of the nth customer (waiting plus service). For $j = 1, 2, \ldots m$, let

$$W_j(t) = P[\, W_n \leq t \,,\, Z_n = j\,] \;, \tag{2.1}$$

$$S_j(t) = P[\, S_n \leq t \,,\, Z_n = j\,] \;. \tag{2.2}$$

Thus, $W_j(t)$ is the probability that a customer of type j arrives and waits in the queue for at least t time units before being served. Similarly, $S_j(t)$ is the probability that a customer of type j arrives and spends at least t time units in the system before departing.

Also, define the row vectors $\mathbf{w}(t)$ and $\mathbf{s}(t)$ as

$$\mathbf{w}(t) = [\ W_1(t)\ ,\ W_2(t)\ ,\ \ldots\ ,\ W_m(t)\ ]\ , \tag{2.3}$$

$$\mathbf{s}(t) = [\ S_1(t)\ ,\ S_2(t)\ ,\ \ldots\ ,\ S_m(t)\ ]\ , \tag{2.4}$$

respectively.

THEOREM 2.5 .

$$\mathbf{w}(t) = \pi\ [\ \mathbf{I} - \mathbf{R}\ \ e^{-\mu(\ \mathbf{I}-\mathbf{R}\ )\ t}\ ]\ , \tag{2.6}$$

and

$$\mathbf{s}(t) = \pi\ [\ \mathbf{I} - e^{-\mu(\ \mathbf{I}-\mathbf{R}\ )\ t}\ ]\ . \tag{2.7}$$

Proof. If the arriving customer finds the system empty (this occurs with probability $\mathbf{x}_0$), it immediately goes into service. If it finds n customers, $n \geq 1$, already in the system, its waiting time is Erlang distributed with parameters μ and n (the convolution of n exponential random variables each with parameter μ). Thus,

$$\begin{aligned}
\mathbf{w}(t) &= \pi\ (\mathbf{I}-\mathbf{R}) + \sum_{n=1}^{\infty} \pi\ (\mathbf{I}-\mathbf{R})\ \mathbf{R}^{n} \int_0^t \frac{\mu\ e^{-\mu x}\ (\mu x)^{n-1}}{(n-1)\ !}\ dx \\
&= \pi\ (\ \mathbf{I}-\mathbf{R}) + \pi\ (\ \mathbf{I}-\mathbf{R})\ \ \mathbf{R} \int_0^t \mu\ e^{-\mu \mathbf{I} x} \sum_{n=1}^{\infty} \frac{(\mu x \mathbf{R})^{n-1}}{(n-1)\ !}\ dx
\end{aligned}$$

$$= \pi\,(\mathbf{I-R}) + \pi\,(\mathbf{I-R})\;\mathbf{R}\int_0^t \mu\, e^{-\mu(\mathbf{I-R})x}\,dx$$

$$= \pi\,(\mathbf{I-R}) + \pi\,(\mathbf{I-R})\;\mathbf{R}\,(\mathbf{I-R})^{-1}\int_0^{\mathbf{t}} \mu(\mathbf{I-R})\, e^{-\mu(\mathbf{I-R})x}\,dx$$

$$= \pi\,(\mathbf{I-R}) + \pi\,\mathbf{R}\,\{\,\mathbf{I} - e^{-\mu(\mathbf{I-R})t}\,\}$$

$$= \pi\,\{\,\mathbf{I-R}\,e^{-\mu(\mathbf{I-R})t}\,\}\,.$$

Similarly, if the arriving customer finds n customers already in the system ($n \geq 0$), its sojourn time equals the sum of $(n+1)$ exponential service times :

$$\mathbf{s}(t) = \sum_{n=0}^{\infty} \pi\,(\mathbf{I-R})\,\mathbf{R}^n \int_0^t \frac{\mu\, e^{-\mu x}\,(\mu x)^n}{n\,!}\,dx$$

$$= \pi\,(\mathbf{I-R}) \int_0^t \mu\, e^{-\mu \mathbf{I} x} \sum_{n=0}^{\infty} \frac{(\mu x \mathbf{R})^n}{n\,!}\,dx$$

$$= \pi\,(\mathbf{I-R}) \int_0^t \mu\, e^{-\mu(\mathbf{I-R})x}\,dx$$

$$= \pi\,(\mathbf{I-R})\;(\mathbf{I-R})^{-1}\int_0^t \mu(\mathbf{I-R})\, e^{-\mu(\mathbf{I-R})x}\,dx$$

$$= \pi\,\{\mathbf{I} - e^{-\mu(\mathbf{I-R})t}\,\}\,.$$

□

Consequently, the marginal waiting time and the marginal sojourn time distributions are

$$W(t) = P\,[\,W_n(t) \leq t\,] = \mathbf{w}(t)\,.\,\mathbf{e} = 1 - \pi\,\mathbf{R}\,e^{-\mu(\mathbf{I-R})t}\,.\,\mathbf{e}\;, \qquad (2.8)$$

and

$$S(t) = P\,[\,S_n(t) \leq t\,] = \mathbf{s}(t)\,.\,\mathbf{e} = 1 - \pi\,e^{-\mu(\mathbf{I-R})t}\,.\,\mathbf{e}\;. \qquad (2.9)$$

THEOREM 2.10 . The expected waiting time and the expected sojourn time are given by

$$E[\ W_n\ ] = \tfrac{1}{\mu}\ \pi\ (\mathbf{I}-\mathbf{R})^{-1}\ \mathbf{R}\ .\ \mathbf{e} = \tfrac{1}{\mu}\ .\ L^a\ , \tag{2.11}$$

and

$$E[\ S_n\ ] = \tfrac{1}{\mu}\ \pi\ (\mathbf{I}-\mathbf{R})^{-1}.\ \mathbf{e}\ , \tag{2.12}$$

or

$$E[\ S_n\ ] = E[\ W_n\ ] + \tfrac{1}{\mu}\ . \tag{2.13}$$

Proof.

$$\begin{aligned} E[\ W_n\ ] &= \int_0^\infty P[\ W_n > t\ ]\ dt \\ &= \int_0^\infty \pi\ \mathbf{R}\ e^{-\mu(\mathbf{I}-\mathbf{R})t} dt\ .\ \mathbf{e} \\ &= \tfrac{1}{\mu}\ \pi\ (\mathbf{I}-\mathbf{R})^{-1}\ \mathbf{R} \int_0^\infty \mu(\mathbf{I}-\mathbf{R})\ e^{-\mu(\mathbf{I}-\mathbf{R})t} dt\ .\ \mathbf{e} \\ &= \tfrac{1}{\mu}\ \pi\ (\mathbf{I}-\mathbf{R})^{-1}\ \mathbf{R}\ .\ \mathbf{e}\ , \end{aligned}$$

and

$$\begin{aligned} E[\ S_n\ ] &= \int_0^\infty P[\ S_n > t\ ]\ dt \\ &= \int_0^\infty \pi\ e^{-\mu(\mathbf{I}-\mathbf{R})t} dt\ .\ \mathbf{e} \\ &= \tfrac{1}{\mu}\ \pi\ (\mathbf{I}-\mathbf{R})^{-1} \int_0^\infty \mu(\mathbf{I}-\mathbf{R})\ e^{-\mu(\mathbf{I}-\mathbf{R})t} dt\ .\ \mathbf{e} \\ &= \tfrac{1}{\mu}\ \pi\ (\mathbf{I}-\mathbf{R})^{-1}\ .\ \mathbf{e}\ . \end{aligned}$$

□

Identity (2.13) comes from the fact that $S_n = W_n + V_n$, where V_n is the exponential service time with rate μ. It is easy to verify the equivalence of (2.12) and (2.13) algebraically.

The waiting time and the sojourn time results given in (2.6)-(2.13) are the matrix analog of the scalar results for the GI/M/1 queue (see Gross and Harris p.309).

Given that a customer of type j arrives at the system, its waiting time and its sojourn time probabilities and expected values are given by

$$P[W_n \le t \mid Z_n = j] = \frac{P[W_n \le t, Z_n = j]}{P[Z_n = j]} = 1 - \frac{1}{\pi_j} \cdot [\pi \mathbf{R} e^{-\mu(\mathbf{I}-\mathbf{R})t}]_j , \tag{2.14}$$

$$P[S_n \le t \mid Z_n = j] = \frac{P[S_n \le t, Z_n = j]}{P[Z_n = j]} = 1 - \frac{1}{\pi_j} \cdot [\pi e^{-\mu(\mathbf{I}-\mathbf{R})t}]_j , \tag{2.15}$$

$$E[W_n \mid Z_n = j] = \frac{1}{\mu} \cdot \frac{1}{\pi_j} \cdot [\pi (\mathbf{I}-\mathbf{R})^{-1} \mathbf{R}]_j = \frac{1}{\mu} \cdot E[N_n^a \mid Z_n = j] , \tag{2.16}$$

$$E[S_n \mid Z_n = j] = \frac{1}{\mu} \cdot \frac{1}{\pi_j} \cdot [\pi (\mathbf{I}-\mathbf{R})^{-1}]_j . \tag{2.17}$$

3.3 The Queue Length Process at Arbitrary Times.

We use the semi-regenerative process found in Çinlar (1975b, chapter 10 sections 6-7) to obtain the queue length probability distributions at arbitrary times. These results are also found in Neuts (1978a).

Let $N(t)$ be the number of customers in the system at time t and $Z(t)$ be the type of the last customer to arrive before time t. By considering the arrival times $\{T_n^a$, $n=0, 1, 2, \ldots\}$ as the stopping times, $(\mathcal{N}(t), \mathcal{Z}(t))$ is a (two-dimensional) semi-regenerative process with $(\mathcal{N}^a, \mathcal{Z}; \mathcal{T}^a)$ as the embedded Markov renewal process. Assume also that the semi-Markov kernel $\mathbf{P}(t)$ of $(\mathcal{N}^a, \mathcal{Z}; \mathcal{T}^a)$ given in (1.16) to be non-lattice.

Let

$$\mathbf{x}_n = [\ \mathbf{x}_{n1}, \mathbf{x}_{n2}, \ldots, \mathbf{x}_{nm}\ ]\ , \quad n = 0, 1, 2, \ldots \tag{3.1}$$

be the stationary probability vector that there are n customers in the system at arrival times. Denote the expected sojourn time in state (i,j) of the Markov renewal process $(\mathcal{N}^a, \mathcal{Z}; \mathcal{T}^a)$ by m_{ij} and let α_i the column vector

$$\alpha_i = [\ m_{i1}\ , m_{i2}\ , \ldots , m_{im}\ ]^{\mathrm{T}}\ . \tag{3.2}$$

Then

$$\begin{aligned} \alpha_i &= \sum_{j=0}^{\infty} \int_0^{\infty} t \ \ d\mathbf{P}_{ij}(t)\ .\ \mathbf{e} \\ &= \int_0^{\infty} t \ \ \sum_{j=0}^{\infty} \mathbf{A}_j(t)\ .\ \mathbf{e} \\ &= \int_0^{\infty} t\ d\mathbf{A}(t)\ .\ \mathbf{e} = \mathbf{A}^{(1)}\ .\ \mathbf{e}\ . \end{aligned} \tag{3.3}$$

Here, $\mathbf{P}_{ij}(t)$ is the (i,j) block of $\mathbf{P}(t)$ given in (1.16), and (3.3) follows from the fact that

$$\sum_{j=0}^{\infty} \mathbf{A}_j(t) \ = \ \mathbf{A}(t)\ ,$$

where $\mathbf{A}_j(t)$, $\mathbf{A}(t)$ and $\mathbf{A}^{(1)}$are defined in (1.17), (2.1.9) and (2.1.15), respectively. Notice that (3.3) implies that α_i is independent of i and hence we can drop the subscript i from α_i :

$$\alpha = \alpha_i = \mathbf{A}^{(1)}\ .\ \mathbf{e}\ . \tag{3.4}$$

Define the *fundamental mean*

$$\varepsilon = \sum_{i=0}^{\infty} \mathbf{x}_i \cdot \alpha_i \ . \tag{3.5}$$

Then

$$\varepsilon = \sum_{i=0}^{\infty} \mathbf{x}_i \cdot \alpha = \pi \cdot \mathbf{A}^{(1)} \cdot \mathbf{e} \ ,$$

or

$$\varepsilon = \frac{1}{\mu \, \rho} \ . \tag{3.6}$$

This follows from $\sum_{i=0}^{\infty} \mathbf{x}_i = \pi$ and (1.53) , where μ is the service rate and ρ is the traffic intensity.

Let

$$y_{nl}(t) = P[\, N(t) = n \, , \, Z(t) = l \,] \tag{3.7}$$

and

$$y_{nl} = \lim_{t \to \infty} y_{nl}(t) \ . \tag{3.8}$$

Here, $y_{nl}(t)$ is the probability that n customers are present in the system at time t and the last customer to arrive before t was of type l.

The limiting probability y_{nl} is given in the next theorem using *the key renewal theorem for Markov renewal process* (see Çinlar (1975b, pp.346-347).

THEOREM 3.9 . Let $\mathbf{y}_n$ *be the row vector*

$$\mathbf{y}_n = [\, y_{n1} \, , \, y_{n2} \, , \ldots , \, y_{nl} \,] \ , \quad n = 0, 1, 2, \ldots \ . \tag{3.10}$$

Then

$$\mathbf{y}_n = \mu \, \rho \, \pi \, (\mathbf{I} - \mathbf{R}) \, \mathbf{R}^{n-1} \, \Phi(\mathbf{R}) \ , \tag{3.11}$$

and

$$\mathbf{y}_0 = \mu\ \rho\ \pi\ [\ \Delta\alpha - \Phi(\mathbf{R})\ ]\ , \tag{3.12}$$

where

$$\Phi(\mathbf{R}) = \sum_{j=0}^{\infty} \mathbf{R}^j \int_0^{\infty} \frac{e^{-\mu t}(\mu t)^j}{j\,!}\ [\ \mathbf{I} - \Delta\mathbf{A}(t)\ ]\ dt\ , \tag{3.13}$$

$\Delta\mathbf{A}(t)$ $\doteq$ *the m by m diagonal matrix with diagonal elements given by the probability distributions* $[\ \mathbf{A}(t)\ .\ \mathbf{e}\ ]_l$, l=1, 2, ..., m , (3.14)

$\Delta\alpha$ $\doteq$ *the m by m diagonal matrix with diagonal elements* α_1, α_2, ..., α_m, *(the elements of the column vector* α *in* (3.4)) . (3.15)

Proof. For $l = 1\ , 2\ , \ldots\ , m$, let

$$K_t(j,n)_l = P\ [\ N(t) = n\ ,\ T_1^a > t\ |\ N(0) = j\ ,\ Z(0) = l\ ]\ . \tag{3.16}$$

Then

$$K_t(j,n)_l = P\ [\ T_1^a > t\ |\ N(0) = j\ ,\ N(t) = n\ ,\ Z(0) = l\ ]\ .\ P\ [\ N(t) = n\ |\ N(0) = j\ ,\ Z(0) = l\ ]\ . \tag{3.17}$$

or

$$K_t(j,n)_l = \begin{cases} [\mathbf{e}.\mathbf{A}(t).\mathbf{e}\ ]_l . \dfrac{e^{-\mu t}(\mu t)^{j+n-1}}{(j+n-1)\ !} & \textit{for } n=1,\ 2,\ \ldots \textit{ and } j=n-1,\ n,\ n+1, \\ [\ \mathbf{e}.\mathbf{A}(t).\mathbf{e}\ ]_l \cdot \displaystyle\sum_{i=j+1}^{\infty} \frac{e^{-\mu t}(\mu t)^i}{i\ !} & \textit{for } n=0 \textit{ and } j=0,\ 1,\ 2,\ \ldots \\ 0 & \textit{otherwise.} \end{cases} \tag{3.18}$$

Using the key renewal theorem for Markov renewal process (Çinlar 1975b, Theorem 10.6.2, p.347) we get for n=0, 1, 2, ... and l=1, 2, ..., m

$$y_{nl} = \frac{1}{\varepsilon} \sum_{j=0}^{\infty} (\mathbf{x}_j)_l \int_0^{\infty} K_t(j,n)_l \ dt \ . \qquad (3.19)$$

Hence, for n=1, 2, ... and l=1, 2, ..., m ,

$$y_{nl} = \mu \rho \sum_{j=n-1}^{\infty} (\mathbf{x}_j)_l \int_0^{\infty} \frac{e^{-\mu t} (\mu t)^{j+n-1}}{(j+n-1)\,!} [\, \mathbf{e} - \mathbf{A}(t) \, . \, \mathbf{e} \,]_l \ dt \ , \qquad (3.20)$$

and for n=0 , l=1, 2, ..., m ,

$$y_{0l} = \mu \rho \sum_{j=0}^{\infty} (\mathbf{x}_j)_l \int_0^{\infty} \sum_{i=j+1}^{\infty} \frac{e^{-\mu t} (\mu t)^i}{i\,!} [\, \mathbf{e} - \mathbf{A}(t) \, . \, \mathbf{e} \,]_l \ dt \ , \qquad (3.21)$$

where

$$\mathbf{x}_j = \pi \, (\mathbf{I} - \mathbf{R}) \, \mathbf{R}^j \ , \ j = 0, 1, 2, \dots \ , \qquad (3.22)$$

with

$$\mathbf{R}^0 = \mathbf{I} \ . \qquad (3.23)$$

Since,

$$\alpha = \int_0^{\infty} t \, d\mathbf{A}(t) \, . \, \mathbf{e} \ = \ \int_0^{\infty} [\, \mathbf{I} - \mathbf{A}(t) \,] \, . \, \mathbf{e} \ \ dt \ , \qquad (3.24)$$

the m by m matrix $\Delta\alpha$ can be written as

$$\Delta\alpha = \int_0^{\infty} [\, \mathbf{I} - \Delta\mathbf{A}(t) \,] \, . \, \mathbf{e} \ \ dt \ , \qquad (3.25)$$

where $\Delta\mathbf{A}(t)$ is defined in (3.14).

The vector-matrix forms of (3.20) and (3.21) are

$$\mathbf{y}_n = \mu\,\rho \sum_{j=n-1}^{\infty} \mathbf{x}_j \int_0^{\infty} \frac{e^{-\mu t}(\mu t)^{j+n-1}}{(j+n-1)\,!} \,[\,\mathbf{I}-\Delta\mathbf{A}(t)\,]\; dt \;\;,\;\; n=1,2,\ldots, \quad (3.26)$$

$$\mathbf{y}_0 = \mu\,\rho \sum_{j=0}^{\infty} \mathbf{x}_j \int_0^{\infty} \sum_{i=j+1}^{\infty} \frac{e^{-\mu t}(\mu t)^{i}}{i\,!} \,[\,\mathbf{I}-\Delta\mathbf{A}(t)\,]\; dt \;. \quad (3.27)$$

Further simplification of(3.26) yields

$$\mathbf{y}_n = \mu\,\rho\,\pi\,(\mathbf{I}-\mathbf{R}) \sum_{j=0}^{\infty} \mathbf{R}^{j+n-1} \int_0^{\infty} \frac{e^{-\mu t}(\mu t)^{j+n-1}}{(j+n-1)\,!} \,[\,\mathbf{I}-\Delta\mathbf{A}(t)\,]\; dt \;, \quad (3.28)$$

or

$$\mathbf{y}_n = \mu\,\rho\,\pi\,(\mathbf{I}-\mathbf{R})\,\mathbf{R}^{n-1}\,\Phi(\mathbf{R}) \;\;,\;\; n = 1\,,2\,,\ldots. \quad (3.29)$$

Similarly,

$$\mathbf{y}_0 = \mu\,\rho\,\pi\,(\mathbf{I}-\mathbf{R}) \int_0^{\infty} \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} \mathbf{R}^{j}\, \frac{e^{-\mu t}(\mu t)^{i}}{i\,!} \,[\,\mathbf{I}-\Delta\mathbf{A}(t)\,]\; dt \;. \quad (3.30)$$

Since

$$\sum_{j=1}^{i-1} \mathbf{R}^{j} = (\mathbf{I}-\mathbf{R})^{-1}\,(\mathbf{I}-\mathbf{R}^{i}) \quad (3.31)$$

and

$$\mathbf{I} - \mathbf{R}^0 = \mathbf{0} \;\;, \quad (3.32)$$

$$\mathbf{y}_0 = \mu\,\rho\,\pi\,(\mathbf{I}-\mathbf{R})(\mathbf{I}-\mathbf{R})^{-1} \int_0^{\infty} e^{-\mu t} \sum_{i=1}^{\infty} (\mathbf{I}-\mathbf{R}^{i})\, \frac{(\mu t)^{i}}{i\,!} \,[\,\mathbf{I}-\Delta\mathbf{A}(t)\,]\; dt$$

$$= \mu\,\rho\,\pi \left\{ \int_0^{\infty} e^{-\mu t}.\,\mathbf{I} \sum_{i=0}^{\infty} \frac{(\mu t)^{i}}{i\,!} \,[\,\mathbf{I} - \Delta\mathbf{A}(t)\,]\; dt \right.$$

$$\left. - \int_0^{\infty} e^{-\mu t} \sum_{i=0}^{\infty} \mathbf{R}^{i}\, \frac{(\mu t)^{i}}{i\,!} \,[\,\mathbf{I}-\Delta\mathbf{A}(t)\,]\; dt \right\}$$

$$= \mu\,\rho\,\pi \int_0^{\infty} [\,\mathbf{I}-\Delta\mathbf{A}(t)\,]\; dt - \sum_{i=0}^{\infty} \mathbf{R}^{i} \int_0^{\infty} e^{-\mu t}\, \frac{(\mu t)^{i}}{i\,!} \,[\,\mathbf{I}-\Delta\mathbf{A}(t)\,]\, dt \quad (3.33)$$

or

$$\mathbf{y}_0 = \mu\, \rho\, \boldsymbol{\pi}\, [\, \Delta\alpha - \Phi(\mathbf{R})\,] \,. \tag{3.34}$$

This completes the proof.

□

LEMMA 3.35 .

$$\Phi(\mathbf{R}) \,.\, \mathbf{e} = \tfrac{1}{\mu} \,.\, \mathbf{e} \;. \tag{3.36}$$

Proof.

$$\Phi(\mathbf{R}) = \sum_{i=0}^{\infty} \mathbf{R}^i \int_0^\infty e^{-\mu t} \frac{(\mu t)^i}{i\,!} [\, \mathbf{I} - \Delta\mathbf{A}(t)\,]\, dt$$

$$= \int_0^\infty e^{-\mu(\mathbf{I}-\mathbf{R})\,t} [\, \mathbf{I} - \Delta\mathbf{A}(t)\,]\, dt \;.$$

Integrating by parts we get

$$\Phi(\mathbf{R}) = \tfrac{1}{\mu} (\mathbf{I}-\mathbf{R})^{-1} \left\{ \mathbf{I} - \int_0^\infty e^{-\mu(\mathbf{I}-\mathbf{R})\,t} d(\Delta\mathbf{A}(t)) \right\} .$$

Since $\Delta\mathbf{A}(t) \,.\, \mathbf{e} = \mathbf{A}(t) \,.\, \mathbf{e}$ and $\mathbf{R} = \int_0^\infty e^{-\mu(\mathbf{I}-\mathbf{R})t} d\mathbf{A}(t)$ (see (1.55)), then

$$\Phi(\mathbf{R}) = \tfrac{1}{\mu} (\mathbf{I}-\mathbf{R})^{-1} (\mathbf{I}-\mathbf{R}) \,.\, \mathbf{e} = \tfrac{1}{\mu} \,.\, \mathbf{e} \;.$$

□

The following theorem gives the marginal distribution of the number of customers in the system at arbitrary times.

THEOREM 3.37 .

$$\mathbf{y}_0 \,.\, \mathbf{e} = 1 - \rho \tag{3.38}$$

and

$$\mathbf{y}_n \cdot \mathbf{e} \;=\; \rho\ \pi\, (\mathbf{I}-\mathbf{R})\, \mathbf{R}^{n-1} \cdot \mathbf{R} \quad , \quad n = 1, 2, 3, \ldots \quad . \tag{3.39}$$

Proof. The proof is obvious from Lemma 3.35 and the fact that $\Delta\alpha \cdot \mathbf{e} = \alpha$.

□

Expected Queue Length - Time property.

Let $L^t = E\,[\;N(t)\;]$ be the expected number of customers in the system at arbitrary times and let L_q^t be the corresponding number of customers waiting in the queue. Then

THEOREM 3.40 .

$$L^t \;=\; \rho\ \pi\, (\mathbf{I}-\mathbf{R})^{-1} \cdot \mathbf{e} \tag{3.41}$$

and

$$L_q^t \;=\; \rho\ \pi\, (\mathbf{I}-\mathbf{R})^{-1}\, \mathbf{R} \cdot \mathbf{e} \;=\; \rho\ L^a \quad . \tag{3.42}$$

Proof. The proof is immediate from

$$L^t \;=\; \sum_{n=1}^{\infty} n \cdot \mathbf{y}_n \cdot \mathbf{e} \qquad \text{and} \quad L_q^a = \sum_{n=2}^{\infty} (n-2) \cdot \mathbf{y}_n \cdot \mathbf{e} \quad .$$

□

The Variance of $N(t)$ is found to be

$$Var\,[\;N(t)\;] \;=\; E\,[\;N(t)^2\;] \;-\; (L^t)^{\,2} \quad , \tag{3.43}$$

where

$$E\,[\;N(t)^2\;] \;=\; \rho\ \pi\, (\mathbf{I}-\mathbf{R})^{-2}\, (\mathbf{I}+\mathbf{R}) \cdot \mathbf{e} \quad . \tag{3.44}$$

3.4 Little's Formulas and Related Results.

In this section, we pull together the mean-value results from sections 3.1 , 3.2 , and 3.3. L and L_q denote the expected number of customers in the system and in the queue, respectively. The superscript a means *at arrival time* and the superscript t means *at arbitrary times*. W and W_q are the expected sojourn time and the expected waiting time, respectively.

- *At arrival time (customer's averages) :*

$$L_q^a = \pi\,(\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{R}^2\,.\,\mathbf{e}\ , \qquad (4.1)$$

$$L^a = \pi\,(\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{R}\,.\,\mathbf{e}\ , \qquad (4.2)$$

$$W_q = \frac{1}{\mu}\,.\,\pi\,(\mathbf{I}-\mathbf{R})^{-1}\mathbf{R}\,.\,\mathbf{e} = \frac{1}{\mu}\,.\,L^a\ , \qquad (4.3)$$

$$W = \frac{1}{\mu}\,.\,\pi\,(\mathbf{I}-\mathbf{R})^{-1}.\,\mathbf{e}\ . \qquad (4.4)$$

- *At arbitrary times (time averages) :*

$$L_q^t = \rho\,\pi\,(\mathbf{I}-\mathbf{R})^{-1}\,\mathbf{R}\,.\,\mathbf{e} = \rho\,L^a\ , \qquad (4.5)$$

$$L^t = \rho\,\pi\,(\mathbf{I}-\mathbf{R})^{-1}.\,\mathbf{e}\ . \qquad (4.6)$$

Below we give 9 formulas that are derived directly from the 6 formulas above. The first 5 of these 9 formulas are known to hold for the more general G/G/1 queue (see Franken, König, Arndt, Schmidt (1982). Formulas (4.8) and (4.9) are known as Little's formulas and they show the relationships between the customer averages and

the time averages. In all of these formulas $\lambda = \frac{1}{E[\ D_n^a\ ]}$ is the arrival rate , μ is the service rate and ρ is the traffic intensity. Here are the formulas :

$$W = W_q + \frac{1}{\mu} \tag{4.7}$$

$$L_q^a = \lambda \cdot W_q \tag{4.8}$$

$$L^t = \lambda \cdot W \tag{4.9}$$

$$L^t = L_q^a + \rho \tag{4.10}$$

$$\lim_{t \to \infty} P[\ N(t) > 0\ ] = \rho \tag{4.11}$$

$$W_q = \frac{1}{\mu} \cdot L^a \tag{4.13}$$

$$W = \frac{1}{\mu} (\ L^a + 1\ ) \tag{4.14}$$

$$L_q^t = \rho\ L^a \tag{4.15}$$

$$L^t = \rho (\ L^a + 1\ ) \tag{4.16}$$

3.5 Computational Procedure for the MR/M/1 Queue.

This section outlines the procedure recommended by Neuts (1981) to compute the matrix $\mathbf{R}$ (see (1.34) and (1.35)). First, we make sure that the limiting probability vector $\mathbf{x}_n$, $n = 0, 1, 2, \dots$, exists. Theorem 1.47 and Lemma 1.51 give such condition: the traffic intensity should be less than 1 ,

$$\rho = \frac{1}{\mu \,.\, E[\, D_n^a \,]} = \frac{1}{\mu.\pi.\mathbf{A}^{(1)}.\,\mathbf{e}} < 1 \ , \tag{5.1}$$

where $E[\, D_n^a \,]$ is the mean interarrival time.

The following theorem summarizes the matrix-geometric solution to the MR/M/1 queue.

THEOREM 5.2 . The initial probability vector $\mathbf{x}_0$ is

$$\mathbf{x}_0 = \pi\,(\mathbf{I}-\mathbf{R}) \ , \tag{5.3}$$

where $\mathbf{R}$ is the minimal non-negative solution to

$$\mathbf{R} = \sum_{n=0}^{\infty} \mathbf{R}^n.\mathbf{A}_n \ . \tag{5.4}$$

Therefore, the partitioned limiting queue length probability vectors are given by

$$\mathbf{x}_n = \pi\,(\mathbf{I}-\mathbf{R})\,\mathbf{R}^n \ , \ n = 0, 1, 2, \dots \ . \tag{5.5}$$

□

The matrix $\mathbf{R}$ is approximated using the iterative function

$$\mathbf{R}(i+1) = \sum_{n=0}^{\infty} \mathbf{R}^n(i)\ \mathbf{A}_n \ , \quad n = 0, 1, 2, \dots \ . \tag{5.6}$$

$\mathbf{R}(i)$ denotes the ith iterate of $\mathbf{R}$ and we choose the starting matrix $\mathbf{R}(\mathbf{0}) = \mathbf{0}$, the zero matrix. Neuts (1981, p.9) showed that the iterative function (5.6) converges monotonically to $\mathbf{R}$, the minimal non-negative solution to (5.4), because $\mathbf{R}(i+1) \geq \mathbf{R}(i)$ can be established through induction using the fact that the matrices $\mathbf{A}_n$'s are non-negative substochastic matrices. The iteration may be stopped as soon a $|\mathbf{R}(i+1) - \mathbf{R}(i)| < \epsilon$, for some norm and for some tolerance ϵ.

Notice that (5.6) can also be written as

$$\mathbf{R}(i+1) = \{ \mathbf{A}_0 + \sum_{n=2}^{\infty} \mathbf{R}^n(i)\, \mathbf{A}_n \} (\mathbf{I}-\mathbf{A}_1)^{-1} \;, \; i = 0, 1, 2, \ldots \; . \tag{5.7}$$

If (5.7) is used instead of (5.6), one usually saves a few iterations in computing $\mathbf{R}$ to within a given tolerance. This is done only when the dimension of the problem makes the evaluation and the storage of $(\mathbf{I}-\mathbf{A}_1)^{-1}$ convenient.

The problem of truncating the infinite summation in (5.6) or (5.7) needs to be considered also. If for some index K, all the elements of $\mathbf{A}_n$ are negligibly small for all $n > K$, then we truncate the summation to K. In the proof of Lemma 1.51, we show that the column vector β defined in (1.46) to be

$$\beta = \mu \,.\, \mathbf{A}^{(1)}.\, \mathbf{e} \;, \tag{5.8}$$

which can be evaluated explicitly from (5.8). Now let

$$\beta_k = \sum_{n=0}^{K} \mathbf{B}_n \,.\, \mathbf{e} \;. \tag{5.9}$$

Then, the *truncation index* K is found by choosing the smallest index K such that

$$\max_j \; \{ |\beta - \beta_K| < \epsilon \}_j \;, \tag{5.10}$$

where $\epsilon = 10^{-8}$ was found to be adequate (see Neuts (1981, p.37)).

Finally, we can also use the fact that $\mathbf{B}[\mathbf{R}]$ is stochastic as an *accuracy check* on $\mathbf{R}$. Below we summarize the numerical procedure to compute $\mathbf{R}$ and $\mathbf{x}_n$ of the MR/M/1 queue.

1. *Compute the traffic intensity* ρ *from* (5.1). *If* $\rho \geq 1$, *stop. This indicates that the Markov chain* $(\mathcal{N}^a, \mathcal{Z})$ *is not positive recurrent. If* $\rho < 1$, *proceed.*
2. *Find the truncation index* K.
3. *Compute* $\mathbf{R}$ *iteratively using* (5.6) *or* (5.7) *until convergence.*
4. *Verify that* $\mathbf{B}[\mathbf{R}]$ *is stochastic (i.e.* $\mathbf{B}[\mathbf{R}] \cdot \mathbf{e} = \mathbf{e}$ *) as an accuracy check.*
5. *Compute* $\mathbf{x}_0 = \boldsymbol{\pi} (\mathbf{I} - \mathbf{R})$.
6. *Compute* $\mathbf{x}_n = \mathbf{x}_{n-1} \cdot \mathbf{R}$, $n = 1, 2, \ldots$.

This procedure is implemented as SUBROUTINE MRM1 and SUBROUTINE NEUTS in APPENDIX A.

Chapter 4

THE EFFECTS OF THE MOMENTS AND THE SERIAL CORRELATIONS OF THE MRAP ON THE QUEUEING PERFORMANCE MEASURES

4.0 Introduction.

The main objective of this paper is to investigate how the dependency in the arrival process effects the queueing performance measures. To accomplish this, the 2-state Markov renewal arrival process (MRAP) described in section 2.3 is used. Subsequently, the queueing properties of the MR/M/1 queue are derived in chapter 3. Our interest is not restricted to the MR/M/1 queue, rather we are interested in queueing systems with dependent arrivals. We believe that the observations made on the MR/M/1 queue are indicative of what happens more generally.

This chapter is organized as follows. In section 4.1, we decide on which performance measures and which parameters of the arrival process to use in this investigation. The numerical methods used in the investigation are discussed in section 4.2, and in section 4.3, we comment on the plotting routine used to produced the graphs in this chapter.

We started the investigation by looking at the uncorrelated GI/M/1 queue and find out how the first 3 moments of the arrival process affect the mean queue length and the standard deviation of the queue length. This is done in section 4.4. In section 4.5, we consider a specific example of the MR/M/1 queue and call it *the correlated*

M/M/1 queue. There, we compare the mean queue length of the correlated M/M/1 queue to that of the ordinary uncorrelated M/M/1 queue.

Then, in section 4.6, we make a thorough investigation on how the moments of the arrival process and the dependency in the arrival process affect the mean queue length and the standard deviation of the queue length. We also look at the caudal characteristic of the MR/M/1 queue and show how it is affected by the moments and the dependency of the arrival process. This is done in section 4.7.

We close this chapter with a discussion on the *wiggles* (anomalies) that appear in some of the curves and we what did trying to explain them.

4.1 Selecting the Parameters.

There are two things that must be decided. First, what queueing performance measure(s) should be used. Second, which moments and cross moments of the arrival process must be included in the investigation.

The identities given by (3.4.7) through (3.4.16) simplify our choice of what queueing performance measures to use. These identities show that the mean queue length (the number of customers in the system) at arrival times, L^a, the expected queue length observed at arbitrary times, L^t, the expected waiting time, W_q, and the expected sojourn time, W, of the MR/M/1 queue are interrelated. In particular, if L^t is known, then L^a , W and W_q can be found using (3.4.16), (3.4.9) and (3.4.7), respectively. Therefore, it was decided to use L^t as the performance measure.

Also, we will include the standard deviation, σ, of the queue length at arbitrary times in our investigation. Here,

$$\sigma = \sqrt{Var[\ N(t)\ ]}\ , \qquad (1.1)$$

where $Var[\ N(t)\ ]$ is given by (3.3.4) and (3.3.44). We will show in sections 4.1 and 4.2 that the σ and the L^t behave similarly.

Recall that η, the dominant value of the matrix $\mathbf{R}$, is called *the caudal characteristic* (Remark 5, section 3.1) because it is a measure of the behavior of the tail distribution of the queue length. We will investigate the effect of the moments and the serial correlation of the arrival process on the caudal characteristic in section 4.7.

Next, we need to decide which of the moments and the cross moments of the arrival process affect the L^t. It is well-known that the first two moments (the mean and the variance) of the interarrival times affect L^t in the GI/M/1 queue and hence, various approximation formulas for the L^t using the first two moments have been developed (see, for example, Shanthikumar and Buzacott (1980)). These formulas use the mean and the squared coefficient of variation (*scv*) of the arrival process to estimate the L^t of the GI/M/1 queue. Altiok (1985) suggests that when the $scv>1.0$, the third moment of the arrival process must be included in the formula to improve the estimate value of the L^t.

With the dependency added to the arrival process, some way(s) of measuring this dependency is needed. We use the serial correlation (or the correlation function), $corr(r)$, as the measure of dependency between the interarrival times (see, (2.3.20)−(2.3.22)). This serial correlation is a measure of linear association between any two interarrival times. It may not be the proper measure of dependency for our purposes but we use it because we know of no other better measure. This serial correlation is characterized by $corr$ and ξ (see (1.5)−(1.7)).

Therefore, the following parameters of the arrival process will be used to investigate the behavior of the L^t and the σ of the MR/M/1 queue. They are:

1. $\rho = \frac{1}{z_1 \cdot \mu}$, *the traffic intensity* (z_1 *is the mean interarrival time* (2.3.26)),(1.2)

2. $svc = \frac{z_2}{z_1^2}$, *the squared coefficient of variation* (z_2 *is the variance of the*

   interarrival time (2.3.27)), (1.3)

3. γ , *the coefficient of skewness* (*see* (2.3.29)) , (1.4)

4. *the correlation function* (*the serial correlation*) *given by*

   $$corr\,(r) = corr\,.\,\xi^{\,r-1} \text{ , the lag-}r\text{ correlation,} \qquad (1.5)$$

   where

   $$corr = corr\,(1) = \frac{z_3}{z_2} \text{ , the lag-1 correlation}$$

   (z_3 *is the lag-1 covariance* (2.3.28)), (1.6)

   and

   $$\xi = a+b-1 \quad \text{is the subdominant eigenvalue of } \mathbf{A}. \qquad (1.7)$$

In general, the L^t of the GI/M/1 queue depends on all of the moments of the interarrival time. This is apparent because

$$L^t = \frac{\rho}{1-z} , \qquad (1.8)$$

where z is the unique solution in the unit interval (0,1) of

$$z = \sum_{k=0}^{\infty} a_k z^k , \qquad (1.9)$$

where

$$a_k = \int_0^{\infty} \frac{e^{-\mu t} (\mu t)^k}{k\,!} \, dA(t) , \qquad (1.10)$$

and $A(t)$ is the interarrival time distribution (see Remark 1 in section 3.1).

Therefore, it is somewhat unrealistic to expect ρ, *scv*, γ, *corr* and ξ to explain completely the behavior of the L^t of the MR/M/1 queue. However, in most (not all) of the cases we examined, these 5 parameters seem to be adequate in explaining the behavior of the L^t. The anomalies appear as wiggles in the curves in Figures 6.16D, 6.17–20. There, the curves of the L^t do not strictly decrease from left to right, rather they fluctuate at low values of γ. We do not have a good explanation for these anomalies. We only refer to the earlier comment that ρ, *scv*, γ, *corr* and ξ are not adequate in explaining completely the behavior of the L^t's for these cases and the usual correlations are not appropriate to this nonlinear responses. Similar anomalies (wiggles) are also observed in the corresponding curves for the standard deviation σ (see Figure 6.24D). We will discuss this later in section 4.7.

4.2 The Numerical Methods.

Here, we describe the numerical procedures used to vary, systematically, the values of ρ, *scv* and *corr* and to compute the L^t, the σ and the η of the MR/M/1 queue. First, the traffic intensity ρ=0.3, 0.5, 0.7, and 0.9 are chosen. The mean interarrival time was set to 1 (i.e. $z_1=1$). Hence, the service rate $\mu=\frac{1}{\rho}$(see (0.2)). We also chose scv=0.5, 1.0, 2.0 and 4.0. Since $z_1=1$, then $scv=z_2$ and $z_3=corr.scv$. The lag-1 correlations are chosen to be $corr$=-0.8, -0.5, -0.3, 0 , 0.3 , 0.5, 0.7 and 0.9.

For the prescribed values of z_1, z_2 and z_3 , Algorithm 2.4.34 is used to calculate a, b, m_1, m_2 and k, the parameters of the semi-Markov kernel $\mathbf{A}(t)$ given by (2.3.7). In this algorithm, first, a minimum value of k ($kmin$) is chosen using the left hand side of the inequality (2.4.35). The other k values are chosen such that

k=*kmin* , *kmin*+5, *kmin*+10, *kmin*+15, *kmin*+20 *and* 500.

The last large value of k=500 was chosen because k does not have an upper bound. Once k is fixed, ξ is calculated using (2.4.36) and (2.4.37). Notice that the ξ value generated depends on the k values chosen (i.e. we do not have a direct control over the ξ value).

Next, the interval containing the feasible values of m_2 is found using either (2.4.38) or (2.4.39). Then, 10 values of m_2 are chosen from within this interval. There are many cases where no feasible interval for m_2 is found when $z_3<0$. For example, when ρ=0.5, z_2=2.0, *corr*=-0.8, no feasible value of m_2 is found for all values of k between 7 and 500. Once, k, ξ *and* m_2 are found, m_1, a and b are calculated from (2.4.40), (2.4.43) and (2.4.44), respectively. Finally, the coefficient of skewness γ is calculated using (2.3.29).

These parameters: a, b, m_1, m_2, k and μ, are used as input to the program that calculates the L^t and the σ of the MR/M/1 queue (the description of this algorithm is given in section 3.5). In this algorithm, we first compute the matrix $\mathbf{R}$, using the iterative function (3.5.6) or (3.5.7). To do that, we need to truncate the infinite sum in (3.5.6) and (3.5.7) to a finite sum with upper limit K. We call this K, the truncation index. In the algorithm used, we set the maximum value of the truncation index K at 300 (see (3.5.9−10) on how K is calculated). Also, the maximum number of iterations to compute the matrix $\mathbf{R}$ is set to 500.

Whenever the truncation index K exceeds 300 or the number of iterations exceeds 500, a warning message is issued. The latter is more frequently encountered than the former. For example, at ρ=0.9, *scv*=0.5, 1.0, 2.0, 4.0 , *corr*=0.9, most of the iterations exceeds 500 and hence were terminated at the 500th iteration. For this reason no graph was produced for these cases (see Figures 2.17–20 (no graphs for *corr*=0.9)).

4.3 Plotting the Numerical Results.

All of the graphs in this chapter were produced using the SASGRAPH graphic package. The caption and the label on each of the graph tell what parameters are fixed and what parameters are on the x-axis and on the y-axis, respectively. To join the discrete points with a smooth line, SASGRAPH uses a spline interpolation routine.

When the xy-coordinates of the points are not properly spaced, this interpolation routine may produce curves that wiggle. This is a typical problem encountered by most interpolating algorithm. For example, the GI/M/1 curves in Figures 6.5, 6.6, 6.13, 6.17–20 and the curves in Figures 6.14D, 6.15D show this phenomenon. However, the wiggles in Figures 6.16D, 6.17–20 are not caused by the spline interpolation routine used by SASGRAPH. There, the L^t values actually decrease and increase again in several places. We mentioned this anomalies at the end of section 4.1 and we will discuss this more in section 2.8.

4.4 Results for the GI/M/1 Queue.

In section 2.3 we showed that when $\xi=a+b-1=0$, the interarrival times D_n^a, $n=1, 2, \ldots$, are both uncorrelated and mutually independent (i.e. the arrival process is a renewal process). In this case, the MR/M/1 queue is equivalent to the GI/M/1 queue and the interarrival time distribution is given by the mixtures of Erlang distributions. In this section, we show how the L^t and the σ of this GI/M/1 queue are affected by ρ, *scv and* γ. In section 4.6, we will compare the L^t and the σ of the MR/M/1 queue to the L^t and the σ of this GI/M/1.

To demonstrate how the traffic intensity (ρ), the squared coefficient of variation (*scv*), and the coefficient of skewness (γ) of the arrival process affect the L^t of the GI/M/1 queue, 4 graphs in Figure 4.1 are plotted for fixed values of $\rho=0.7$ and $scv=0.5$, 1.0, 2.0 and 4.0, respectively. In each of the graph, the L^t is on the y-axis, the γ is on the x-axis and each curve in the graph corresponds to a k value, where k is the parameter of the Erlang distribution.

Recall that the L^t of the M/M/1 queue is given by $L^t=\frac{\rho}{1-\rho}$; and hence, $L^t=2.33$ for $\rho=0.7$. Notice that the L^t for the M/M/1 queue depends only on the first moment of the interarrival time and on the first moment of the service time. The horizontal line, $L^t=2.33$, is drawn on all of the 4 graphs in Figure 4.1.

Notice that for fixed values of ρ, *scv* and γ, the L^t of the GI/M/1 queue is not constant. This is expected because this L^t depends on all of the moments of the interarrival time (see (1.8)–(1.10)). Klincewicz and Whitt (1984) showed numerically that if ρ, *scv* and γ are fixed and the restriction on the shape of the distribution of the interarrival time $A(t)$ is imposed (the shape constraint), the range of the possible values of L^t is small. Our graphs in Figure 4.1 are consistent with their result. We also

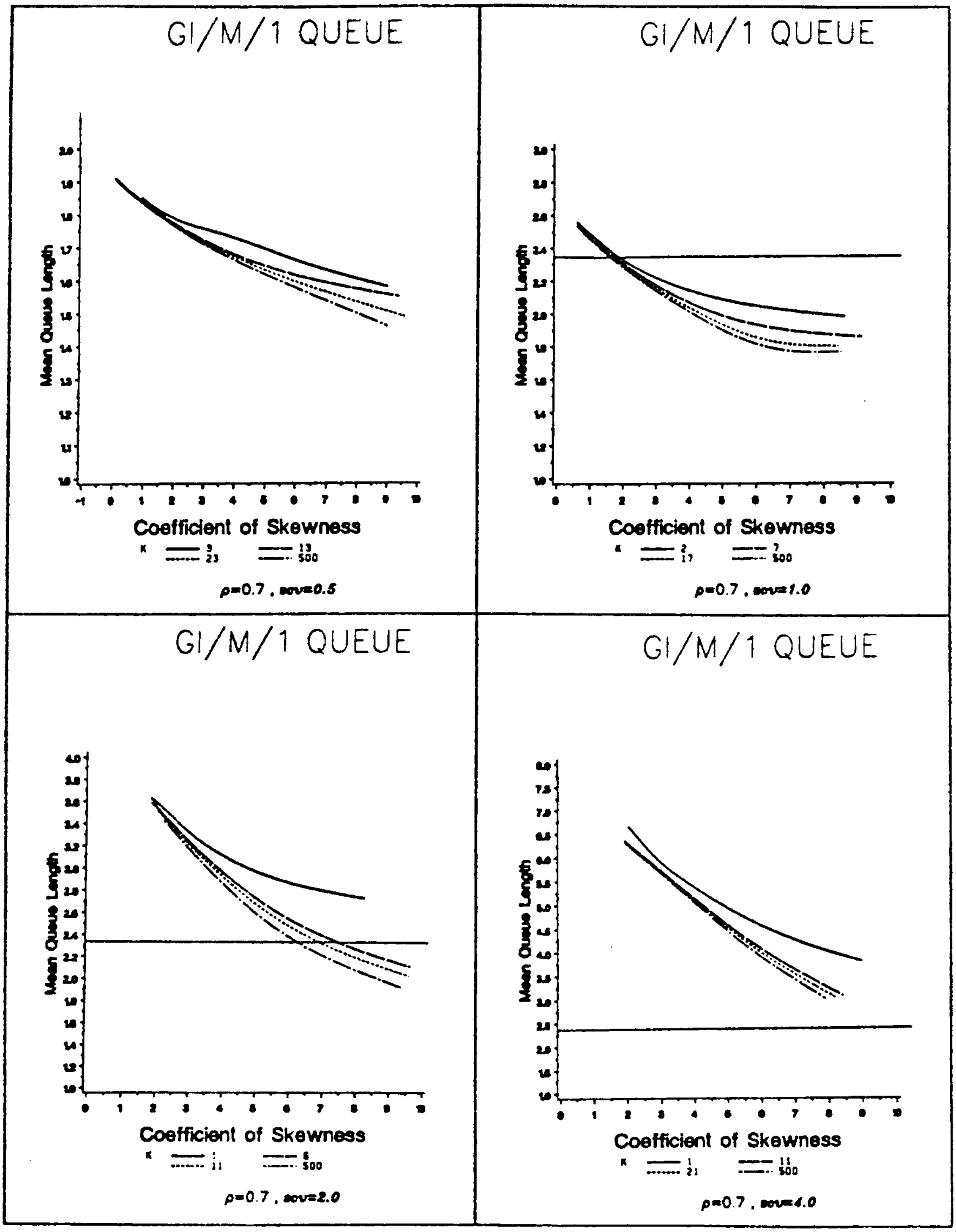

Figure 4.1 The Mean Queue Length of the GI/M/1 Queue.

notice that the range of L^t is smaller for large values of γ.

For the exponential distribution, we know that scv=1.0 and γ=2.0. Figure 4.1B shows that the L^t value of the GI/M/1 queue for scv=1.0 and γ=2.0 are very close to that of the M/M/1 queue (L^t=2.33); i.e., the M/M/1 queue is a good approximation for the GI/M/1 queue with scv=1.0 and γ=2.0.

In practice, when there are only two moments (the mean and the scv) of the interarrival time available to estimate the L^t of the GI/M/1 queue, different formulas are used according to whether scv=1.0 (here the L^t of the M/M/1 queue is used), scv<1.0 or scv>1.0 (see, for example, Shantikumar and Buzacott (1980)). These formulas ensure that when scv<1.0, the estimated L^t will be smaller than the L^t of the M/M/1 queue, and when scv>1.0, the estimated L^t will be larger than that of the M/M/1 queue .

In Figure 4.1A, all of the curves of the L^t of the GI/M/1 queue lie below that of the M/M/1 queue. However, for scv=2.0 and scv=4.0, only the curves for k=1 lie entirely above the L^t=2.33 line. The other curves cross the L^t=2.33 line at $\gamma \sim 0.8$ for scv=2.0 and at $\gamma \sim 12.0$ for scv=4.0. This shows that the two-moment approximation formula for scv>1.0 may not work well when γ is large. The formula corrects the L^t in the wrong direction.

In Figure 4.2, we replace the L^t in Figure 4.1 with the standard deviation, σ, of the queue length. The shapes of the curves in these graphs follow those of the L^t curves in Figure 4.1. For the GI/M/1 queue, the standard deviation σ is larger than the corresponding L^t and for the M/M/1 queue, this is apparent because $L^t=\frac{\rho}{1-\rho}$ and $\sigma=\frac{\sqrt{\rho}}{1-\rho}$ (e.g., for ρ=0.7, L^t=2.33 and σ=2.79).

In Figures 6.5−20 (section 4.6), we superimpose the upper and the lower L^t

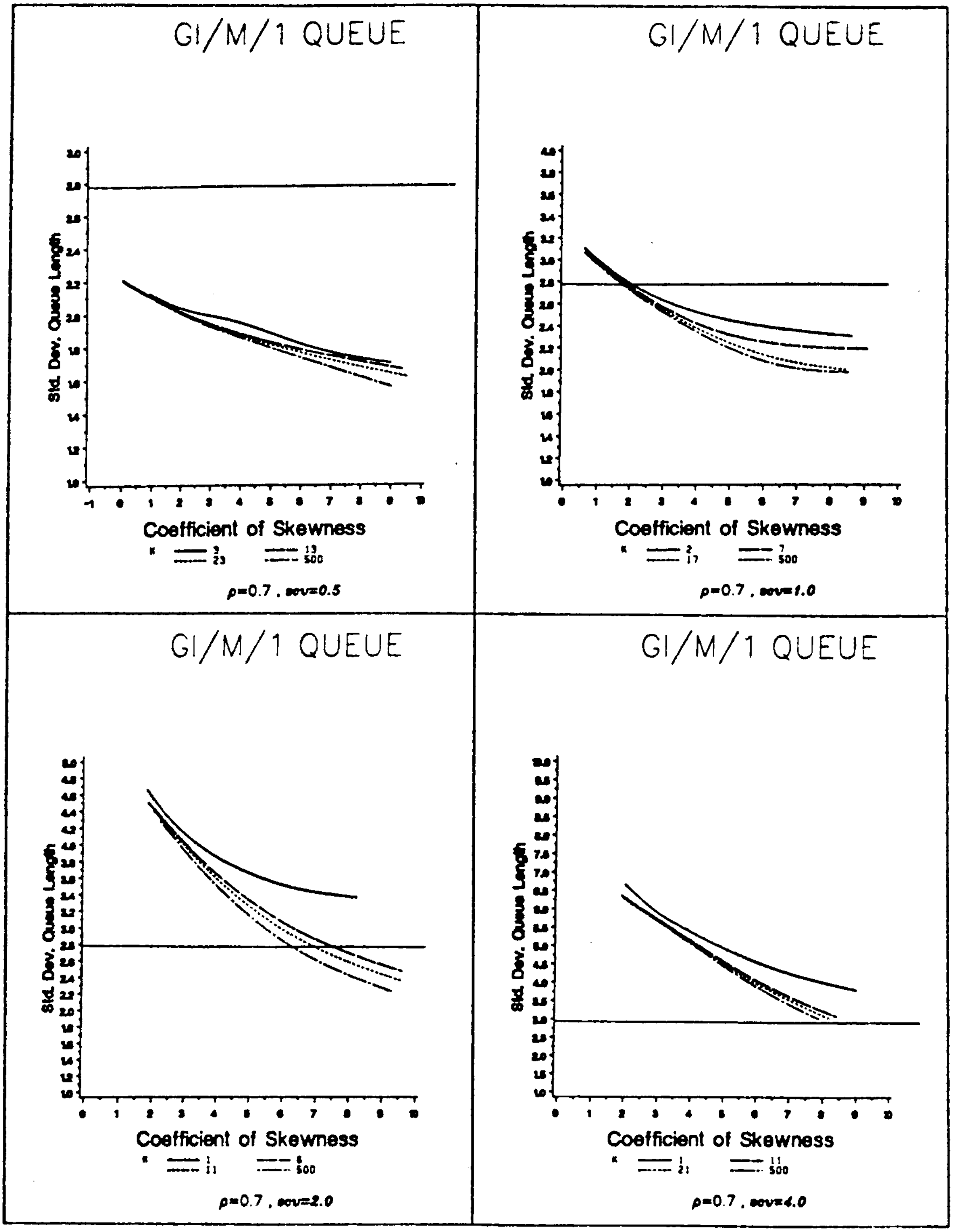

Figure 4.2 The Standard Deviation of the Queue Length of the GI/M/1 Queue.

curves of the GI/M/1 queue on the other L^t curves of the MR/M/1 queue. For example, the two curves labeled k=3 and k=500 of Figure 4.1A appear in all of the 4 graphs in Figure 6.13 as the curves labeled GI/M/1. This way, by looking at the graphs in Figures 6.5−20, one can compare the L^t of correlated queue to the corresponding L^t of the uncorrelated queue for various values of ρ, *scv*, *corr*, ξ and γ. We do the same with the graph for the standard deviation σ (see Figure 4.2 and Figures 4.24−27).

4.5 The Correlated M/M/1 Queue.

Now moving away from the uncorrelated queue, we consider the MR/M/1 queue with ρ=0.7, *scv*=1.0 and γ=2.0. Although the *scv* and the γ of this MRAP match those of the Poisson arrival process, the interarrival distribution is not exponential (it is the mixtures of Erlang distributions) and furthermore, the interarrival times are correlated. We will call this MR/M/1 queue the *correlated M/M/1 queue* to distinguish it from the ordinary M/M/1 queue.

Recall that the serial correlation of this MRAP is completely determined by *corr* and ξ (see (1.5)−(1.7)). Notice also that the signs of *corr* and ξ are the same, i.e., *corr* and ξ are either both positive or both negative. Next, we will compare the L^t of this correlated M/M/1 queue to the uncorrelated M/M/1 queue.

The graph in Figure 5.1 is plotted with the L^t on the y-axis and ξ on the x-axis. The 5 curves in Figure 1.3 correspond to *corr*=0.9, 0.7, 0.5, 0.3 and −0.3. We plot the curves using whatever ξ values are generated by algorithm 2.3.34. Recall that

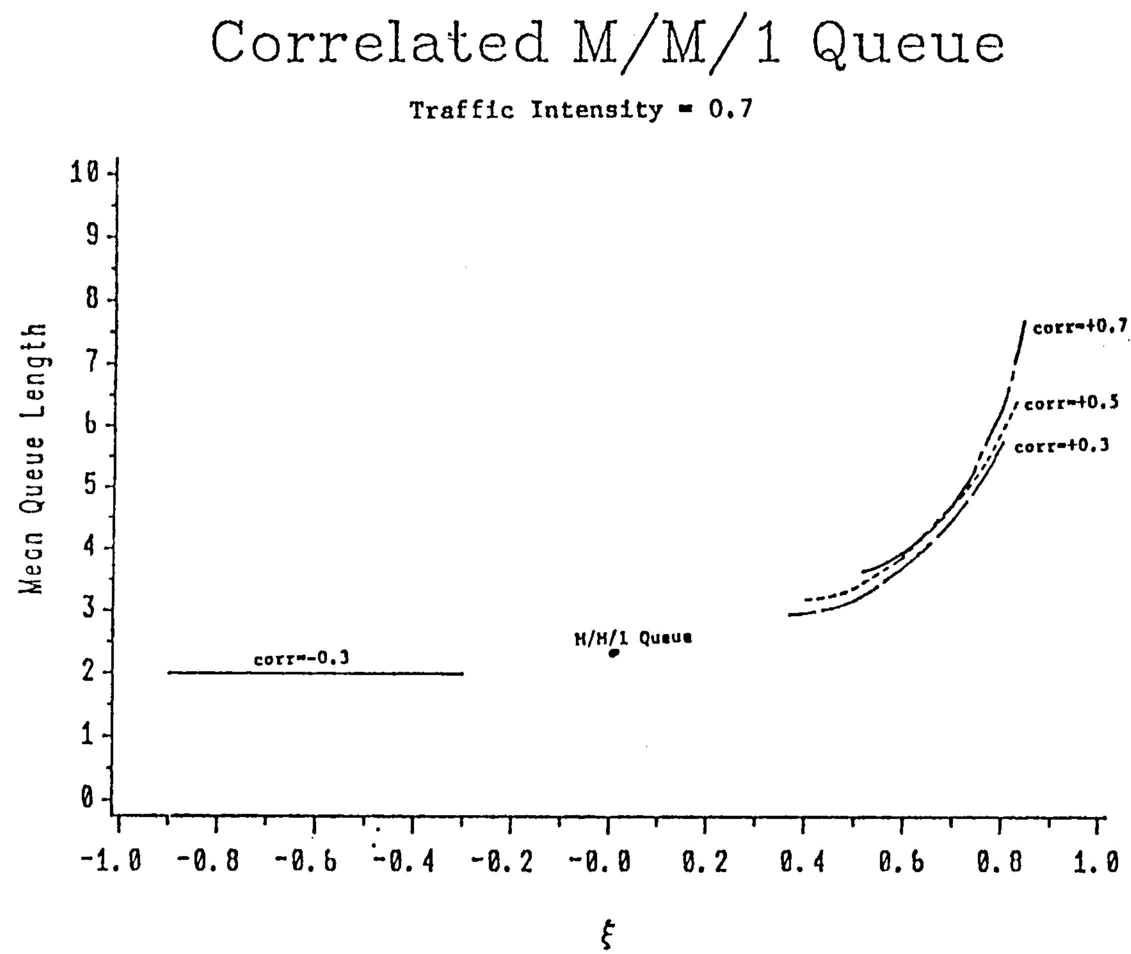

Figure 5.1 Correlated M/M/1 Queue

the L^t for the uncorrelated M/M/1 queue is 2.33. and it appears in Figure 5.1 as a point with label M/M/1.

For *corr*=−0.3, the L^t is approximately 2.0 for all values of ξ. At *corr*=−0.8 (this curve is not plotted), $L^t \sim 1.7$ for all ξ. This shows that the negative lag-1 correlation in the arrival process reduces the mean queue length (when compared to the L^t of the uncorrelated arrival) but not by much. In all of the cases we examined, the $L^t$of the correlated MR/M/1 queue with negative lag-1 correlation is smaller than the L^t of the uncorrelated queue. However, the difference between these L^t's is small. Therefore, in section 4.6 we will look only at the MR/M/1 queue with *corr*>0.

A possible explanation of the above observation is as follows: when *corr*<0, the serial correlation, *corr*(r), alternates in signs because $\xi<0$. Thus, if the lag-1 correlation is negative, then the lag-2 correlation is positive, and so on. This means that a long interarrival time is likely to be followed by a short interarrival time. The next interarrival time will likely be long again and is followed by a short one, etc.. The behavior of this arrival process prevents the congestion in the queue and hence reduces the mean queue length.

When both the *corr* and the ξ are large, the L^t's in Figure 5.1 are large. In particular, when *corr*=0.7 and ξ=0.84, $L^t \sim 7.7$, which is more than 3 times larger than the L^t of the uncorrelated M/M/1 queue. For *corr*=0.9 and ξ=0.98 (this curve is not plotted), $L^t \sim 40$. Here, the L^t is almost 17 times larger than that of the uncorrelated M/M/1 queue. We will elaborate more on this in section 4.6. There, comparisons are made between the L^t of the correlated MR/M/1 queue and the the corresponding L^t of the uncorrelated GI/M/1 queue for various values of ρ, *scv*, *corr*, ξ and γ. The L^t of the correlated M/M/1 queue for ρ=0.3, 0.5, 0.7 and 0.9 can be obtained from the points on the curves corresponding to γ=2.0 in Figures 6.6, 6.10, 6,14 and 6.18, respectively.

4.6 The Effects of the Moments and the Serial Correlation of the MRAP on the Mean and the Standard Deviation of the Queue Length.

There are 5 parameters of the arrival process used in this investigation: ρ, *scv*, *corr*, ξ and γ (see (4.1.2)–(4.1.7) for the definitions of these parameters). Thus, to see how the L^t of the MR/M/1 varies with respect to these 5 parameters, one needs to draw a six-dimensional surface which , of course, is impossible. So, we cut *slices* through this six-dimensional surface and plot these slices as two-dimensional graphs in Figures 6.5–6.20. Each of the curves in these graphs corresponds to fixed values of ρ, *scv*, *corr*, and ξ.

The graphs in Figures 6.5–6.20 are organized as follows: Figures 6.5–6.8 are for traffic intensity ρ=0.3, Figures 6.9–6.12 are for ρ=0.5, Figures 6.13–6.16 are for ρ=0.7 and Figures 6.17–6.20 are for ρ=0.9. For each traffic intensity, there are 4 figures and each of these figures corresponds to *scv*=0.5, 1.0, 2.0 and 4.0, respectively. Within each figure, there are 4 graphs. Each graph is for *corr*=0.3, 0.5, 0.8 and 0.9 (for ρ=0.7 and 0.9, we use *corr*=0.7 instead of 0.8). The missing graphs in Figures 6.17–6.20 for *corr*=0.9 were be explained in section 4.2.

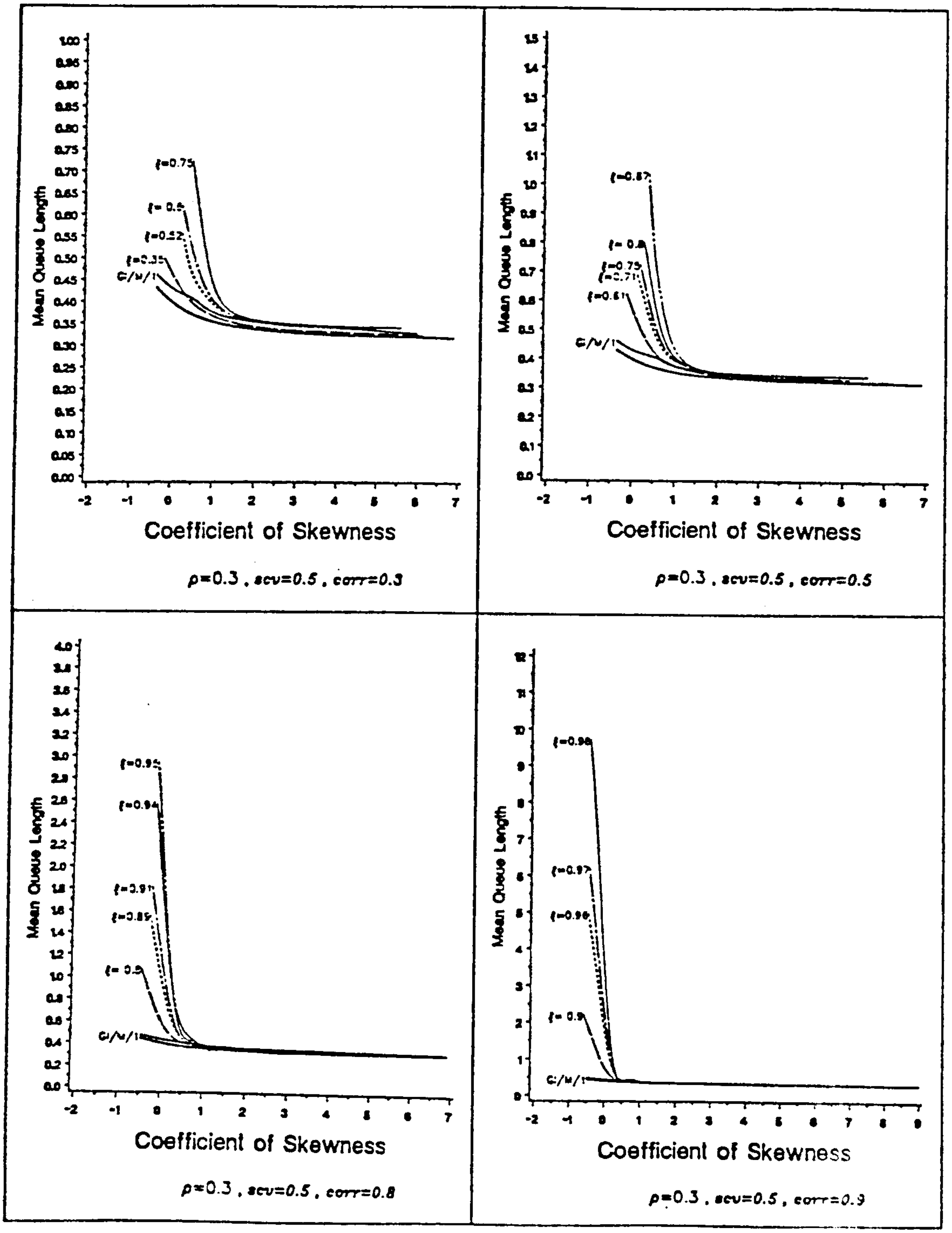

Figure 6.5 The mean queue length of the MR/M/1 queue: ρ=0.3, scv=0.5

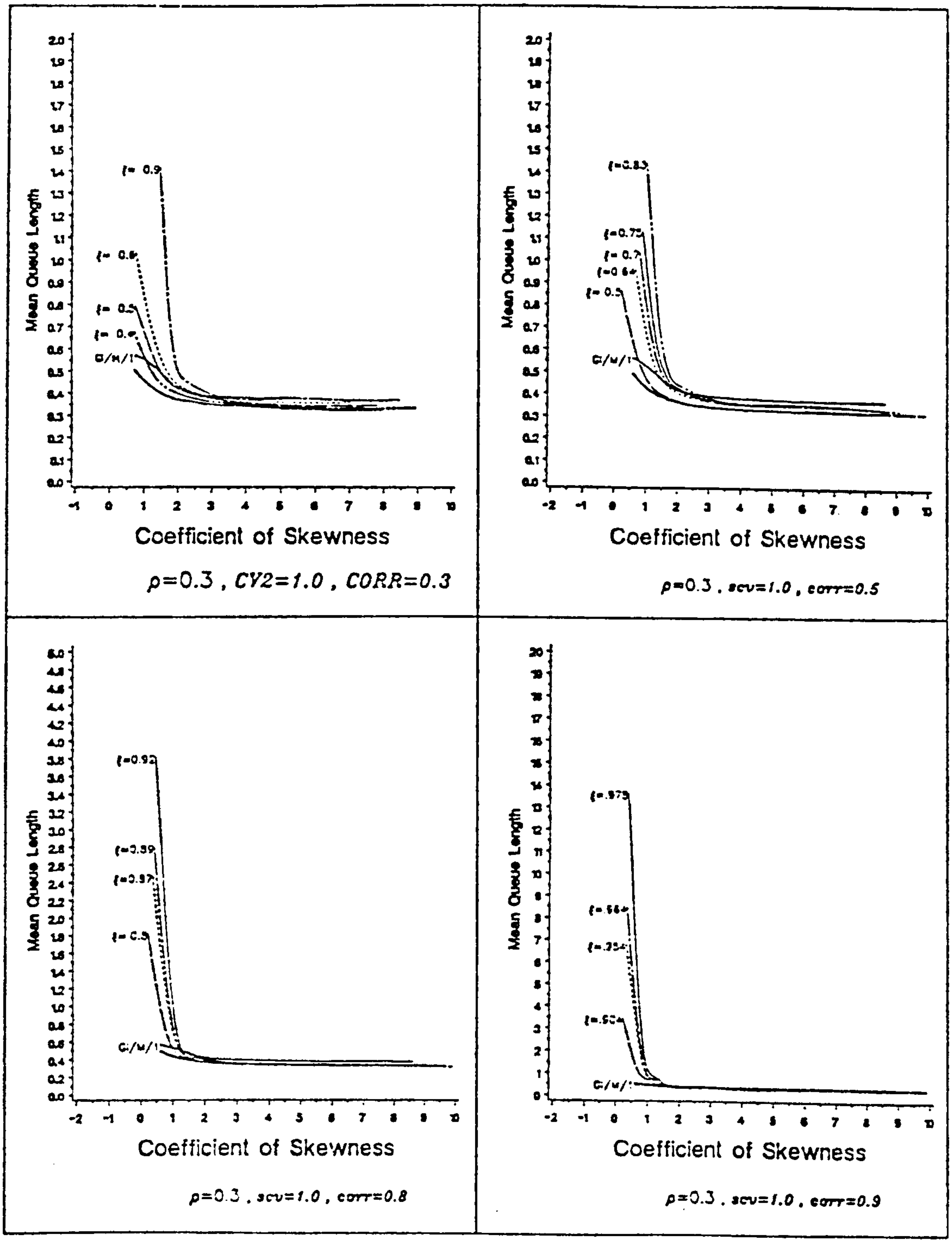

Figure 6.6 The mean queue length of the MR/M/1 queue: ρ=0.3, scv=1.0

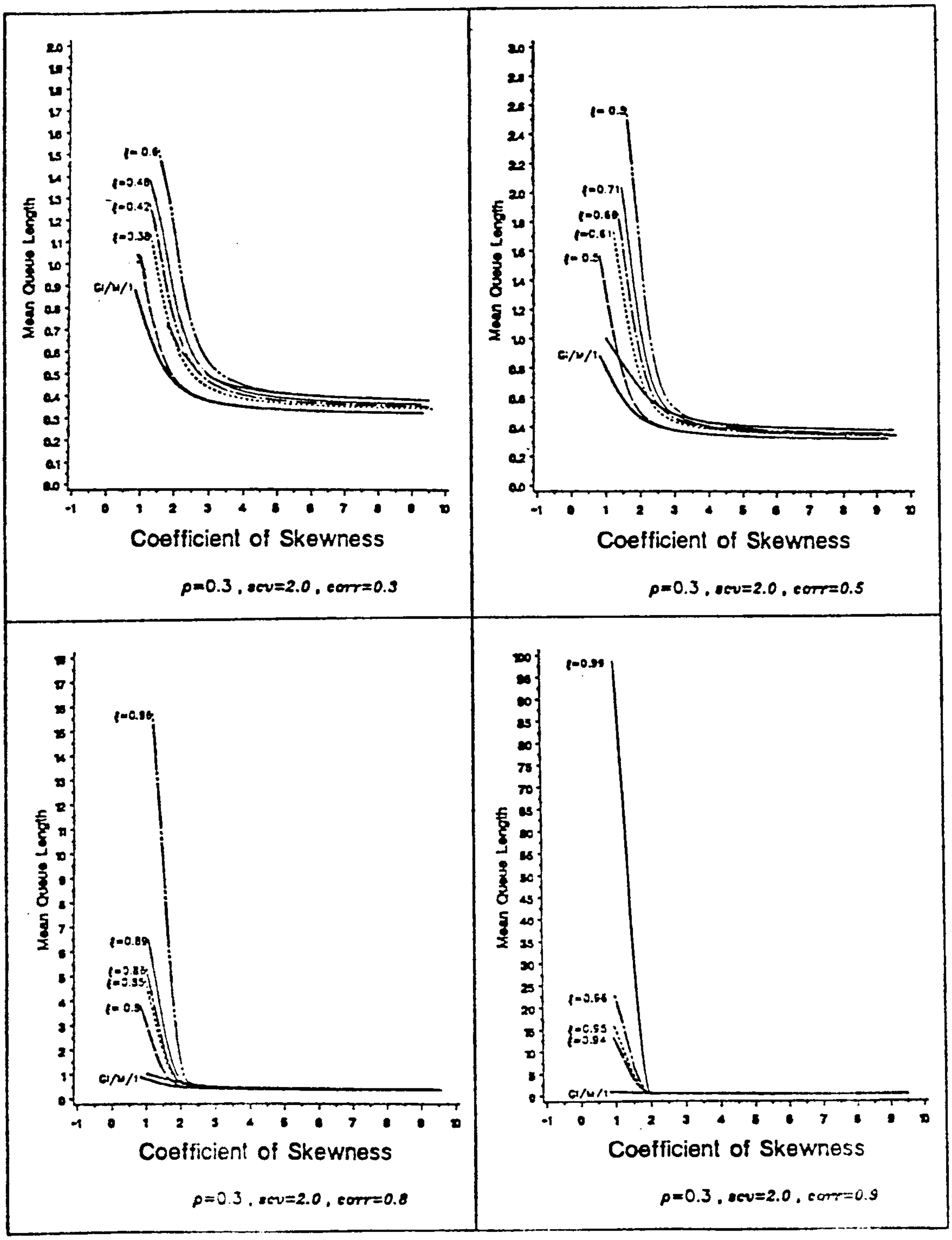

Figure 6.7 The mean queue length of the MR/M/1 queue: $\rho=0.3$, $scv=2.0$

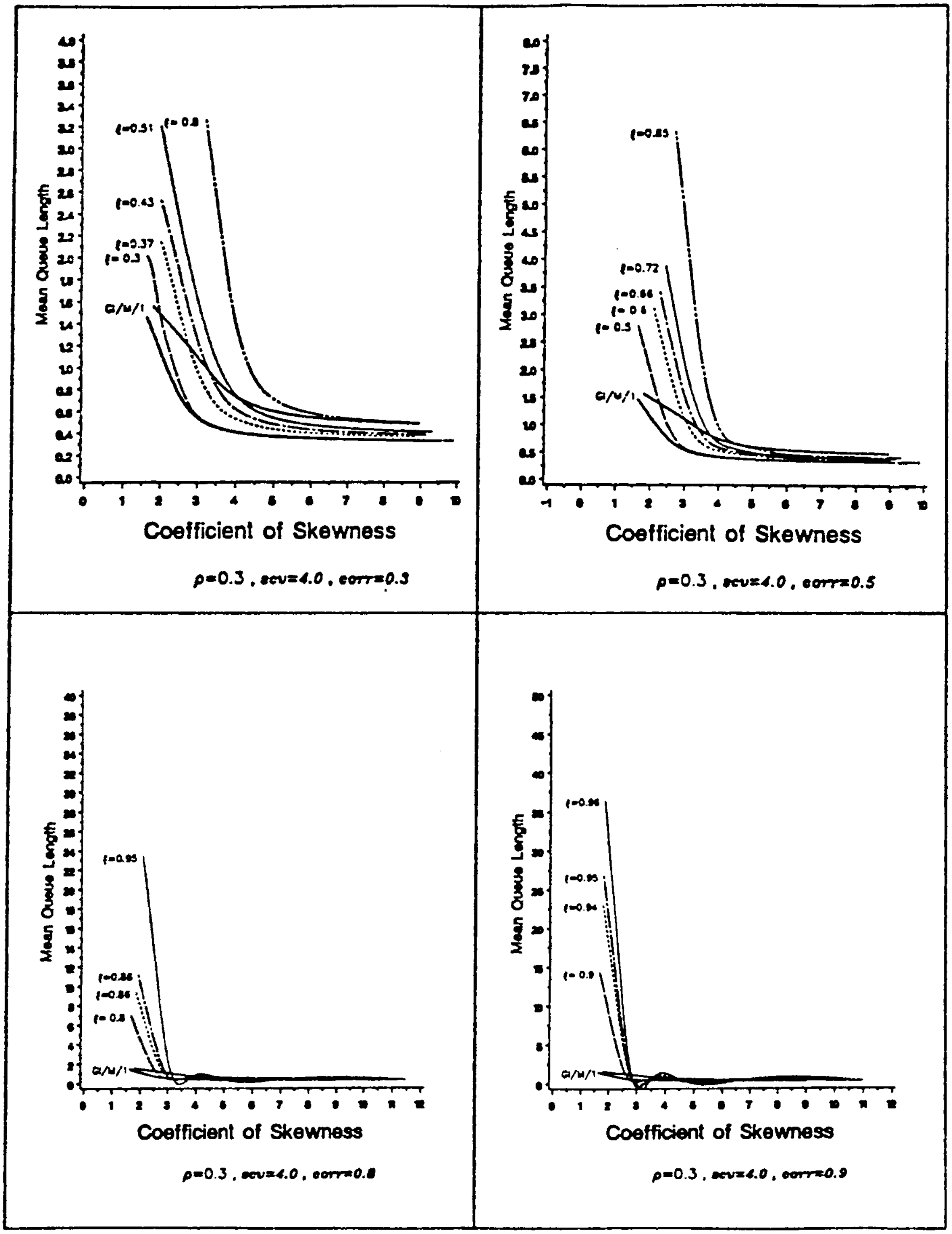

Figure 6.8 The mean queue length of the MR/M/1 queue: $\rho=0.3$, $scv=4.0$

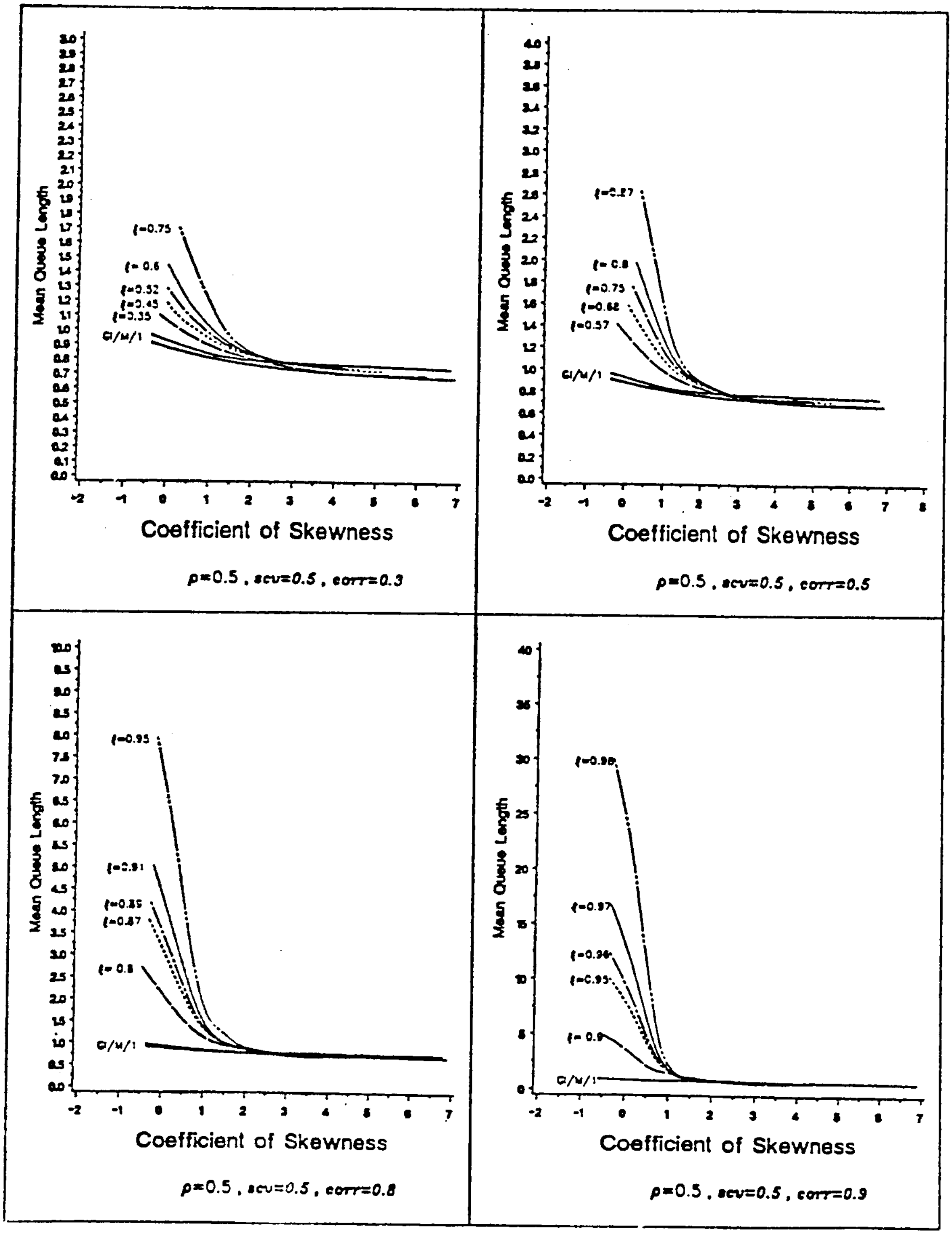

Figure 6.9 The mean queue length of the MR/M/1 queue: ρ=0.5, scv=0.5

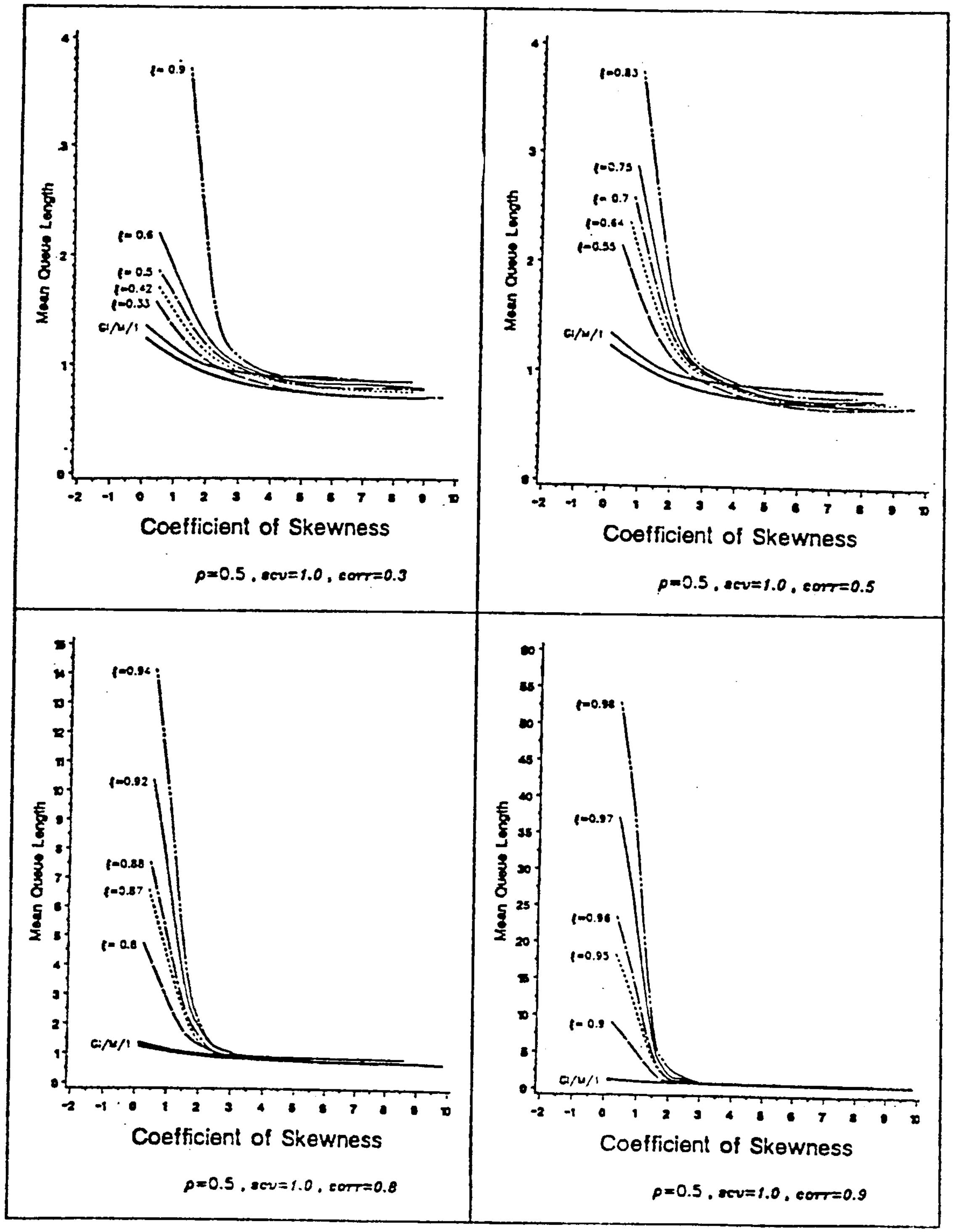

Figure 6.10 The mean queue length of the MR/M/1 queue: $\rho=0.5$, $scv=1.0$

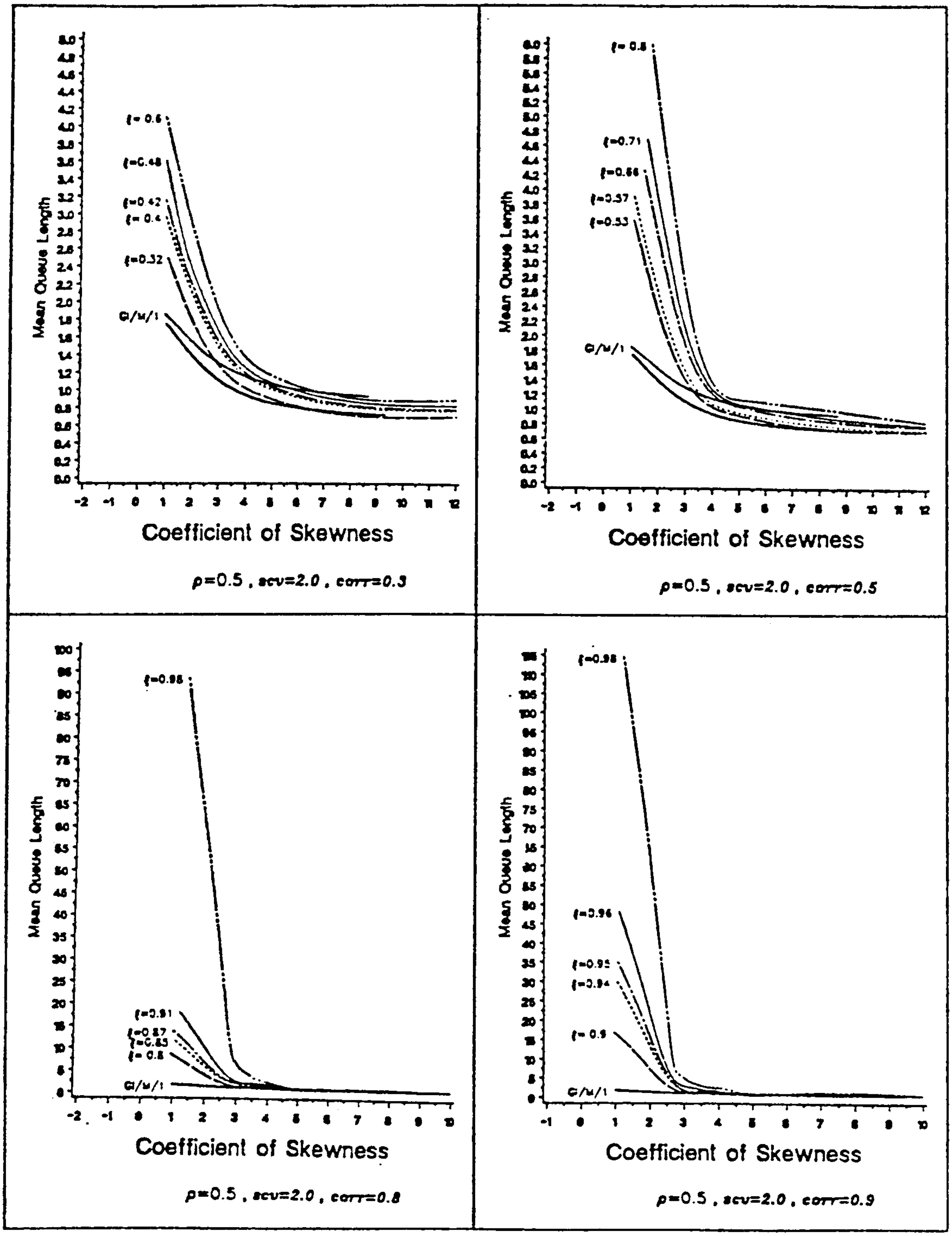

Figure 6.11 The mean queue length of the MR/M/1 queue: ρ=0.5, scv=2.0

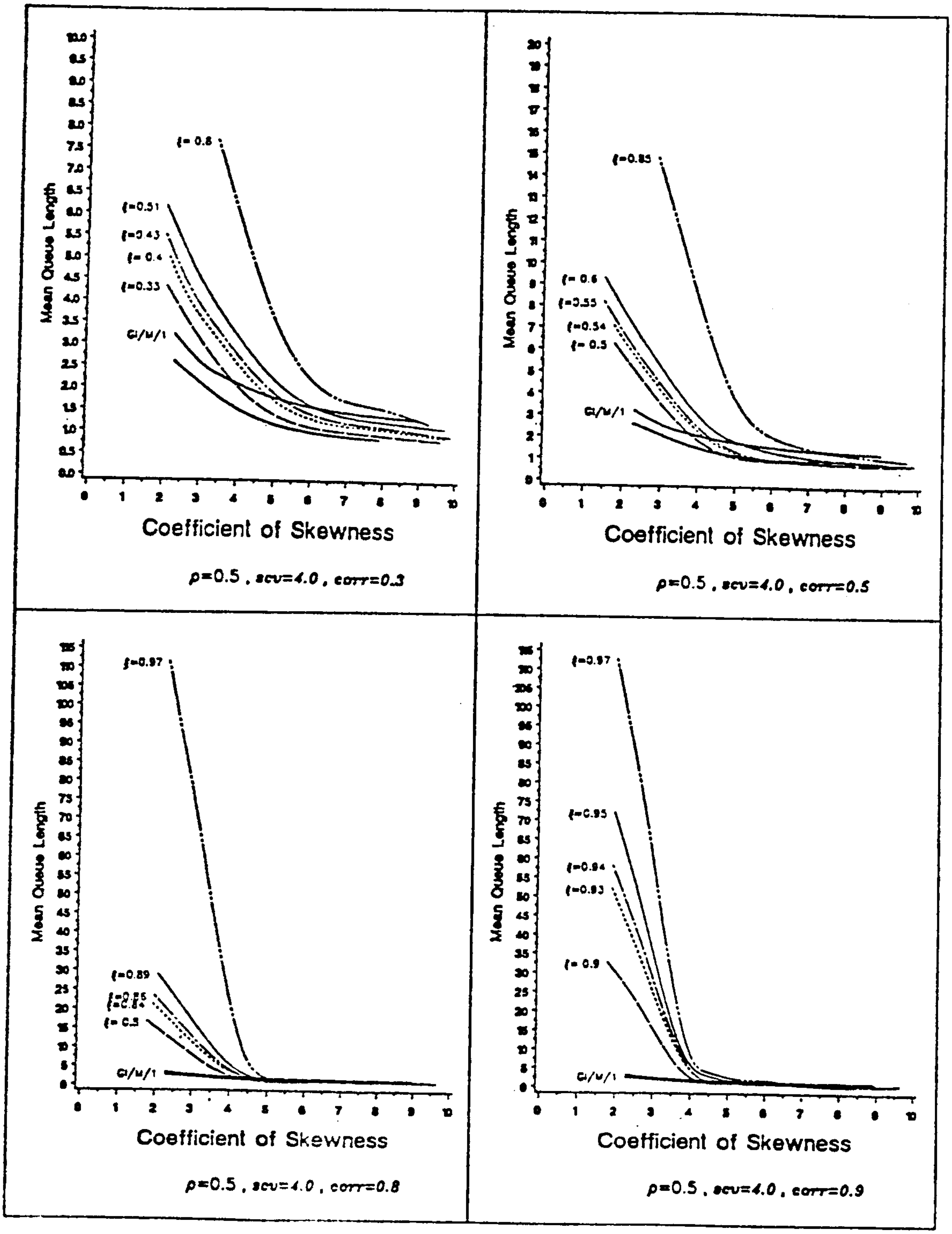

Figure 6.12 The mean queue length of the MR/M/1 queue: $\rho=0.5$, $scv=4.0$

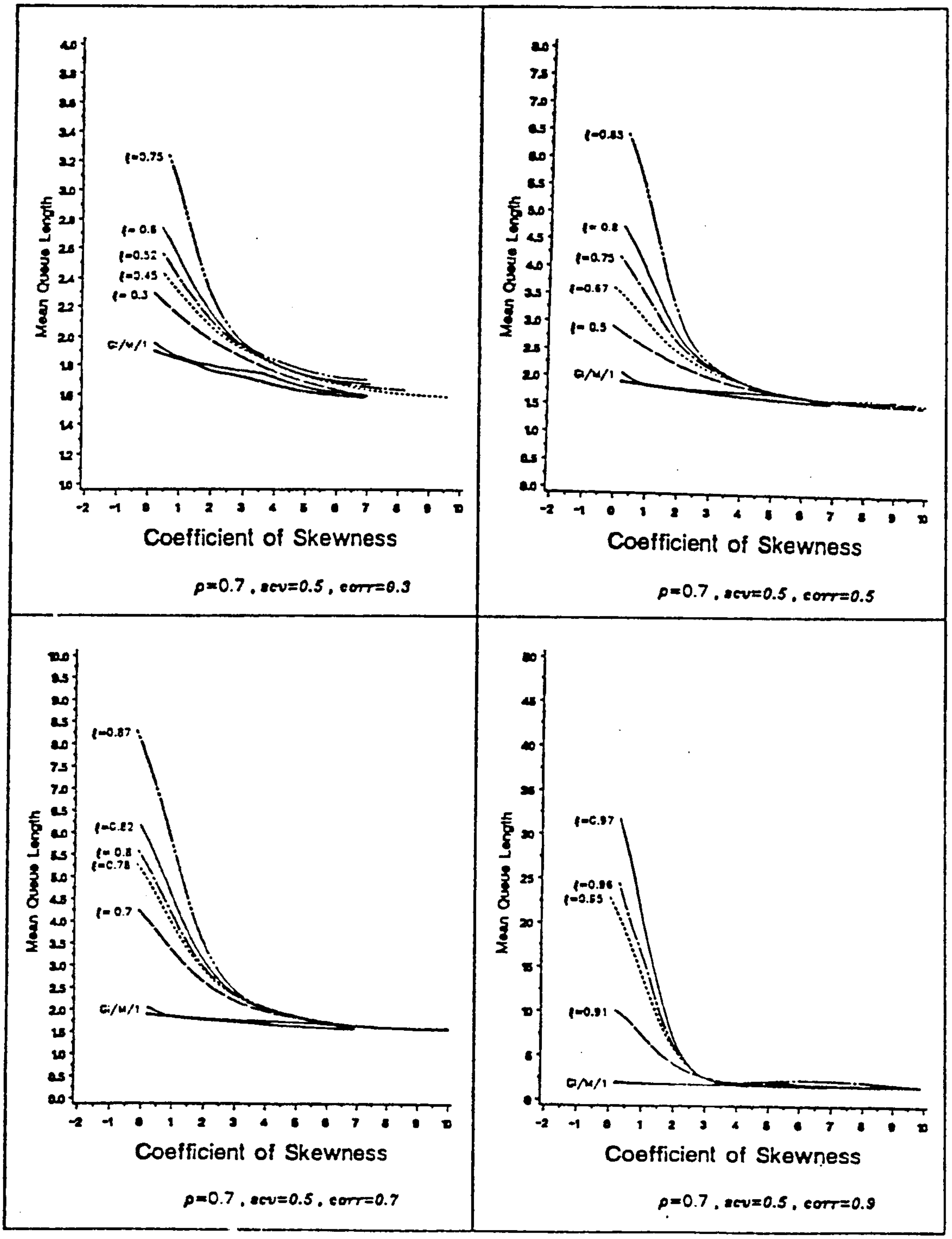

Figure 6.13 The mean queue length of the MR/M/1 queue: $\rho=0.7$, $scv=0.5$

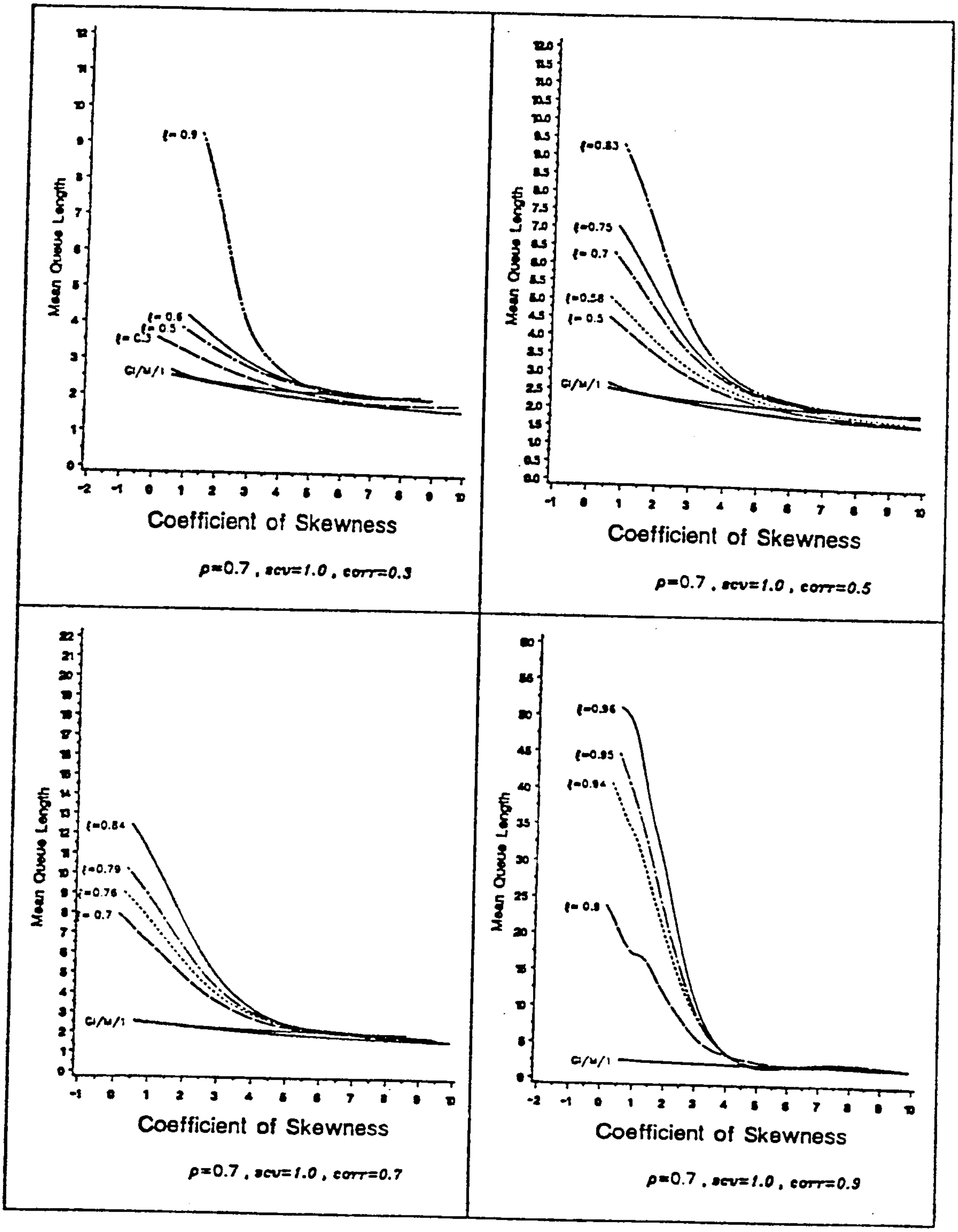

Figure 6.14 The mean queue length of the MR/M/1 queue: $\rho=0.7$, $scv=1.0$

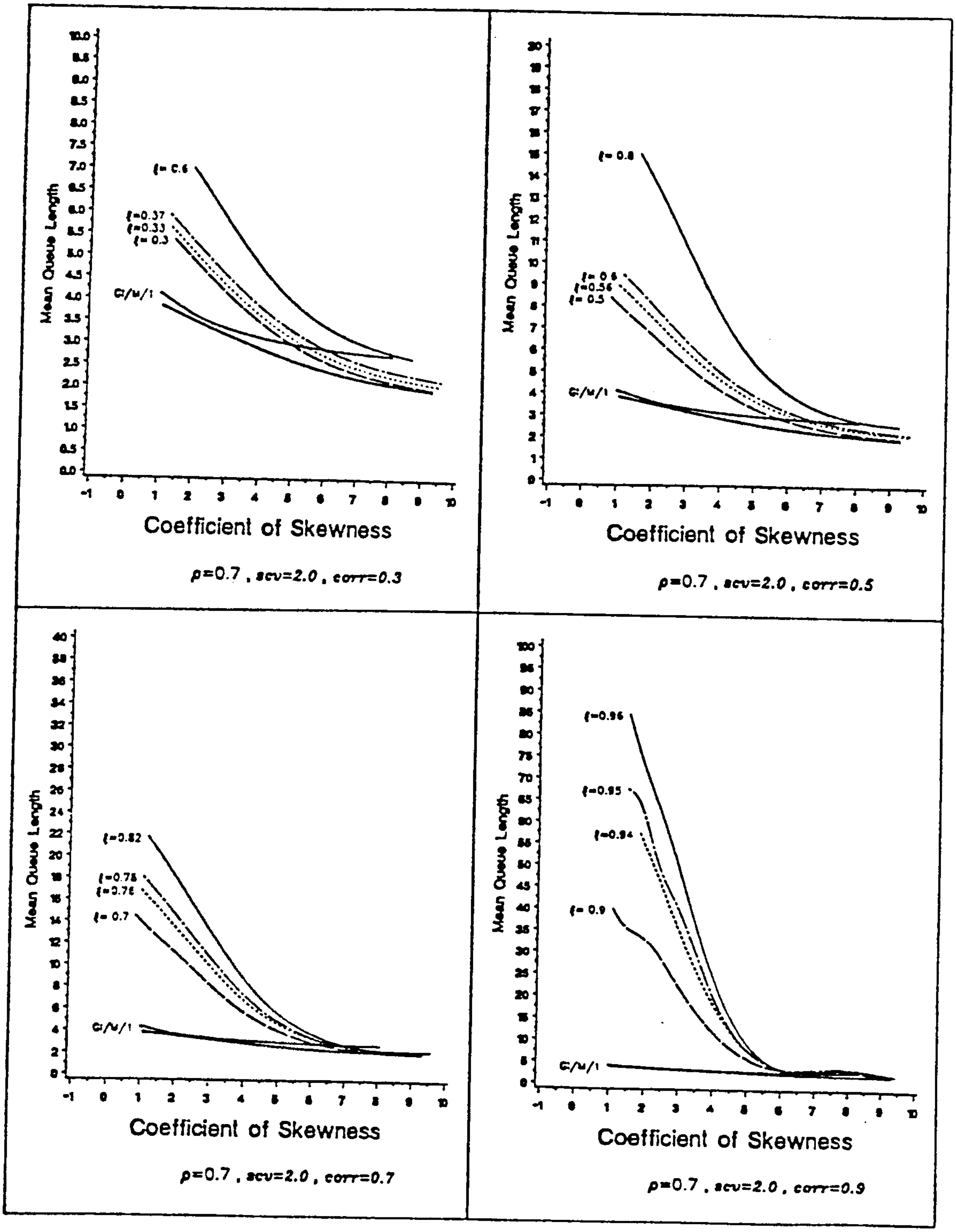

Figure 6.15 The mean queue length of the MR/M/1 queue: ρ=0.7, scv=2.0

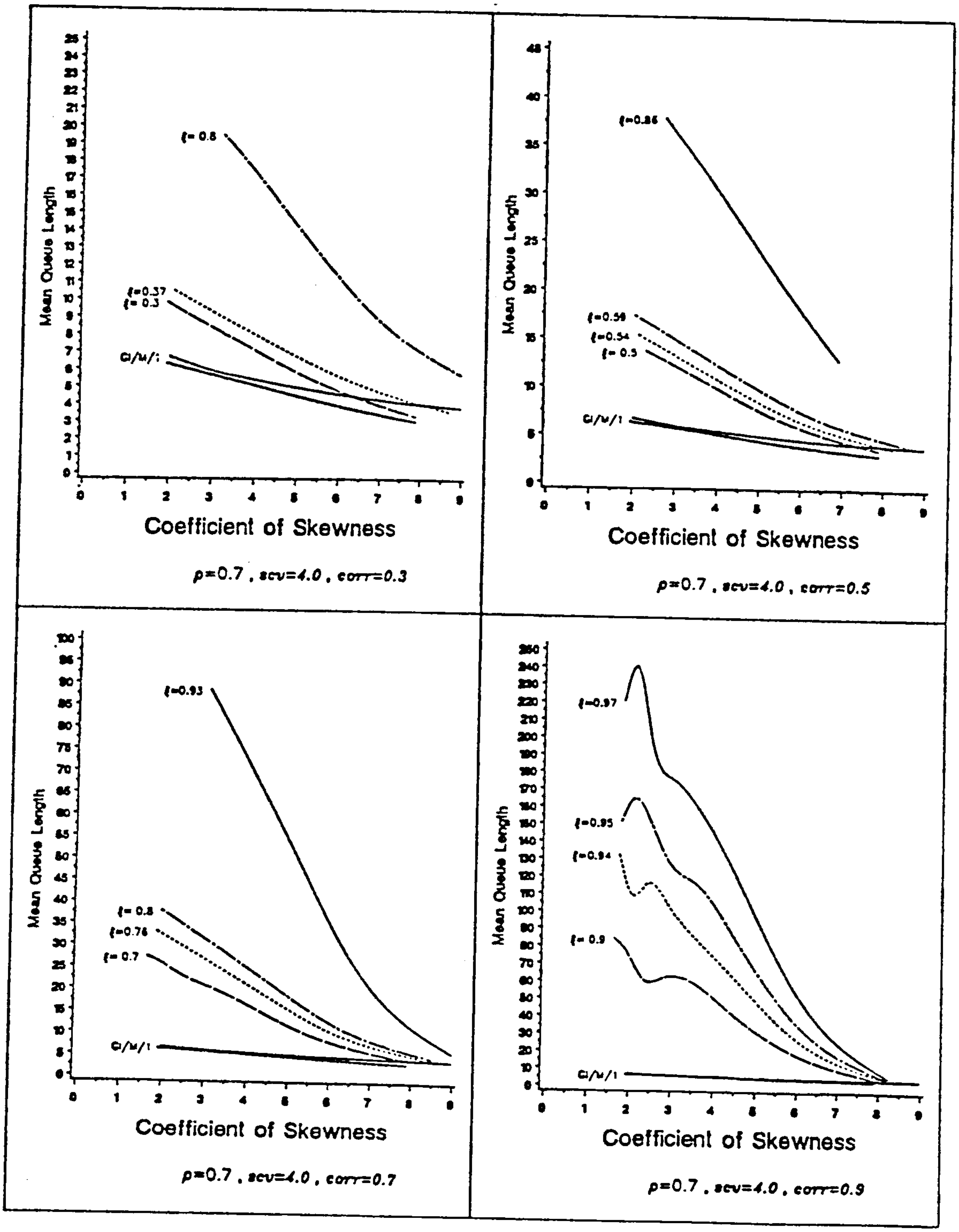

Figure 6.16 The mean queue length of the MR/M/1 queue: $\rho=0.7$, $scv=4.0$

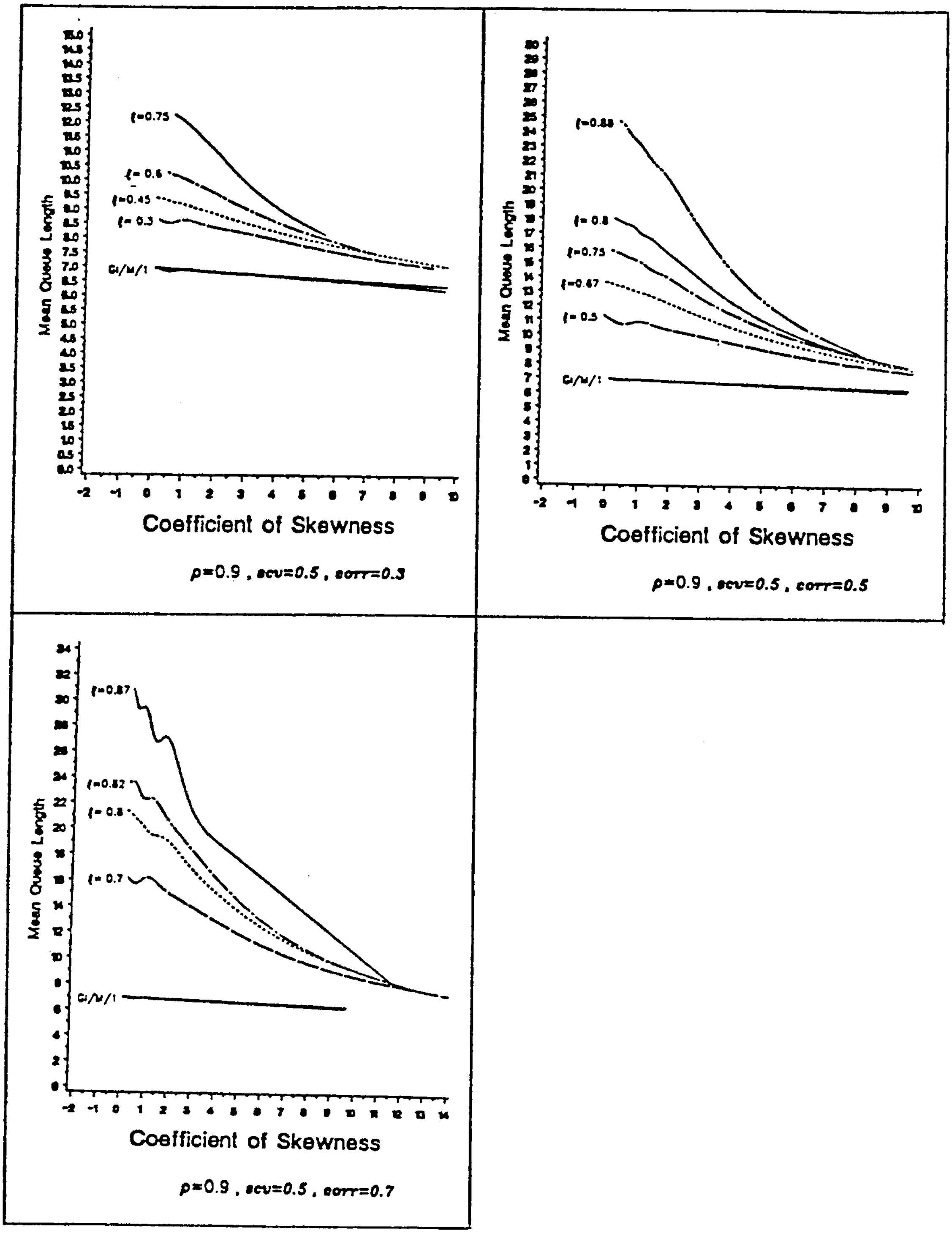

Figure 6.17 The mean queue length of the MR/M/1 queue: ρ=0.9, scv=0.5

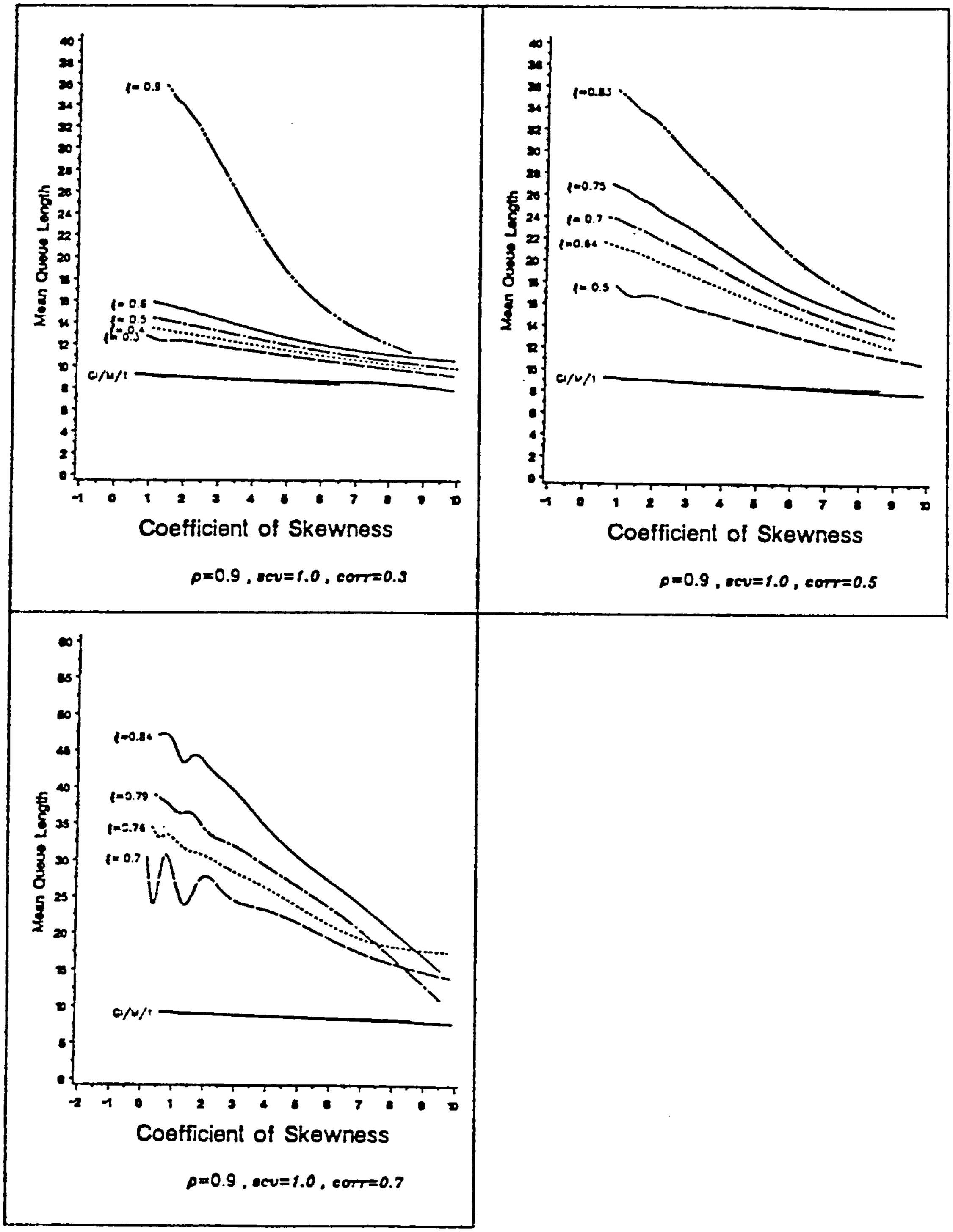

Figure 6.18 The mean queue length of the MR/M/1 queue: ρ=0.9, scv=1.0

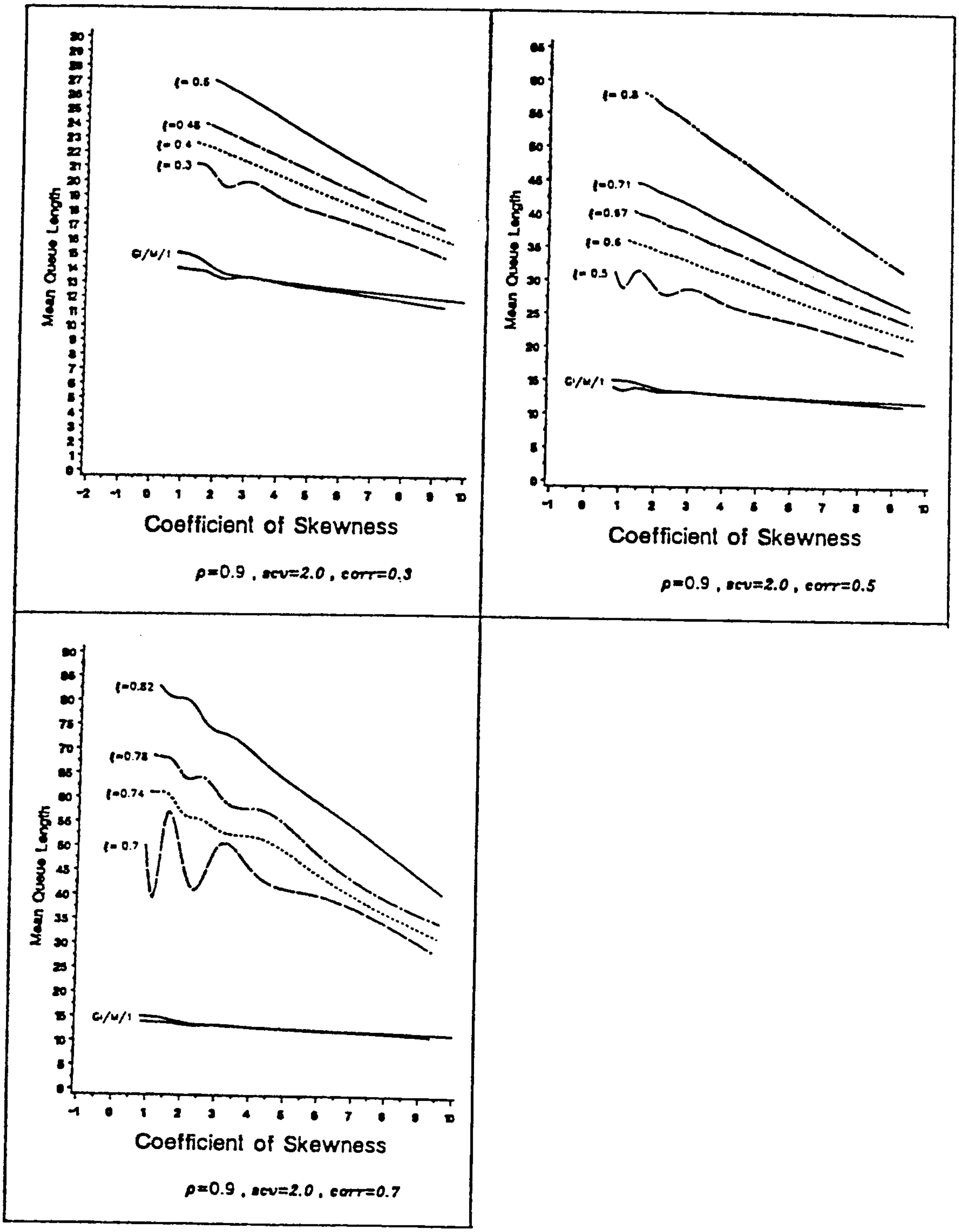

Figure 6.19 The mean queue length of the MR/M/1 queue: $\rho=0.9$, $scv=2.0$

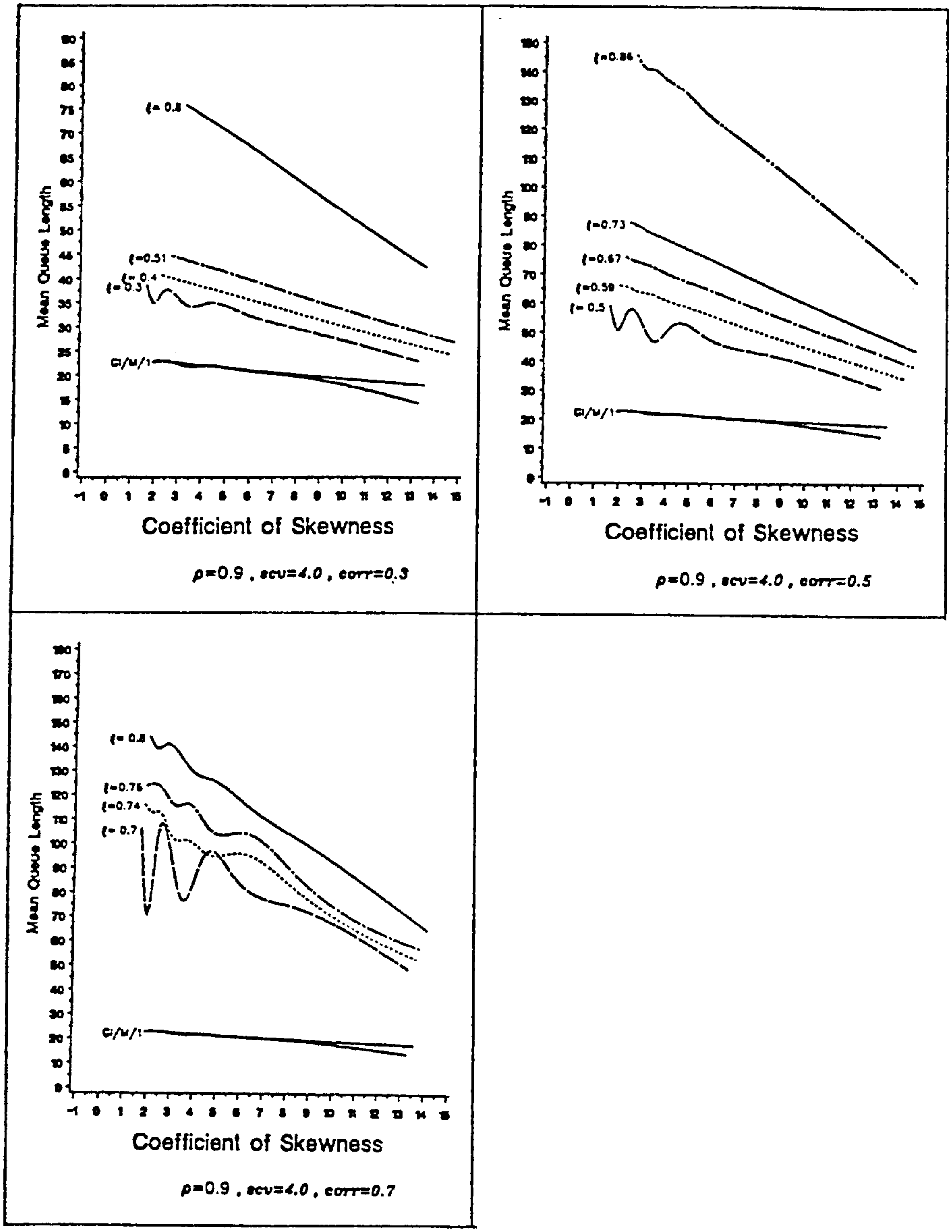

Figure 6.20 The mean queue length of the MR/M/1 queue: ρ=0.9, scv=4.0

4.6.1 The Asymptotic Behavior of L^t as $\gamma \rightarrow \infty$.

One striking observation of the graphs in Figure 6.5 through Figure 6.20 is that the L^t seems to converge to a certain value for large values of γ regardless of the values of *scv*, *corr* and ξ. For example, the L^t curve labeled ξ=0.94 in Figure 6.10C decreases rapidly as γ increases from 0.0 to approximately 3.0 and for γ>3.0, the L^t curve appears almost horizontal. Therefore, we conjecture the following:

CONJECTURE 6.1. As $\gamma \rightarrow \infty$, the L^t of the MR/M/1 queue converges to the L^t of the D/M/1 queue regardless of the values of other moments and the serial correlations.

Since the L^t of the D/M/1 queue depends only on the traffic intensity ρ (see (6.2)−(6.4)), then the L^t of the MR/M/1 queue as $\gamma \rightarrow \infty$ depends only on ρ. Therefore, as $\gamma \rightarrow \infty$, the L^t becomes insensitive to the second and the higher moments of the interarrival time and to the dependency between the interarrival times.

We based our conjecture on the following observation. The interarrival time D_n^a is a positive random variable with distribution function the mixtures of Erlang distributions. As $\gamma \rightarrow \infty$, the shape of the density function of D_n^a becomes very skewed to the right and eventually the mass (area) under the density function will lump around the mean interarrival time z_1. That is, the density of D_n^a becomes a Dirac delta function $u_0(t-z_1)$ (see Kleinrock (1975, pp.341-343)).

The Laplace-Stieltjes transform of $u_0(t-z_1)$ is $e^{-z_1 s}$. Therefore, for the D/M/1 queue,

$$L^t = \frac{\rho}{1-z}, \tag{6.2}$$

where z is the unique solution in (0,1) to

$$z = A^*(\ \mu(1-z)) = e^{-z_1 \cdot \mu(1-z)}\ , \tag{6.3}$$

or

$$z = e^{-\frac{1}{\rho}(1-z)}\ . \tag{6.4}$$

For ρ=0.3, 0.5, 0.7 and 0.9, the L^t's of the D/M/1 queue are found to be 0.31, 0.63, 1.23 and 4.50, respectively. The graphs in Figures 6.5–6.16 agree with the above numbers. For ρ=0.9 , the γ values generated were not large enough to indicate that $L^t \to 4.50$. Nevertheless, the graphs in Figures 6.17–6.20 show the trend of convergence and we believe that $L^t \to 4.50$ as $\gamma \to \infty$. How large γ should be before the L^t becomes *almost insensitive* to *corr* and ξ, will be discussed in section 4.6.2.

When *corr*<0, the L^t of the MR/M/1 queue is smaller than the corresponding GI/M/1 queue but not by much. We mentioned this earlier in section 4.5. However, when *corr*>0 and γ is small, we see that L^t is very sensitive to both *corr* and ξ. This is the subject of the discussion in section 4.6.3.

4.6.2 The Almost Insensitivity Property of the Mean Queue Length.

In all of the graphs in Figure 6.5–20, we notice that the L^t curves decrease rapidly from left to right and at some values of γ they make a turn. After the turn, the curves become almost flat. Let us call the γ values at which the L^t curve makes its turn, *the turning region.*

The range of the L^t values to the right of the turning region is small. This indicates that at these γ values, the L^t is almost insensitive to *corr* and ξ. On the other hand, to the left of the turning region, the L^t is very sensitive to *corr* and ξ. In Table 6.21–23, we tabulate the turning regions for each ρ, *scv* and *corr*.

Table 6.21. The turning regions for $\rho=0.3$.

scv	corr			
	corr=0.3	corr=0.5	corr=0.8	corr=0.9
0.5	(1.0−1.8)	(0.9−1.5)	(0.4−0.9)	(0.2−0.4)
1.0	(1.7−2.7)	(1.5−2.5)	(1.0−2.0)	(0.7−1.5)
2.0	(2.3−4.0)	(2.2−3.7)	(1.7−2.2)	(1.6−1.9)
4.0	(2.5−6.0)	(2.5−4.7)	(2.0−4.0)	(2.0−3.0)

Table 6.22. The turning regions for $\rho=0.5$.

scv	corr			
	corr=0.3	corr=0.5	corr=0.8	corr=0.9
0.5	(1.0−2.7)	(1.0−2.5)	(0.8−2.0)	(0.7−1.4)
1.0	(2.2−3.5)	(1.5−3.2)	(1.5−2.8)	(1.7−2.5)
2.0	(2.5−6.0)	(3.0−5.0)	(2.7−4.6)	(2.5−3.5)
4.0	(4.9−9.0)	(4.0−7.0)	(4.0−5.0)	(3.7−4.5)

Table 6.23. The turning regions for $\rho=0.7$.

scv	corr			
	corr=0.3	corr=0.5	corr=0.7	corr=0.9
0.5	(2.0−6.0)	(2.0−5.5)	(1.5−5.0)	(2.0−3.5)
1.0	(3.0−6.0)	(3.0−6.0)	(3.0−5.0)	(3.0−4.5)
2.0	(5.0−10.0)	(1.0−7.0)	(4.5−7.0)	(4.5−6.0)
4.0	(>10.0)	(5.0−9.0)	(6.0−8.5)	(6.0−8.2)

For ρ=0.9, the turning regions are larger than 10 and are not seen in Figures 2.17−2.20, and hence no table for the turning regions is produced. Nevertheless, Tables 6.21−6.23 give indication how the turning region is affected by *scv*, *corr* and ξ.

As *corr* increases, the turning region shifts to the left (i.e. the turning region occurs at lower γ values). As *scv* increases, the turning region shifts to the right. Increasing ξ values also shifts the turning region to the right. The information about the location of the turning regions is important because it tells at what values of γ the L^t is very sensitive and is almost insensitive to corr and ξ. We will illustrate these cases as we go through specific examples in section 4.6.3.

4.6.3 The Effect of ρ, *scv*, *corr*>0 , ξ>0 and γ on the L^t.

In this section, we investigate how the individual parameters of the arrival process (ρ, *scv*, *corr*, ξ and γ) affects the L^t of the MR/M/1 queue. The graphs in Figures 6.5−20 are the source of our information.

To see how the L^t varies with respect to the traffic intensity ρ (for fixed values of *scv*, *corr*, ξ and γ), one only needs to look at every 4th figure in Figures 6.5−20. For example, to see how the L^t varies with respect to ρ for *scv*=1.0, *corr*=0.5, ξ=0.7 and γ=2, we examine Figures 6.6B, 6.10B, 6.14B and 6.18B and pick the L^t values from the curves labeled ξ=0.7 at γ=2. Here, L^t=0.4, 1.4, 4.5, 22.0 for ρ=0.3, 0.5, 0.7 and 0.9, respectively. This shows that the L^t increases as the traffic intensity gets higher.

This example corresponds to the correlated M/M/1 queue because *scv*=1.0 and γ=2.0 (see section 4.5). For the M/M/1 queue, L^t=0.4, 1.0, 2.3 and 9 when ρ=0.3, 0.5, 0.7 and 0.9, respectively. Notice that the L^t for ρ=0.7 and 0.9 are significantly larger for the correlated M/M/1 queue (with *corr*=0.5, ξ=0.7) than those of the uncorrelated M/M/1 queue. This shows the combined multiplicative effect of high ρ value with *corr* and ξ on the L^t.

To see how the L^t varies with respect to the *scv* for fixed values of ρ, *corr*, ξ and γ, we look at the 4 consecutive figures with the same ρ values and find the 4 graphs with the prescribed *corr* value. For example, holding ρ=0.5, *corr*=0.9, ξ=0.5 and γ=2 constant, we find that L^t=0.9, 1.5, 9 and 30 for *scv*=0.5, 1.0, 2.0, 4.0, respectively. These values are obtained from Figures 6.9D, 6.10D, 6.11D and 6.12D. This again shows that the L^t increases when the *scv* increases.

The corresponding L^t value of the GI/M/1 queue are L^t=0.8, 1.0, 1.5 and 2.8, respectively, for the above 4 values of *scv*. Notice that the L^t values for *scv*=2.0 and 4.0 are significantly larger than those of the uncorrelated queue: 6 time and 10 times larger, respectively. This shows the combined effect of the large *scv* together with the serial correlation (*corr*=0.9 and ξ=0.5) on the L^t is significant.

Each curves in Figures 6.5−20 shows how the L^t varies with respect to γ for fixed values of ρ, *scv*, *corr* and ξ. There, we defined *the turning region* as the γ values at which the L^t curve, from left to right, makes a turn and becomes almost flat. The approximate values of the turning regions for various ρ, *scv* and *corr* are given in Tables 6.21−23.

To the right of the turning region, the L^t is very sensitive to both *corr* and ξ (i.e., the serial correlation). There the L^t increases rapidly as both *corr* and ξ increase. For example, consider the curve corresponding to ξ=0.97 in Figure 6.12D. Here, the traffic intensity, ρ=0.5, is moderate and the *scv*=4.0, *corr*=0.9, ξ=0.97 (the values of *scv*, *corr* and ξ are high). As γ decreases: γ=6.8, 4.0, 2.7, 1.97, we observe the spectacular increase of the L^t: L^t=1.0, 8.3, 74.0, 112.7. The L^t at γ=1.97 is 112 times larger than the L^t at γ=6.8. This shows how a low value of γ combined with high values of *scv* and *corr* can have pronounced effect on the L^t.

For the GI/M/1 queue, L^t=0.9, 2.0, 2.4 and 2.9, respectively, for the above γ values. The differences between these two sets of L^t values are due to the dependency in the arrival process. At γ=6.8, the L^t is about the same for both the correlated and the uncorrelated queue. However, at low γ value, we see a significant combined effect of high values of *corr* and ξ and low value of γ. At γ=1.97, the L^t of the correlated queue is 39 times larger than the uncorrelated L^t. Notice that this spectacular larger L^t occurs at the γ values to the left of the turning region.

To the left of the turning region, the L^t flattens. This implies that at those γ values, the L^t is almost insensitive to the serial correlation. This leads us to conjecture (Conjecture 6.1) that as $\gamma \rightarrow \infty$, the L^t of the MR/M/1 queue converges to the L^t of the D/M/1 queue regardless of the serial correlation and *scv* of the arrival process. Recall that the L^t of the D/M/1 queue depends only on ρ (see (4.2.2−4)). We have not found any literature that mentions this insensitivity of the L^t for large γ. It would be worthwhile to prove or disprove this conjecture.

The 4 graphs within each figure shows how the L^t varies with respect to *corr* when all of the other parameters are fixed. For example, from Figure 6.15A−D (ρ=0.7, scv=1.0, γ=2 and ξ=0.7), L^t=4.3, 4.6, 4.9 for *corr*=0.3, 0.5 and 0.7. Although we do not have the L^t value for *corr*=0.9, Figure 6.15D gives indication that the L^t is larger than 5. Again, increasing the value of *corr* increases the L^t for γ values to the left of the turning region. To the right of the turning region, the L^t is almost insensitive to *corr* and ξ.

Earlier we mentioned that combined multiplicative effect of low γ value and high *scv*, *corr* and ξ values. Here is an example: in Figure 6.16D, at ρ=0.7, *scv*=4.0, γ=1.8, *corr*=0.9 and ξ=0.97, the mean queue length L^t=220.5. The corresponding uncorrelated L^t is only 6.4.

Finally, to see how the L^t varies with respect to ξ for fixed ρ, *scv*, *corr* and γ, we look at the graph that matches the ρ, *scv* and *corr* values. For example, when ρ=0.9, *scv*=1.0, *corr*=0.9 and γ=3, the curves in Figure 6.18B show that L^t=15.5, 18.5, 20, 22 and 29 for ξ=0.50, 0.64, 0.70, 0.75 and 0.83, respectively. The L^t increases with the increasing value of ξ when γ is to the left of the turning region; and to the right of the turning region, the L^t is almost insensitive to ξ

In conclusion, when the coefficient of skewness of the interarrival time is to the left of the turning region, one can expect a significant effect of the dependency in the arrival process on the queueing performance measures. On the other hand, when the coefficient of skewness is to the right of the turning region, the queueing performance measures is almost insensitive to the dependency in the arrival process.

4.6.4 Results for the Standard Deviation of the Queue Length.

For the MR/M/1 queue, $Var[N(t)]$ is given by (3.3.43-44) and hence the standard deviation of the queue length, $\sigma = \sqrt{Var[N(t)]}$. For all of the L^t values computed in Figures 6.5−20, the corresponding σ values are also computed. We observed that the value of the σ is larger than the corresponding L^t but the behavior of the standard deviation σ with respect to ρ, *scv*, *corr*, ξ, and γ follows that of the L^t.

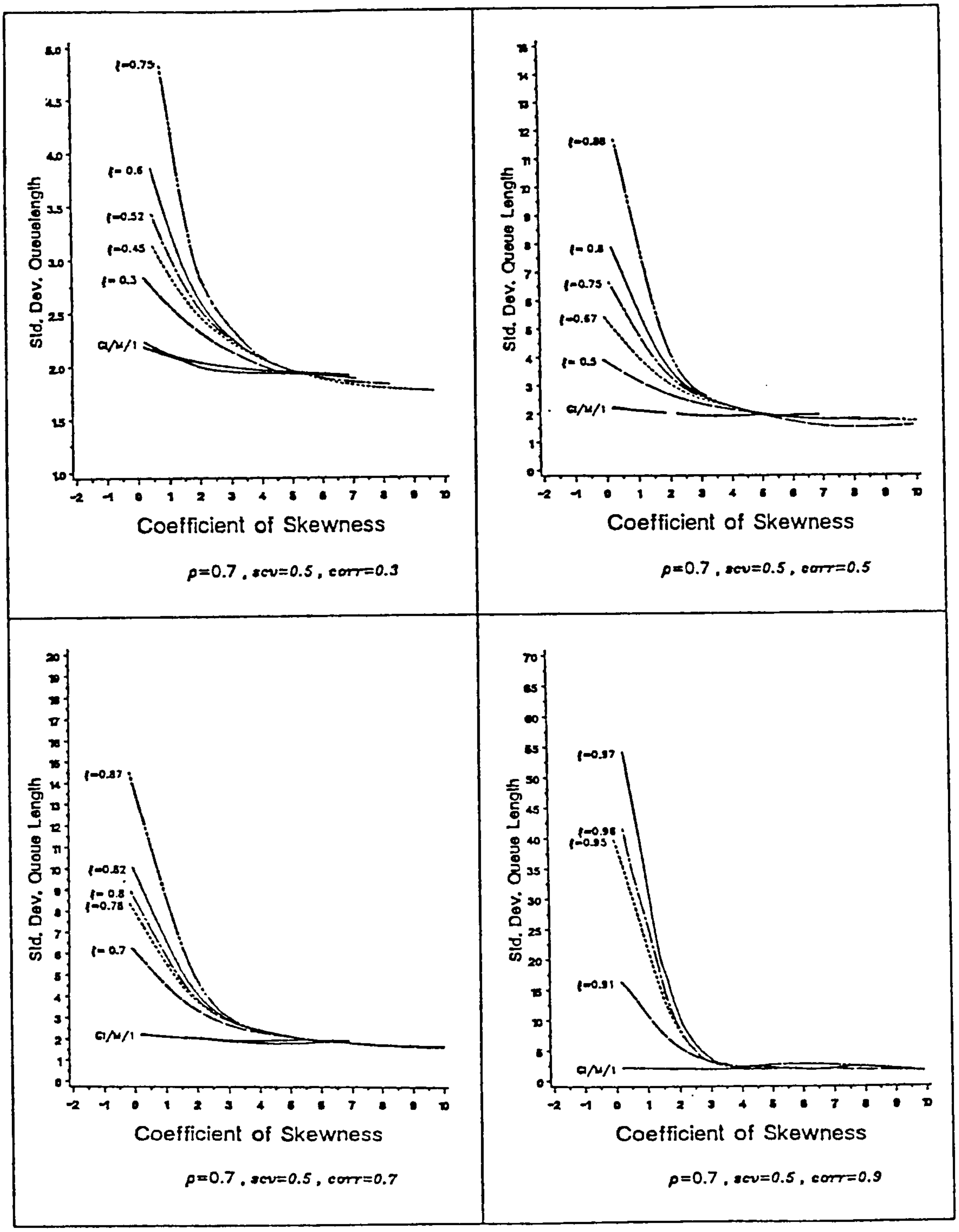

Figure 6.24 The standard deviation of the MR/M/1 queue: ρ=0.7, scv=0.5

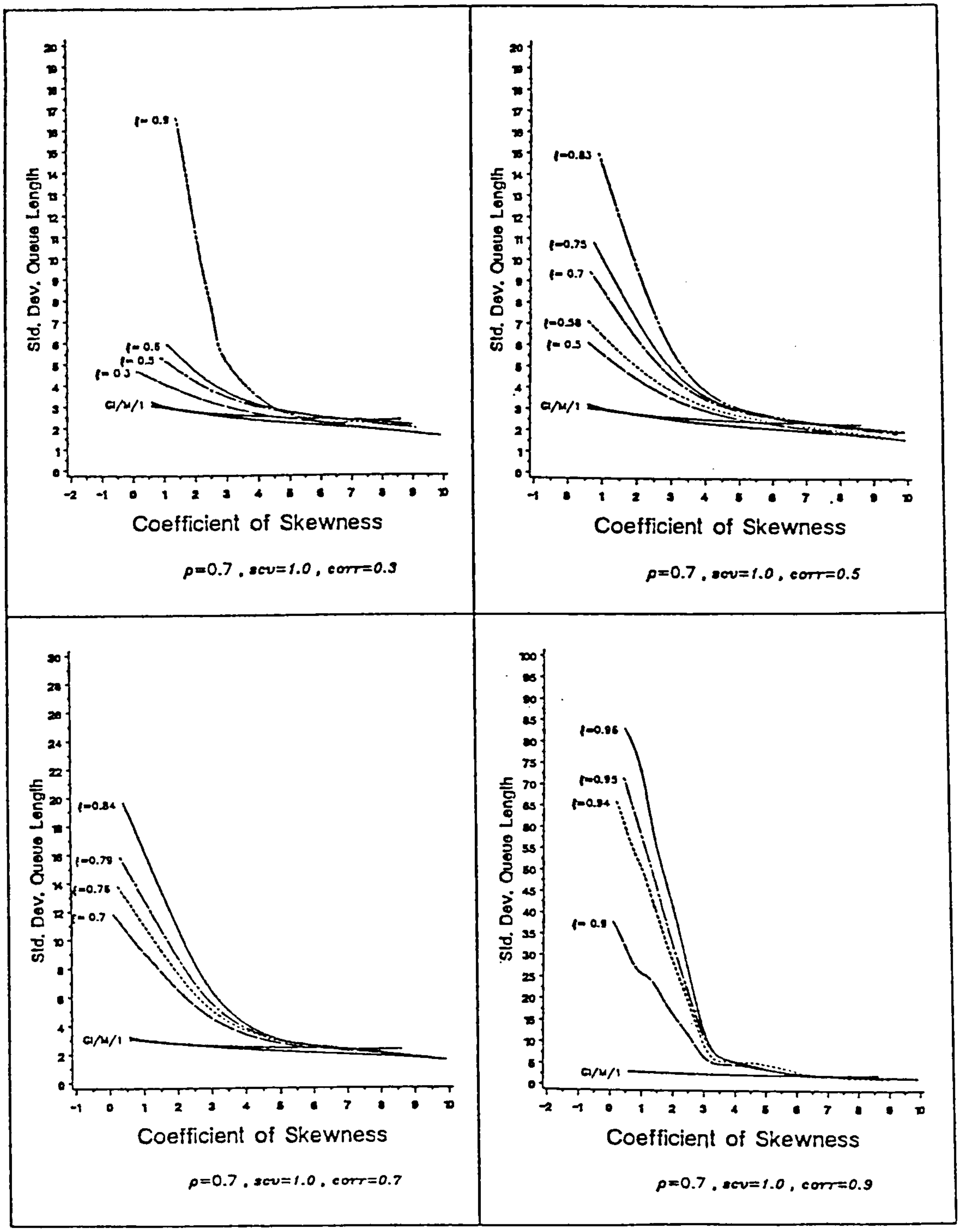

Figure 6.25 The standard deviation of the MR/M/1 queue: ρ=0.7, scv=1.0

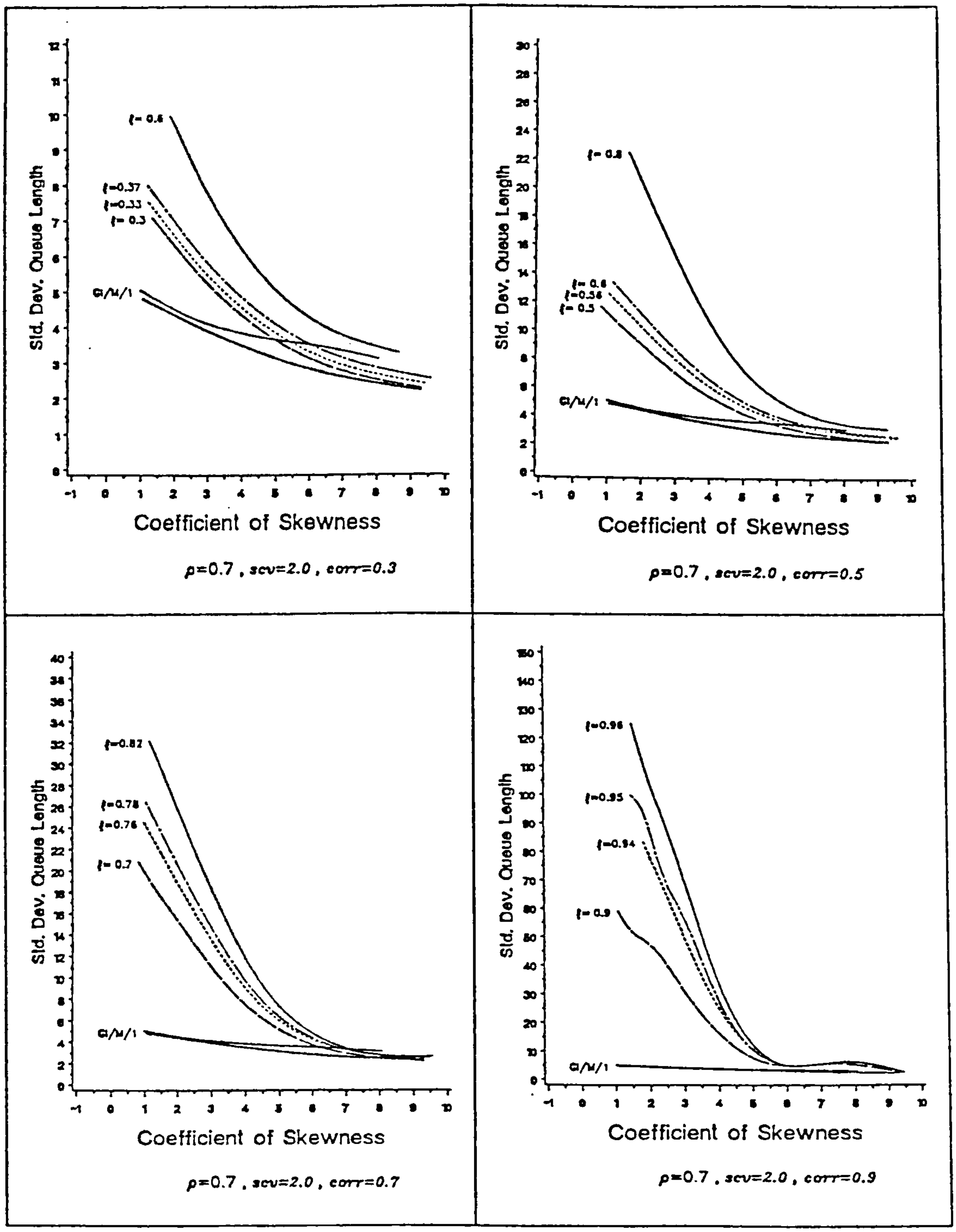

Figure 6.26 The standard deviation of the MR/M/1 queue: $\rho=0.7$, $scv=2.0$

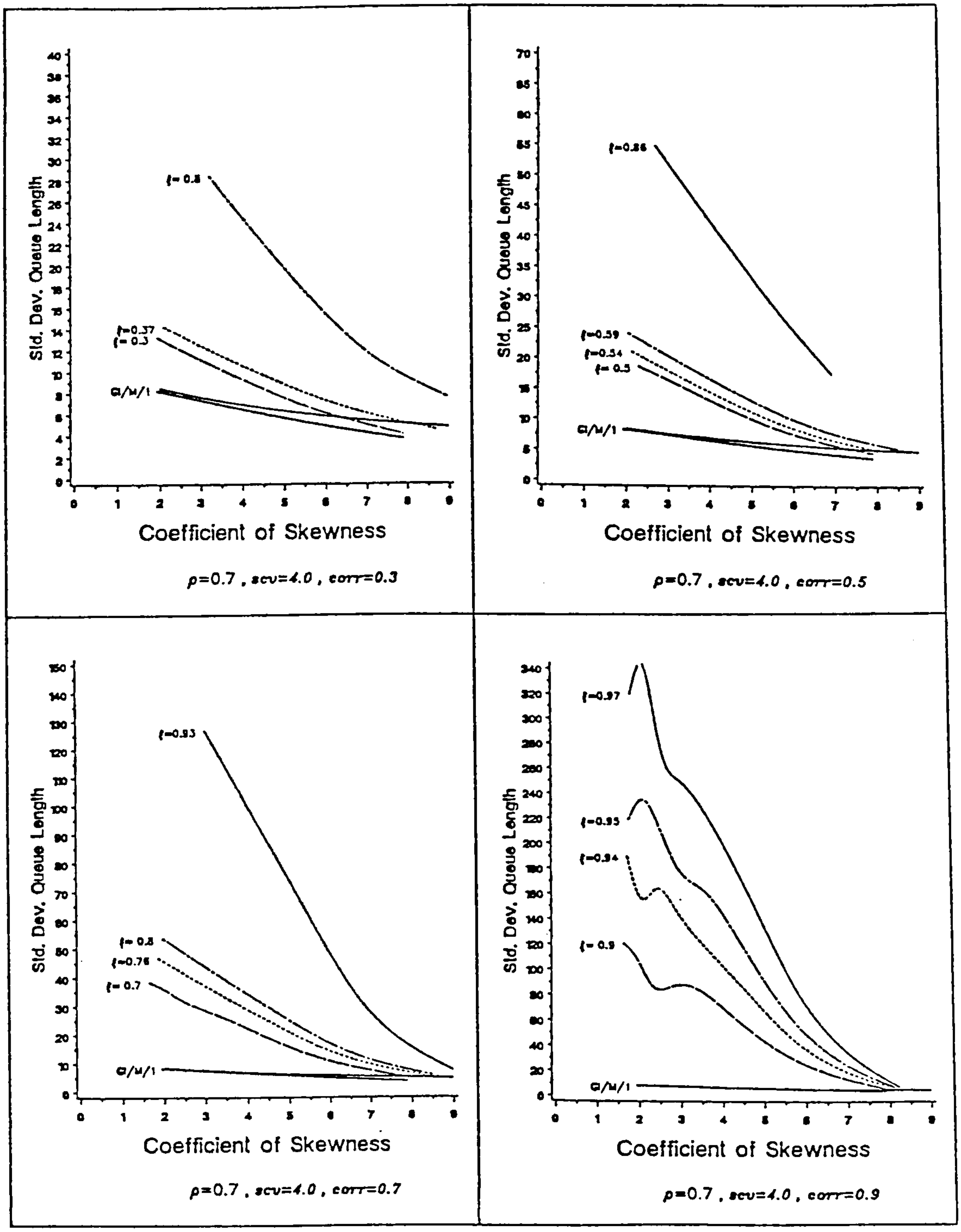

Figure 6.27 The standard deviation of the MR/M/1 queue: $\rho=0.7$, $scv=4.0$

Hence, all of the observations for the L^t also apply to the σ.

The graphs in Figure 6.24 through Figure 6.27 are given in order to illustrate the above observations. These graphs are for ρ=0.7 only. We did not produce the graphs for other traffic intensities because the graphs for ρ=0.7 are sufficient to illustrate what happens generally to the standard deviations σ of the MR/M/1 queues. The graphs are plotted using the same format as those in Figure 6.13 through Figure 6.16 with σ, the standard deviation of the queue length $N(t)$, on the y-axis (instead of the mean queue length).

Pair-wise comparisons of Figure 6.13 and Figure 6.24, Figure 6.14 and Figure 6.25, Figure 6.15 and Figure 6.26, Figure 6.16 and Figure 6.27 reveal the similarity of the shapes of the curves within each of the graphs. Hence, all of the observations for the L^t apply to the standard deviation σ.

Here is an example. Consider the 2 curves corresponds to ξ=0.8: one in Figure 6.15B for the L^t and the other in Figure 6.26B for the σ. Notice that he shape of these two curves are identical. Table 6.28 shows how the L^t increases from 2.6 to 15.0 and the σ increases from 3.1 to 22.5 as γ decreases from 9.29 to 1.62. We also calculate the coefficient of variations of the queue length ($\frac{\sigma}{L^t}$) for these cases and put these values in the last column of Table 6.28. The $\frac{\sigma}{L^t}$ increases from 1.19 to 1.50 for the same decreasing γ values. Notice also that $\frac{\sigma}{L^t}>1$.

Table 6.28. The Mean, the standard deviation and the coefficient of variation of the queue length for $\rho=0.7$, $scv=2.0$, $corr=0.5$ and $\xi=0.8$.

γ	L^t	σ	σ/L^t
9.29	2.6	3.1	1.19
5.92	4.1	5.3	1.29
4.42	6.7	9.0	1.34
3.54	9.1	12.7	1.40
2.95	11.0	15.5	1.41
2.51	12.4	17.8	1.44
2.15	13.5	19.6	1.45
1.87	14.3	21.2	1.48
1.62	15.0	22.5	1.50

4.7 Results for the Caudal Characteristic η.

The caudal characteristic η is a by-product of queues which have matrix-geometric stationary probability vector: η is the dominant eigenvalue of the rate matrix $\mathbf{R}$ (see Latouche (1985), Neuts (1986), Lemma 3.1.36 and Remark 5 in chapter 3).

The Uncorrelated GI/M/1 Case.

For the case where the Markov renewal process is equivalent to a renewal process (MR=GI), the $\mathbf{R}$ matrix of this degenerate MR/M/1 queue are all zero except one, which is, of course, the caudal characteristic η. Here, $\eta=z$, where z is the unique solution to the characteristic equation $z=A^*(\mu(1-z))$ (see Remark 1 in chapter 3). For the M/M/1 queue, $\eta=\rho$; i.e., the caudal characteristic is the traffic intensity. The caudal characteristic η is a better indicator of the congestion in the queue than the traffic intensity (see Remark 5, chapter 3).

In Figure 7.1−7.4, we plot the caudal characteristic using the same layout as the graphs in Figures 5−20. The 16 graphs for the caudal characteristic are only for $\rho=0.7$; they are sufficient to illustrate our observations.

Consider the GI/M/1 curves in Figure 7.1−7.4 (they correspond to $corr=0$, $\xi=0$). For $scv=0.5$, we would expect the MR/M/1 queue to behave like the Erlang queue with $\eta< \rho$. Figure 7.1 shows that this is the case for all the γ values on the graphs. However, for $scv=1.0$, 2.0 and 4.0, the η decreases as γ increases and the η curves crosses the $\eta=\rho=0.7$ line at around 2.0, 6.0 and 11.0, respectively. Recall that for $scv=1.0$ and $\gamma=2.0$ (see Figure 7.2), the GI/M/1 queue is approximately the M/M/1 queue and hence $\eta\sim0.7$ for these scv and γ values is expected.

For the case where $scv=2.0$ and 4.0, we have a rather surprising result. We would expect that a queue with $scv>1.0$ would behave like a H/M/1 queue (Hyperexponential arrivals) with $\eta>\rho$. What the GI/M/1 curves in Figures 7.3 and 7.4 show is that if the coefficient of skewness γ is large enough, it can reduce the caudal characteristic η and make it smaller than ρ (in a sense, the queue would behave like an $E_k/M/1$ queue instead of the H/M/1 queue). For example, for the GI/M/1 curve in Figure 7.3C ($scv=2.0$), $\eta=0.82$ for $\gamma=0.9$. As γ increases to 9.5, η decreases to 0.65. This η curve crosses the $\eta=0.7$ line at around $\gamma=6.0$. This observation reinforces the previous observation made in section 4.4 which shows that for the GI/M/1 queue, the

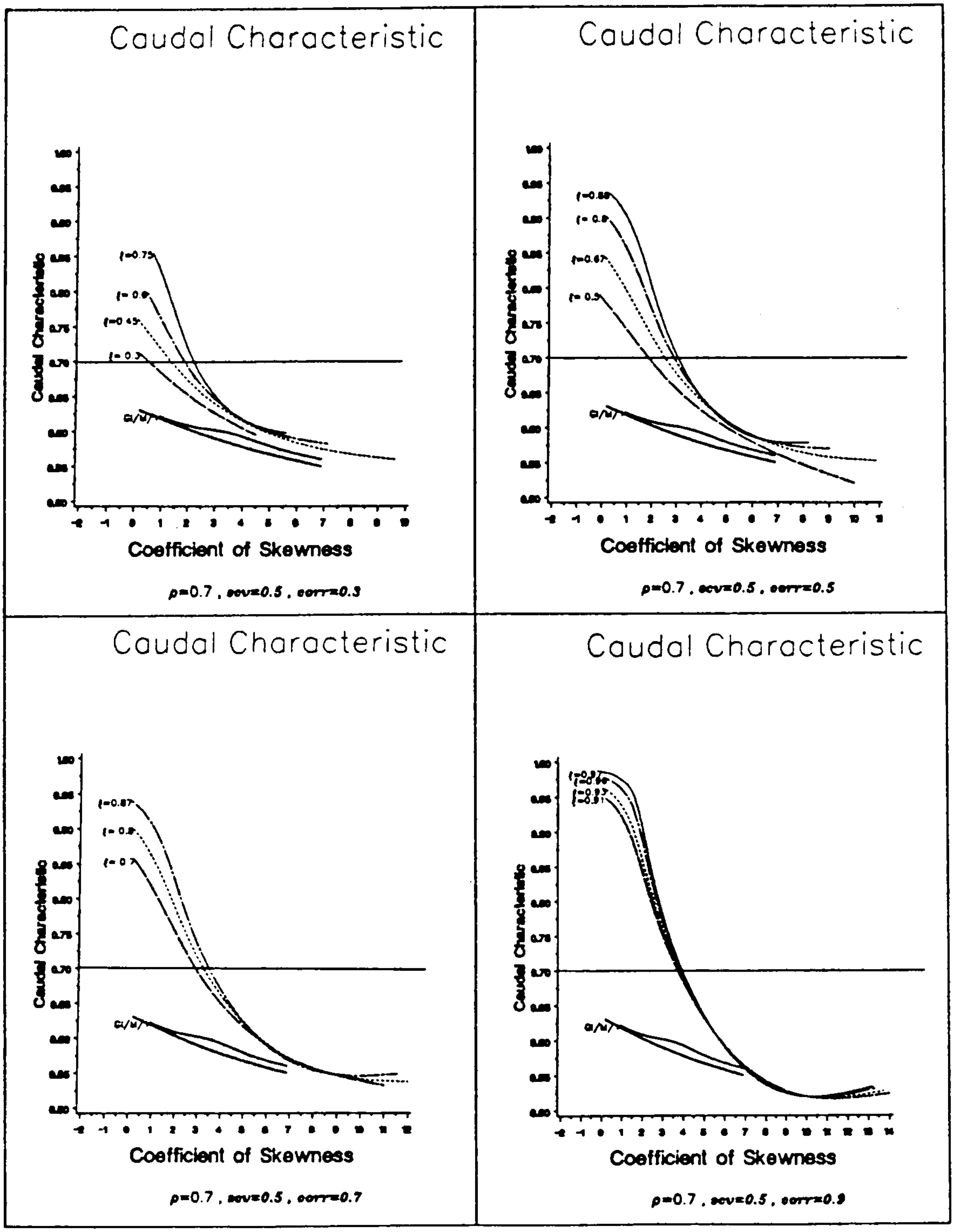

Figure 7.1 The caudal characteristic of the MR/M/1 queue: ρ=0.7, scv=0.5

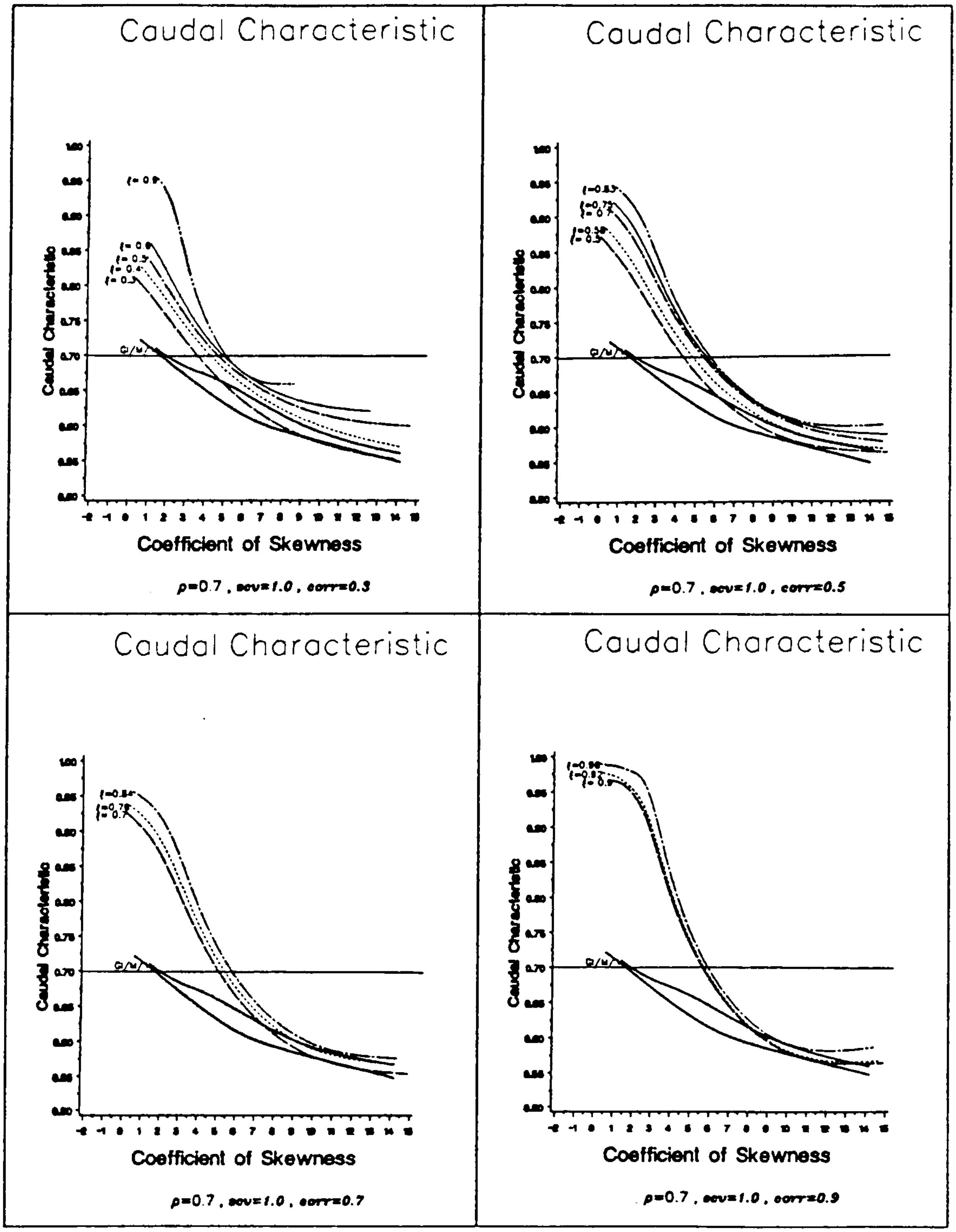

Figure 7.2 The caudal characteristic of the MR/M/1 queue: ρ=0.7, scv=1.0

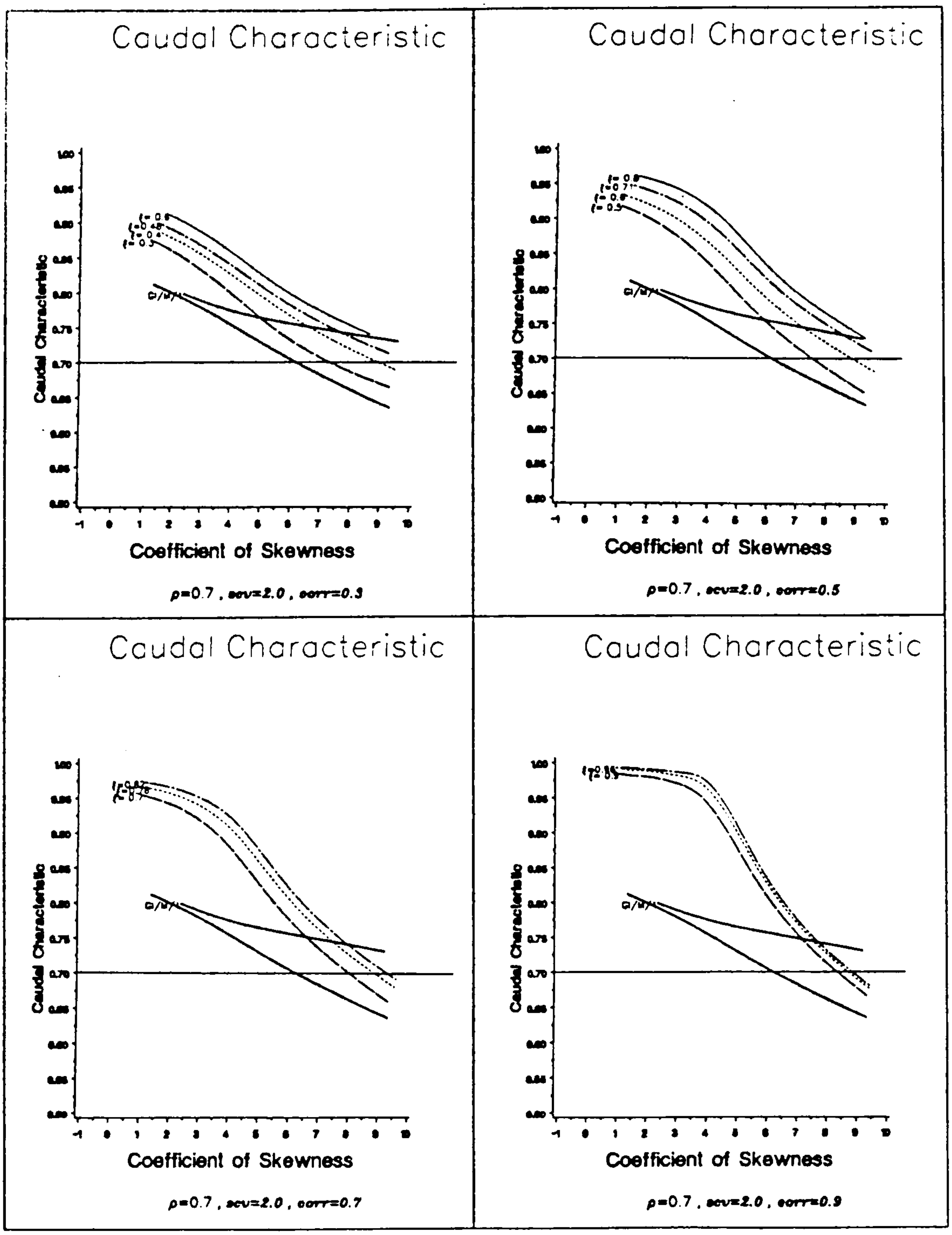

Figure 7.3 The caudal characteristic of the MR/M/1 queue: ρ=0.7, scv=2.0

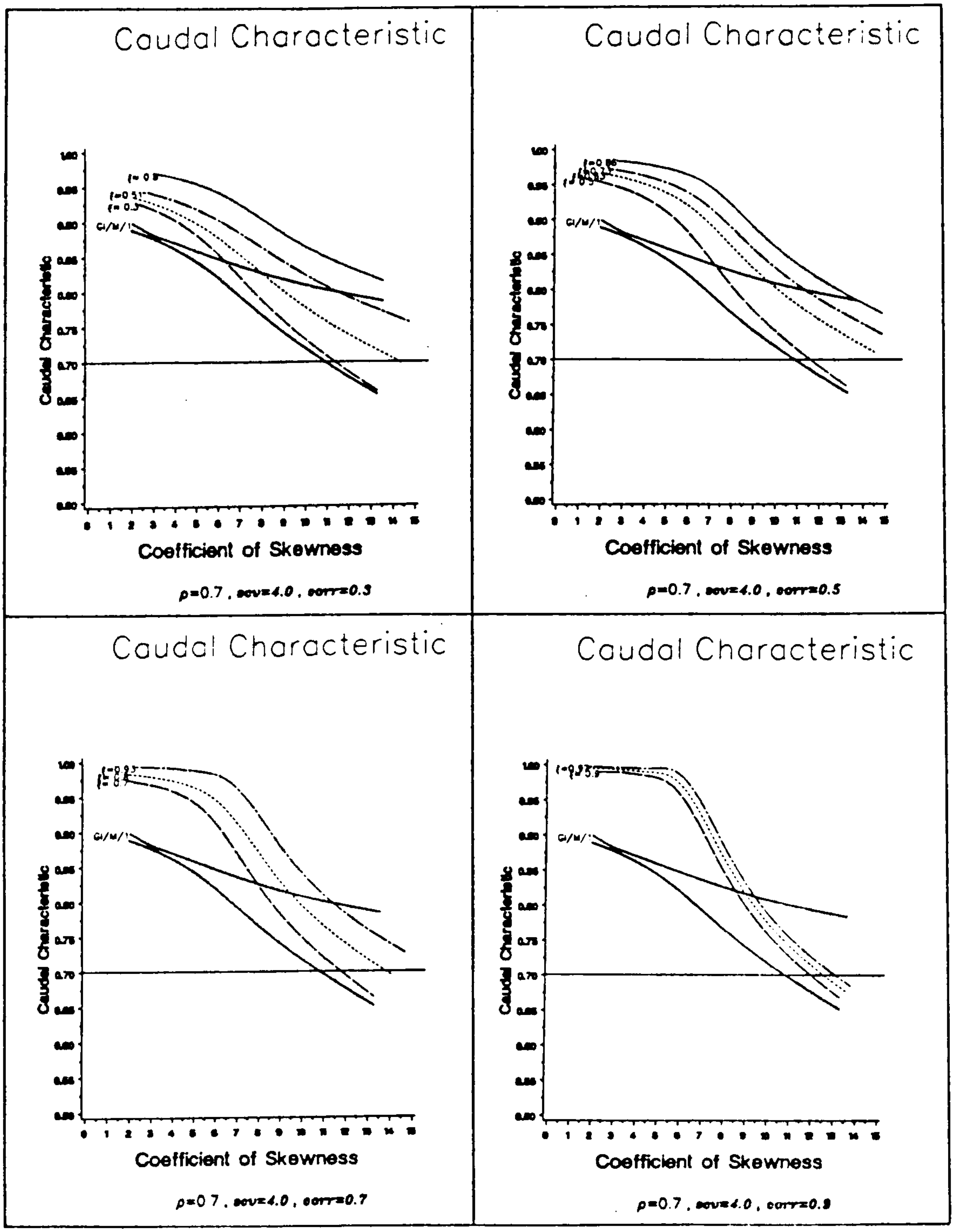

Figure 7.4 The caudal characteristic of the MR/M/1 queue: ρ=0.7, scv=4.0

γ plays an important role for η as it does for L^t and σ.

An obvious corrolary to Conjecture 4.6.1 is: as $\gamma \rightarrow \infty$, $\eta \rightarrow z$, where z is the unique solution in (0,1) to the characteristic equation (6.4) of the D/M/1 queue. This is so, because when MR=GI, $\eta = z$ and all of the eigenvalues of the matrix $\mathbf{R}$ are zero.

The Correlated MR/M/1 Case.

Let us look at the η for the correlated MR/M/1 case (notice that all of the curves in Figures 7.1−7.4 are for $corr \geq 0$ and $\xi \geq 0$). For $scv=0.5$ (Figure 7.1), η is not strictly less than ρ for values of γ. The low γ values combined with $corr>0$, $\xi>0$ (the positive serial correlation) can cause η to be larger than ρ. As γ increase, η decreases and at some γ value, η crosses the $\eta=0.7$ line. For example, in Figure 7.1B ($corr=0.5$), the η curve for $\xi=0.88$ decreases from 0.93 to 0.58 as γ increases from 0.83 to 8.2.

The following observations apply to all of the MR/M/1 queues: as γ increases; η decreases; as *corr* increases, η increases; as ξ increases, η increases. The specific values of γ at which η crosses the $\eta=0.7$ line for various values of *corr* and ξ can be obtained from the individual curve in Figure 7.1−7.5.

When $scv=1.0$ (Figure 7.2), the η values are larger than the η values for $scv=0.5$ (for the same values of γ, *corr* and ξ). This shows that as *scv* increases, η also increases. Notice again that the curves in Figure 7.2 decrease from left to right and at some γ value, they cross the $\eta=0.7$ line. For example, in Figure 7.2C ($corr=0.7$), the η curve for $\xi=0.84$ decreases from 0.96 to 0.58 as γ increases from 0.43 to 18.4.

Recall that for $scv=1.0$ and $\gamma=2.0$ we have the correlated M/M/1 queue. Here, $\eta>\rho$, which implies that the positive serial correlation in the arrival process causes η to be larger than ρ (ρ the caudal characteristic of the uncorrelated M/M/1 queue). We did not plot the η curves for the negative lag-1 correlation cases ($corr<0$, $\xi<0$). However, looking at the computer output we have for these cases, we found that $\eta<\rho$, which implies that the negative lag-1 correlation (the lags of the serial correlation alternate in signs) in the arrival process causes the η to be smaller than the traffic intensity ρ. This observation is consistent with some of the results of Latouche (1985).

When $scv>1.0$, $corr>0$ and $\xi>0$, we would expect that $\eta>\rho$. However, Figures 7.3 and 7.4 show that this is not the case for all values of γ. Again, the η curves cross the $\eta=0.7$ line at some γ values. For example, in Figure 7.4C ($scv=4.0$, corr=0.7), the

η curve for ξ =0.93 decreases from 0.96 to 0.73 as γ increases from 2.3 to 14.7. This observation shows that a large γ value can counteract the effect of *scv*, *corr* and ξ on the η and can cause the η to be smaller than the traffic intensity ρ.

4.8. The Anomalies.

In Figures 6.6D and 6.17–20, we observe that some of the L^tcurves are not monotonically decreasing. These wiggles are not caused by the spline interpolating routine use by SASGRAPH, rather the computed values for these L^t's show this non-monotonic behavior (see, for example Table 8.4).

At the end section 4.1, we commented that although ρ, *scv*, *corr*, ξ and γ seem to explain most (but not all) of the behavior of the L^t, there are many other parameters in the arrival process that were not included in this investigation. Could some of these unaccounted parameters cause the problem? We also mentioned that it could be that the serial correlation (which is a measure of linear association between random variables) used may not be appropriate for the nonlinear responses observed in our investigation.

In trying to explain these anomalies, we rechecked all of the mathematical derivations in chapters 2 and 3 to ensure there is no error there. We also reexamined the computer codes and made sure also they are correct. These codes have been tested extensively over the period of two years and hand calculations were performed to validate the codes for some special cases of the MR/M/1 queue.

Therefore, we have concluded that the higher moments of the arrival process are having effects here that our numerical methods cannot expose. To test this conjecture, we calculated the 4th moment (the coefficient of kurtosis), the 5th moment and a partial covariance of three consecutive interarrival times of the arrival process for the anomalous curve ρ=0.9, *scv*=2.0, *corr*=0.7 and ξ=0.7 (see the bottom curve of Figure 6.19C):

$$Kurt = \frac{E[\,(D_n - z_1)^4\,]}{\sqrt{z_2^4}} \quad (\textit{the coefficient of kurtosis}), \tag{8.1}$$

$$M_5 = \frac{E[\,(D_n^a - z_1)^5\,]}{\sqrt{z_2^5}} \quad (\textit{the fifth moment})\ , \tag{8.2}$$

$$pcov(1,2,3) = E[\,(D_n^a - z_1)\,(D_{n+1}^a - z_1)\,(D_{n+2}^a - z_1)\,]\ , \tag{8.3}$$

where $pcov(1,2,3)$ is the partial covariance among the three consecutive interarrival times). The results are tabulated in Table 8.4.

We observe that the values of *Kurt*, $pcov(1,2,3)$ and M_5 are monotonically increasing as the γ value increases. What these numerical values in Table 8.4 do not show is how each individual moment: *Kurt*, $pcov(1,2,3)$ and M_5 affects the L^t when all of the other parameters are held constant.

In section 4.6, we show that ρ, *scv*, *corr* and ξ have positive effect on the L^t. That is, if all of the others parameters are held constant, as ρ, *scv*, *corr* or ξ increases, the L^t also increases. On the other hand, the coefficient of skewness γ has negative effect on the L^t (the L^t decreases as γ increases when all of the other parameters are held constant. We do not know the individual effect of each of the other parameters in the arrival process on the L^t. Some of them could have positive effect on the L^t and some could have negative effect on the L^t; and hence the combined effects of all of these parameters on the L^t cause the L^t to be non-monotonic when the L^t is viewed as a function of γ alone.

The plausible explanations given above in trying to make sense the anomalous behavior of the L^t may not be completely satisfactory. The scenario used in our numerical investigation is rather limited: we only have 5 parameters (a, b, m_1, m_2 ,k) in the semi-Markov kernel to characterize the Markov renewal arrival process. But in the absence of analytic results, this numerical investigation has shed light on the properties of queues with dependent arrivals that were not known prior to this investigation.

Table 8.4. The Anomalous curve: $\rho=0.9$, scv=2.0, corr=0.7, $\xi=0.7$.

γ	$Kurt$	$pcov(1,2,3)$	M_5	L^t
0.82	1.69	1.13	2.25	50.15
1.07	2.16	1.47	3.42	39.22
1.36	2.87	1.88	5.33	54.41
1.72	3.99	2.38	8.68	52.50
2.19	5.82	3.02	15.06	41.15
2.84	9.06	3.91	28.68	48.87
3.80	15.50	5.26	63.12	46.00
5.49	31.22	7.59	177.85	40.66
9.33	88.32	12.92	837.88	28.90
28.26	801.86	39.15	22801	8.68

Chapter 5

SUMMARY, CONCLUSIONS AND EXTENSIONS

5.1 Summary.

The thesis of this paper is to investigate how the dependency in the arrival process affects the queueing performance measures. We assume that the queue is a single server queue with exponential service time. For the arrival process, we chose the Markov renewal arrival process (MRAP). This choice was made because many of the typical arrival processes can be obtained as special cases of the MRAP. But the main reason behind this choice is that the interarrival times of the MRAP are dependent. To reduce the number of parameters involved in the arrival process, we chose a 2-state MRAP with the semi-Markov kernel $\mathbf{A}(t)$ given by (2.3.7−8). The moments and the serial correlation of this MRAP were derived in chapter 2.

Given the mean (z_1), the variance (z_2), and the lag-1 covariance (z_3), of the 2-state MRAP, the parameters, a, b, m_1, m_2 *and* k, of the semi-Markov kernel $\mathbf{A}(t)$ are found using Algorithm 2.4.34. This allows us to vary, systematically, the traffic intensity ($\rho=\frac{1}{\mu . z_1}$), the squared coefficient of variation ($scv=\frac{z_2}{z_1^2}$) and the lag-1 correlation ($corr=\frac{z_3}{z_2}$) , and compute the corresponding a, b, m_1, m_2 and k. Then for each set of parameters: a, b, m_1, m_2, k, we compute the mean queue length at arbitrary times (L^t), the standard deviation (σ) of the queue length at arbitrary times and the caudal characteristic η of the MR/M/1 queue (see (3.3.41), (3.3.43), (3.3.44) and Lemma 3.1.36). The other performance measures of the MR/M/1 queue (e.g., the mean waiting time and the mean sojourn time) are related to the L^t through the formulas given in section 3.4. The performance measures of the MR/M/1 queue and the algorithm to compute them were given in chapter 3.

The investigation of the effect of the dependency in the arrival process on the L^t and the σ of the MR/M/1 queue was carried out in chapter 4. The L^t and the σ were found numerically because no closed form solutions of L^t and σ exist. We summarize the conclusions from chapter 4 in the next section and in section 5.3 we mention the future extensions to this work.

5.2 Conclusions.

There are 5 parameters of the arrival process used in this investigation: ρ, *scv*, *corr*, ξ and γ (see (4.1.2)–(4.1.7) for the definitions of these parameters). Thus, to see how the L^t of the MR/M/1 varies with respect to these 5 parameters, one needs to draw a six-dimensional surface which , of course, is impossible. So, we cut *slices* through this six-dimensional surface and plot these slices as two-dimensional graphs in all of the graphs in chapter 4. Each of the curves in these graphs corresponds to fixed values of ρ, *scv*, *corr*, and ξ.

The Mean Queue Length at Arbitrary Times (L^t).

In section 4.4, we show that for fixed values of ρ, *scv* and γ, the range of possible values of the L^t for the (uncorrelated) GI/M/1 queue is small (see Figure 4.4.1). When the lag-1 correlation of the arrival process is negative (this means that the lags of the serial correlation alternate in signs), the L^t of the MR/M/1 queue is smaller (but not by much) than the L^t of the GI/M/1 queue (section 4.5). Therefore, in section 4.6, we focus our attention to the MR/M/1 queue with positive serial correlation.

We find that the coefficient of skewness of the arrival process (γ) plays a very important role. All of the curves in Figures 4.6.5–20 decreases rapidly as γ increases and after certain values of γ called *the turning region* (see section 4.6.2), the L^t curves flatten. This important observation indicates that to the left of the turning region, the L^t is *almost insensitive* to the dependency in the arrival process.

However, to the right of the turning region, the L^t is sensitive to the positive serial correlation in the arrival process. Highly correlated arrival process (large *corr* and ξ) can cause the L^t to be significantly larger than the L^t for the uncorrelated queue. In section 4.6.3, we cite an example where L^t=220.5 for *corr*=0.7 and ξ=0.97; where as the corresponding uncorrelated L^t is only 6.4.

The Standard Deviation of the Queue Length (σ).

For the MR/M/1 queue, the magnitude of the standard deviation σ is larger than the corresponding L^t(see Figures 4.6.24–27). However, the shapes of the σ curves are similar to those of the L^t curves. So, all of the conclusions drawn for the L^t also apply to the standard deviation σ.

We can find the turning regions of the σ curves from Figures 4.6.24–27. To the right of the turning region, the σ is *almost insensitive* to the dependency in the arrival process; to the left of the turning region, the σ is sensitive to the positive serial correlation in the arrival process, analogous to the L^t curves.

The Caudal Characteristic of the MR/M/1 queue.

For the M/M/1 queue, the caudal characteristic η equals to the traffic intensity ρ ($\eta=\rho$). For the uncorrelated GI/M/1 queue, one would expect that when $scv<1.0$, $\eta<\rho$ (i.e., the GI/M/1 queue would behave like an E_k/M/1 queue) and when $scv>1.0$, $\eta>\rho$ (i.e., the queue would behave like a H/M/1 queue).

Our observation of Figures 4.7.1−4 indicates that this is not necessarily true. We found again that the coefficient of skewness (γ) plays an important role. For the uncorrelated GI/M/1 queue with $scv>1.0$, η can be smaller than ρ when γ is large enough.

For the correlated MR/M/1 queue, even for $scv<1.0$, a low γ value combined with the positive serial correlation can cause η to be larger than ρ. On the other hand, $scv>1.0$ does not necessarily results in $\eta>\rho$. A large value of γ can cause η to be smaller than ρ, even for the queue with highly correlated interarrival times.

In conclusion, the coefficient of skewness of the interarrival time plays an important role in determining the magnitude of the performance measures of queues with correlated arrival process. When the coefficient of skewness is large (to the right of the turning region), the queueing performance measures are almost insensitive to the dependency in the arrival process. However, when the coefficient of skewness is small (to the left of the turning region) the queueing performance measures are sensitive to the positive serial correlation in the arrival process.

5.3 Extensions.

In this dissertation, we only look at the single server queue. An immediate extension is to allow multiple servers, the MR/M/c queue, where c is the number of servers in the queue. Here, only one extra parameter is introduced. The semi-Markov kernel of the joint queue length and type of the customers can be easily obtained by following the reasoning in Gross and Harris section 5.3.2. The kernel is the matrix-vector analog of the scalar version GI/M/c queue.

We can replace the 2-state MRAP with the 3-state MRAP. In doing so, more parameters are introduced into the arrival process and hence it is more difficult to come up with an algorithm similar to Algorithm 2.4.34 which enabled us to vary systematically the mean, the variance and the covariance of the arrival process.

In the 2-state MRAP, one cannot obtain a strictly negative serial correlation where all lags of the serial correlation are negative. If an arrival process with strictly negative serial correlation is found, it will be interesting to see how this serial correlation affects the queueing performance measures.

For the finite capacity MR/M/1/L queue (with L−1 waiting room capacity), the stationary queue length probability vector can be found numerically by solving the system of linear equations

$$\mathbf{x}.\mathbf{P} = \mathbf{x}\ , \quad \mathbf{x}.\mathbf{e} = 1, \tag{3.1}$$

where $\mathbf{P}$ is the transition matrix of the joint queue length and the customer's type,

$$\mathbf{P} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{A}_0 & & & \\ \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{A}_0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \mathbf{B}_{L-1} & \mathbf{A}_{L-1} & \mathbf{A}_{L-2} & \cdots & \mathbf{A}_0 \\ \mathbf{B}_{L-1} & \mathbf{A}_{L-1} & \mathbf{A}_{L-2} & \cdots & \mathbf{A}_0 \end{bmatrix}, \tag{3.2}$$

$\mathbf{A}_n$ and $\mathbf{B}_n$ for n=0, 1, 2, ..., L-1, are given by (3.1.21) and (3.1.22).

When the queue is full, an arriving customer cannot enter the queue and it *overflows*. Here, one can investigate on how the dependency in the arrival process affects the overflow process. The overflow for the GI/M/1 queue was given by Çinlar and Disney (1967).

Another interesting queueing model is the $MR/M^j/1$ queue where the m types of customers arrive according to the m-state Markov renewal process and a type j customer demands an exponential service time with rate μ_j. One difficulty encountered in the $MR/M^j/1$ queue is that in order to maintain the Markov renewal property of the semi-Markov kernel, one must use a queue length vector that not only keeps track of the number of customers in the system but also keeps track of the type and position of each customer in the queue. Therefore, the state space for this queue can be prohibitly large.

Finally, Disney and Kiessler (1987) showed that in the Markovian queueing network, the traffic processes on the arcs of the network are Markov renewal processes. If one can find a way to estimate the first 3 moments and the serial correlation of the input to each node in the network, then each node may be treated as if it is an MR/M/c queue. This may be a better approximation than that which treats each node as if it is a queue with renewal arrivals (see Whitt (1983a–b)). However, one must realize that the input to each of the node in the network is not a Markov renewal process, rather it is a superposition of Markov renewal processes.

REFERENCES

Albin, S. L. (1984), "Approximating a Point Process by a Renewal Process, II: Superposition Arrival Processes to Queues", *Oper. Res.,* **32**, 1133-1162.

Altiok, T. (1985), "On the Phase-Type Approximations of General Distributions", *IIE Transactions,* **17**, 110–116.

Arjas, E. (1972a), "On a Fundamental Identity in the Theory of Semi-Markov Processes", *J. Appl. Probab.,* **4**, 258-270.

Arjas, E. (1972b), "On the Use of a Fundamental Identity in the Theory of Semi-Markov Queues, *J. Appl. Probab.,* **4**, 271-284.

Baily, D. E. and Neuts, M. F. (1981), "Algorithmic Methods for Multi-server Queues with Group Arrivals and Exponential Services", *European J. Oper. Res.,* **8**, 184-196.

Burke, P. J. (1972), "Output Processes and Tandem Queues", *in J. Fox (ed.), Proceedings of the Symposium on Computer-Communications Networks and Teletraffic,* Polytechnic Press of the Polytechnic Insitute of Brooklyn, John Wiley and Sons, Inc., New York, 419-429.

Çinlar, E. (1965), "Analysis of Systems of Queues in Parallel", *Doctoral Dissertation,* University of Michigan.

Çinlar, E. (1967), "Queues with Semi-Markovian Arrivals", *J. Appl. Probab.,* **4**, 365-379.

Çinlar, E. (1969), "Markov Renewal Theory", *Adv. in Appl. Probab.,* **1**, 123-187.

Çinlar, E. (1975a), "Markov Renewal Theory: a Survey", *Management Sci.,* **21**, 727-752.

Çinlar, E. (1975b), *Introduction to Stochastic Processes,* Prentice Hall, Inc., Englewood Cliff, N.J. .

Çinlar, E. and Disney, R.L. (1967), "Streams of Overflows from a Finite Queue", *Oper. Res.,* **15**, 131-134.

De Smit, J. H. A. (1985), "The Single Server Semi-Markov Queue", *Memorandum nr. 517, Department of Applied Mathematics,* Twente University of Technology, Enschede, The Nederlands.

Disney, R. L. and de Morais, P. R. (1976), "Covariance Properties for the Departure Process of M/Ek/1/N Queues", *AIIE Trans.,* **8**, 169-175.

Disney, R. L. and Kiessler, P. C. (1987), *Traffic Processes in Queueing Networks: a Markov Renewal Approach,* Johns Hopkins Univ. Press, Baltimore.

Finch, P. D. (1963), "The Single Server Queueing System with Non-recurrent Input Process and Erlang Service Time", *J. Austral. Math. Soc.,* **3,** 220-236.

Finch, P. D. and Pearce, C. (1965), "A Second Look at a Queueing System with Moving Average Input Process", *J. Austral. Math. Soc.,* **5,** 100-106.

Franken, P., König, D., Arndt, U., and Schmidt, V. (1982), *Queues and Point Processes,* John Wiley and Sons Inc., New York.

Gopinath, B. and Morrison, J. A. (1977), "Single Server Queues with Correlated Inputs", *Computer Performance,* K.M. Chandy and M. Reiser eds., North-Holland Publ. Co., pp.264-277.

Gross, D. (1975), "Sensitivity of Queueing Models to the Assumption of Exponentiality", *Nav. Res. Logist. Quart.,* **22,** 271-287.

Gross, D. and Harris, C. M. (1985), *Fundamentals of Queueing Theory,* 2nd edition, John Wiley and Sons Inc., New York.

Heyman, D. P. and Sobel, M. J. (1982), *Stochastic Models in Operations Research,* Vol. 1, McGraw-Hill, New York.

Jackson, J. R. (1963), "Networks of Waiting Lines", *Oper. Res.,* **2,** 518-521.Jacobs, P. A. (1979), "Heavy Traffic Results for Single Server Queues with Dependent (EARMA) Service and Interarrival Times", *Naval Postgraduate School Tech. Report* NPS55-79-009.

King, R. A. (1971), "The Covariance Structure of the Departure Process from M/G/1 Queues with Finite Waiting Lines", *J. Roy. Statist. Soc. Ser. B,* **33,** 401-405.

Kleinrock, L. (1975), *Queueing Systems Vol. I: Theory,* John Wiley & Sons Inc., New York.

Klincewicz, J. G. and Whitt, W. (1984), "On Approximations for Queues, II: Shapes Constraints ", *AT&T Bell Lab. Tech. J.,* **63,** 139-161.

Latouche, G. (1981), "On a Markovian Queue with Weakly Correlated Interarrival Times", *J. Appl. Probab.,* **18,** 190-223.

Latouche, G., Gaver, D. P. and Jacobs, P. A. (1982), "Finite Markov Chain Models Skip Free in One Direction", *Naval Postgraduate School* Tech. Report NPS55-82-015.

Latouche, G. (1985), "An Exponential Semi-Markov Process with Applications to Queueing Theory", *Stochastic Models,* **1,** 137-170.

Loynes, R. M. (1962a), "The Stability of a Queue with Non-independent Interarrival and Service Times", *Proc. Cambridge Philos. Soc.* **58**, 497-520.

Loynes, R. M. (1962b), "Stationary Waiting Time Distribution for Single Server Queues", *Ann. Math. Statist.*, **33**, 1323-1339.

McNickle, D. C. (1974), "The Number of Departures from a Semi-Markov Queue", *J. Appl. Probab.*, **11**, 825-828.

Morse, Ph. M. (1958), *Queues, Inventories and Maintenance,* John Wiley and Sons Inc., New York.

Neuts, M. F. (1978a), "Markov Cchains with Applications in Queueing Theory, Which Have a Matrix-Geometric Invariant Probability Vector", *Adv. in Appl. Probab.*, **10**, 185-212.

Neuts, M. F. (1981), *Matrix Geometric Solutions in Stochastic Models,* Johns Hopkins Univ. Press, Baltimore.

Neuts, M. F. (1986), "The Caudal Characteristic Curve of Queues", *Adv. in Appl. Probab.*, **18**, 221-254.

Neuts, M. F. and Chakravarthy, S. (1981), "A Single Server Queue with Platooned Arrivals and Phase Type Services", *European J. Oper. Res.*, **8**, 379-389.

Pearce, C. (1966), "A Queueing System with General Order Moving Avarage Input and Negative Exponential Service Time", *J. Austral. Math. Soc.*, **6**, 223-236.

Pearce, C. (1968), "An Imbedded Chain Approach to a Queue with Moving Average Input", *J. Oper. Res. Soc.*, **16**, 1117-1130.

Ramaswami, V. (1980), "The Busy Period of Queues Which Have a Matrix-geometric Steady State Probability Vector", *Tech. Report no. 3-80,* Dept. of Math. Sciences, Drexel Univ., Philadelphia, Pa.

Runnenburg, J. Th. (1961), "An Example Illustrating the Possibilities of Renewal Theory and Waiting Line Theory for Markov Dependent Arrival Intervals", *Nederl. Akad. Wetensch. Pro. Ser. A,* **64**, 560-576.

Runnenburg, J. Th. (1962), "Some Numerical Results on Waiting Time Distributions for Dependent Arrival-intervals", *Statist. Neerlandica,* **16**, 337-346.

Shanthikumar, J. G. and Buzacott, J. A. (1980), "On the Approximations to the Single Server Queue", *Int. J. Prod. Res.*, **18**, 761-773.

Simon, B. and Disney, R. L. (1984), "Markov Renewal Processes and Renewal Processes: Some Conditions for Equivalence", *J. New Zealand Oper. Res. Soc.*, **12**, 19-29.

Tin, P. (1985), "A Queueing System With Markov Dependent Arrivals", *J. Appl. Probab.*, **22**, 668-677.

Whitt, W. (1983a), "The Queueing Network Analyzer", *Bell Syst. Tech. J.*, **62**, 2779-2815.

Whitt, W. (1983b), "Performance of the Queueing Network Analyzer", *Bell Syst. Tech. J.*, **62**, 2817-2843.

APPENDIX A

PROGRAM LISTINGS

```
C ******************** MAIN PROGRAM ************************
C                                                          *
C THIS PROGRAM FIGURES OUT THE FIVE PARAMETERS IN THE      *
C MARKOV RENEWAL ARRIVAL PROCESS, GIVEN THE                *
C MEAN (Z1), VARIANCE (Z2), AND THE LAG-1 COVARIANCE (Z3). *
C IT ALSO COMPUTES THE COEFFICIENTS OF SKEWNESS AND        *
C KURTOSIS OF THE INTERARRIVAL TIMES.                      *
C CV2 = SQUARE COEFFICIENT OF VARIATION                    *
C CORR = CORRELATION COEFFICIENT                           *
C                                                          *
C ***********************************************************
C
C LAST REVISION: APRIL 12, 1988
C
      REAL M1, M2, MU, KURT
C
 50   CONTINUE
C
      READ (7,*) Z1, RHO, CV2, CORR
C
      IF ( Z1 .LE. 0.0 ) GO TO 900
C
      MU = 1.0 / (RHO * Z1)
      Z2 = CV2 * Z1 * Z1
      Z3 = CORR * Z2
C
      SZ2 = SQRT (Z2)
      CV = SQRT (CV2)
C
      WRITE (8,26) Z1,Z2,Z3,MU,RHO
 26   FORMAT( '1',/,40X,'MAP OF MR/M/1 QUEUEING CHARACTERISTICS',
     &   /,40X,'======================================',
     &   //,20X,'Z1 =',F10.4,3X,'Z2 = ',F10.4,3X,'Z3 =',
     &   F10.4,3X,'MU =',F10.4,3X,'RHO =',F9.4,
     &   //,5X,'K',5X,
     &   'CV*2',6X,'CORR',7X,'PSI',6X,'SKEW',6X,
     &   'KURT',5X,'E[NT]',5X,'STDEVT',4X,'E[NA]',
     &   5X,'STDEVA',5X,'ETA',5X,'P(BUSY)',/)
```

```
C
C     *************************** FIX K ***********************
C
C
      YY = ABS (Z3)
      KK = (YY + Z1*Z1) / (Z2 - YY) + 1.0
      KK5 = KK + 25
        DO 100 K = KK , KK5, 5
          K1 = K
          IF (K1 .EQ. KK5) K1 = 500
          Y = ( K1*Z2 - Z1*Z1) / (K1 + 1.0)
C
          IF (Z3 .EQ. 0.0) THEN
            PSI = 0.0
          ELSE
            PSI = Z3 /Y
          ENDIF
C
C       ***** NEGATIVE CORRELATION CASE ***************
C
        IF (Z3 .LT. 0.0 ) THEN
          ZZ3 = SQRT(-Z3)
          D1 = Z1 + ZZ3 / PSI
          IF ( D1 .LE. 0.0 ) D1 = 0.0
          D2 = Z1 - ZZ3
          IF ( D2 .LE. 0.0 ) D2 = 0.0
          D = D2 - D1
          IF ( D .LE. 0.0 ) THEN
      WRITE (8,36) K1
 36   FORMAT( /,10X,'NO FEASIBLE VALUES FOR M2 FOR K =',
     &            I6,' ..........................')
            GO TO 100
          ENDIF
          H1 = D / 10.0
          M2 = D1 - 0.5 * H1
        ENDIF
C
        IF ( Z3 .GE. 0.0 ) THEN
          H1 = Z1 / 10.0
          M2 = -0.5 * H1
        ENDIF
        IF ( ABS(PSI) .GT. 0.99 ) GO TO 100
```

```
C
C ******** FIX M2 ****************************************
C
      DO 90 I = 1 , 10
         M2 = M2 + H1
C
         IF ( Z3 .LT. 0.0 ) THEN
           IF ( M2 .GT. D2 .OR. M2 .LT. D1 ) GO TO 90
         ENDIF
C
         M1 = ( K1 * (Z2 - Y) - M2 * Z1) / (Z1 - M2)
         PI1 = (Z1 - M2) / (M1 - M2)
         PI2 = 1.0 - PI1
C
         A = PI1 + PI2 * PSI
         B = PI2 + PI1 * PSI
C
C  NOW COMPUTE THE SHAPE CHARACTERISTICS OF THE MARGINAL
C  INTERARRIVAL TIME DISTRIBUTION.....................
C
         E3 = PI1 * M1*M1*M1 + PI2 * M2*M2*M2
         E3 = (K1+1.0) * (K1+2.0) * E3 / (K1 * K1)
         Z4 = E3 - 3.0 * Z1 * Z2 - Z1*Z1*Z1
         SKEW = Z4 / (SZ2*SZ2*SZ2)
C
         E4 = PI1 * M1*M1*M1*M1 + PI2 * M2*M2*M2*M2
         E4 = (K1+1.0)*(K1+2.0)*(K1+3.0) * E4 / (K1*K1*K1)
         Z5 = E4 - 4.0*Z1*E3 + 6.0*Z1*Z1*Z2 + 3.0*Z1*Z1*Z1*Z1
         KURT = Z5 / (Z2*Z2)
C
C        COMPUTE THE CORRESPONDING QUEUEING PROPERTIES .....
C
         CALL INFACE ( K1,A,B,M1,M2,PI1,PI2,MU,ENA,STDEVA,
     &                 ETA, PIDLE, ENT,STDEVT)
         PBUSY = 1.0 - PIDLE
C
         WRITE(8,60) K1,CV2,CORR,PSI,SKEW,KURT,
     &                ENT,STDEVT,ENA,STDEVA,ETA,PBUSY
 60      FORMAT ( I6,2F10.4, F10.2, 8F10.4 )
 90   CONTINUE
 100 CONTINUE
C
   GO TO 50
 900 STOP
   END
```

```
      SUBROUTINE INFACE (KX,AX,BX,M1,M2,PI1,PI2,MU,
     &                   ENA,STDEVA,ETA,PIDLE,ENT,STDEVT)

C ***********************************************************************
C THIS PROGRAM ACCEPT THE PARAMETERS FROM 'MRMAP',                      *
C AND MAKE CALLS TO MRM1, NEUTS, AND PRPROB TO COMPUTE                  *
C THE QUEUEING CHARACTERISTICS OF THE MR/M/1 QUEUE .                    *
C ***********************************************************************
C     LAST REVISION : MARCH 3 , 1988
C
      REAL M1, M2, MU
      INTEGER K(5,5), L(5,5)
      REAL P(5,5),LAMDA(5,5), ALPHA(5,5),A(5,5), PI(5)
      REAL AA(5,5,300),BB(5,5,300),Q1(5,5),R(5,5)
      REAL XX(300,5), TX(300)
C
      N=2
      DO 20 J = 1 , N
        DO 30 I = 1 , N
          K(I,J) = KX
          L(I,J) = 1
          P(I,J) = 1.0
 30     CONTINUE
 20   CONTINUE
C
      LAMDA(1,1) = 1.0 / M1
      LAMDA(1,2) = 1.0 / M2
      LAMDA(2,1) = LAMDA(1,1)
      LAMDA(2,2) = LAMDA(1,2)
      ALPHA(1,1) = 1.0
      ALPHA(1,2) = 1.0
      ALPHA(2,1) = 1.0
      ALPHA(2,2) = 1.0
C
      A(1,1) = AX
      A(1,2) = 1.0 - AX
      A(2,1) = 1.0 - BX
      A(2,2) = BX
      PI(1) = PI1
      PI(2) = PI2
      CALL  MRM1 (N,A,K,L,P,LAMDA,ALPHA,MU,PI,ITRUNC,AA,BB,Q1)
      CALL  NEUTS (N,MU,AA,BB,ITRUNC,PI,Q1,RHO,ETA,R,
     &         ENA,STDEVA,PIDLE,ENT,STDEVT)
      RETURN
      END
      SUBROUTINE MRM1 (N, A, K, L, P, LAMDA, ALPHA, MU, PI,
     &                 ITRUNC, AA, BB, Q1 )

C ***********************************************************************
C                                                                       *
C  THIS SUBROUTINE GENERATES THE BLOCK MATRICES IN                      *
C  THE ONE STEP PROBABILITY TRANSITION MATRIX OF AN                     *
C  N-TYPE MR/M/1 QUEUEING SYSTEM                                        *
```

```
C                                                                         *
C ************************************************************************
C
C   LAST UPDATE: JANUARY 23, 1988
C   IMSL VERSION 10.0 IS USED IN THIS PROGRAM
C
C ************************************************************************
C K,L    = ERLANG STAGES (THE CONDITIONAL INTERARRIVAL DISTRIB.  *
C          IS MIXED ERLANG-K AND ERLANG-L .                      *
C LAMDA = THE MATRIX OF ERLANG PARAMETER: ERLANG(LAMBDA,K)*
C ALPHA= THE MATRIX OF ERLANG PARAMETER: ERLANG(ALPHA,L).   *
C P      = PROB. WEIGHTS OF THE MIXED ERLANGS.                *
C N      = NUMBER OF TYPES OF ARRIVALS                        *
C A      =  TRANSITION MATRIX OF THE UNDERLYING MARKOV CHAIN IN*
C          THE ARRIVAL PROCESS.                               *
C Q1     = THE MATRIX OF THE MEANS OF ERLANG(LAMBDA,K).        *
C MU     = THE SERVICE RATE.                                  *
C RHO  = TRAFFIC INTENSITY.                                   *
C AA,BB = THE BLOCK MATRICES IN THE TRANSITION MATRIX .       *
C ITRUNC = THE TRUNCATION INDEX FOR THE TRANSITION MATRIX     *
C ************************************************************************
C
      INTEGER K(5,5), L(5,5)
      REAL MU, PI(5),LAMDA(5,5),ALPHA(5,5),P(5,5),A(5,5),Q1(5,5)
      REAL AA(5,5,300), BB(5,5,300), X(5,5), Y(5,5)
      REAL R(5,5), S(5,5), X1(5,5), X2(5,5), DIFF(5)
C
      N2 = N*N
C
      DO 30 J = 1 , N
        DO 20 I = 1 , N
          R(I,J) = MU / LAMDA(I,J)
          S(I,J) = MU / ALPHA(I,J)
          Q1(I,J) = A(I,J) * ( P(I,J) / LAMDA(I,J) +
     &                  (1.0-P(I,J)) / ALPHA(I,J) )
 20     CONTINUE
 30     CONTINUE
```

```
C
C ****** COMPUTE AND STORE MATRICES AA'S AND BB'S     ********
C ****** AND DETERMINE THE TRUNCATION INDEX 'ITRUNC' ********
C ****** NOTE: THE INDEX IS OFF BY ONE (E.Q. AA(0)=AA(I,J,1))*
C
      DO 110 I = 1 , N
        DO 100 J = 1 , N
           X(I,J) = R(I,J) / ( K(I,J) + R(I,J) )
           Y(I,J) = S(I,J) / ( L(I,J) + S(I,J) )
          X1(I,J) = A(I,J) * P(I,J) * (1.0-X(I,J))**K(I,J)
          X2(I,J) = A(I,J) * (1.0-P(I,J)) * (1.0-Y(I,J))**L(I,J)
           AA(I,J,1) = X1(I,J) + X2(I,J)
           BB(I,J,1) = A(I,J) - AA(I,J,1)
 100    CONTINUE
 110  CONTINUE
C
C   INITIALIZE THE COMPUTATION FOR THE ITRUNC
C
      DO 130 I = 1 , N
        SUM1 = 0.0
        SUM2 = 0.0
        DO 120 J = 1 , N
           SUM1 = SUM1 +  Q1(I,J)
           SUM2 = SUM2 + BB(I,J,1)
 120    CONTINUE
        DIFF(I) = MU * SUM1 - SUM2
 130  CONTINUE
```

```
C
C     NOW COMPUTE AA AND BB RECURSIVELY
C
      DO 160 LL = 2 , 300
        IFLAG = 0
        IFLAG1 = 0
        DO 150 I = 1 , N
          SUM1 = 0.0
          DO 140 J = 1 , N
            X1(I,J) = X1(I,J) * X(I,J) * (K(I,J)+LL-2.0) / (LL-1.0)
            X2(I,J) = X2(I,J) * Y(I,J) * (L(I,J)+LL-2.0) / (LL-1.0)
            AA(I,J,LL) = X1(I,J) + X2(I,J)
            BB(I,J,LL) = BB(I,J,LL-1) - AA(I,J,LL)
            IF ( AA(I,J,LL) .LE. 1.0E-6) IFLAG1 = IFLAG1 + 1
            SUM1 = SUM1 + BB(I,J,LL)
 140      CONTINUE
          DIFF(I) = DIFF(I) - SUM1
            IF ( DIFF(I) .LE. 0.0) THEN
              ITRUNC = LL-1
C              WRITE(6,156) ITRUNC
              GO TO 200
            ENDIF
          IF (DIFF(I) .LT. 1.0E-8) IFLAG = IFLAG + 1
 150    CONTINUE
        IF (IFLAG .EQ. N .OR. IFLAG1 .EQ. N2) THEN
           ITRUNC = LL
C            WRITE (6,156) ITRUNC
C156          FORMAT (/,5X,'TRUNCATION INDEX =',I5,//)
           GO TO 200
        ENDIF
 160  CONTINUE
C
      ITRUNC = 300
      WRITE (8,166)
 166  FORMAT (/,3X,'WATCH OUT !! THE TRUNCATION INDEX ',
     &          'IS TOO LARGE (BIGGER THAN 300).',/,
     & 'BUT WE LET THE CALCULATION TO CONTINUE WITH ITRUNC=300')
C
C **** BUT WE LET THE CALCULATION TO CONTINUE WITH K = 300 ****
C
 200  RETURN
      END
```

```
      SUBROUTINE NEUTS ( N, MU, AA, BB, ITRUNC, PI, Q1,
     &          RHO,&ETA, R, ENA, STDEVA, PIDLE, ENT, STDEVT )

C ***********************************************************************
C                                                                       *
C THIS ROUTINE IS TO BE USED WITH MRMAP                                 *
C                                                                       *
C THIS SUBROUTINE FINDS THE R MATRIX                                    *
C AND ITS EIGENVALUES FOR THE MR/M/1 QUEUE                              *
C USING NEUTS' MATRIX GEOMETRIC METHOD                                  *
C A CALL TO MRM1 IS NECESSARY BEFORE                                    *
C CALLING THIS SUBROUTINE                                               *
C ***********************************************************************
C
C   LAST UPDATE: MARCH 3, 1988
C   IMSL VERSION 10.0 IS USED IN THIS PROGRAM ........
C
C ***********************************************************************
C N     = NUMBER OF TYPES OF CUSTOMERS IN THE ARRIVAL PROCESS.          *
C A     = TRANSITION MATRIX OF THE UNDERLYING MARKOV CHAIN IN           *
C          THE ARRIVAL PROCESS.                                         *
C  PI   = THE STATIONARY PROBABILITY VECTOR OF A.                       *
C Q1    = THE MATRIX OF MEANS INTERARRIVAL TIMES.                       *
C RHO = TRAFFIC INTENSITY.                                              *
C EIGR= THE (COMPLEX OR REAL) EIGENVALUES OF R.                         *
C ETA  = THE PRINCIPAL EIGENVALUE OF R.                                 *
C ENA  = MEAN QUEUELENGTH EMBEDDED AT ARRIVAL TIMES                     *
C VARNA = VARIANCE OF THE QUEUELENGTH AT ARRIVAL TIMES                  *
C ENT = MEAN QUEUELENGTH ( CONTINUOUS TIME )                            *
C VARNT = VARIANCE OF THE QUEUELENGTH ( CONTINUOUS TIME )               *
C ***********************************************************************
C
      REAL Q1(5,5), MU, EIGR(10), RZ(50), X(5,5), Z(5,5)
      REAL AA(5,5,300), BB(5,5,300), WKMAT(5,5)
      REAL R(5,5), RINV(5,5), PI(5), WKVEC(40), A1INV(5,5)
      COMPLEX W(5), CZ(5,5)
      EQUIVALENCE ( W(1),EIGR(1) ), ( CZ(1,1),RZ(1) )
C
C ********** CHECK THE POSITIVE RECURRENCE CONDITION *******
      SUM = 0.0
      DO 30 J = 1 , N
        DO 20 I = 1 , N
          SUM = SUM + PI(I) * Q1(I,J)
 20     CONTINUE
 30   CONTINUE
```

```
      RHO = 1.0 / ( MU * SUM )
      IF (RHO .GE. 1.0) THEN
         WRITE (8,50)
 50 FORMAT(//,2X,'THE MARKOV CHAIN IS NOT POSITIVE RECURRENT!!',/
&        2X,'============================================ ')
        WRITE (8,60) RHO
 60      FORMAT(///,3X,'TRAFFIC INTENSITY =',F9.5,//)
        STOP
      ENDIF
C
C ****** THE MATRICES AA'S AND BB'S AND   ********************
C ****** THE TRUNCATION INDEX 'ITRUNC' **********************
C ****** ARE COMPUTED IN THE SUBROUTINE MRM1 *****************
C ****** NOTE: THE INDEX IS OFF BY ONE (E.Q. AA(0)=AA(I,J,1))*
C
C ***** INITIALIZE R AND COMPUTE AND STORE (I-A1) INVERSE *****
      DO 210 J = 1 , N
        DO 205 I = 1 , N
          Z(I,J) = -AA(I,J,2)
          R(I,J) = 0.0
 205    CONTINUE
        Z(J,J) = 1.0 + Z(J,J)
 210  CONTINUE
C
      LDA = 5
      LDAINV = 5
      CALL LINRG (N,Z,LDA,A1INV,LDAINV)
C ****************************************************
C   BEGIN THE ITERATIVE PROCEDURE TO FIND R .
C ****************************************************
C
      DO 300 ITER = 1 , 500
C        *** EVALUATE A0 + R**2*A2 + R**3*A3 + ... + R**K*AK   ***
C        *** USING HORNER'S ALGORITHM (NESTED MULTIPLICATION). ***
        DO 220 J = 1 , N
        DO 220 I = 1 , N
          Z(I,J) = AA(I,J,ITRUNC)
 220    CONTINUE
        DO 260 L = ITRUNC-1, 3, -1
          CALL MATMUL (N,R,Z,WKMAT)
          DO 250 J = 1 , N
          DO 250 I = 1 , N
            Z(I,J) = AA(I,J,L) + WKMAT(I,J)
 250      CONTINUE
 260    CONTINUE
        CALL MATMUL(N,R,Z,WKMAT)
        CALL MATMUL(N,R,WKMAT,Z)
        DO 270 J = 1 , N
        DO 270 I = 1 , N
          WKMAT(I,J) = AA(I,J,1) + Z(I,J)
 270    CONTINUE
        CALL MATMUL ( N,WKMAT,A1INV,Z)
C
```

```
C       **** TEST FOR CONVERGENCE ************
C
       DO 280 J = 1 , N
       DO 280 I = 1 , N
         IF ( ABS(Z(I,J) - R(I,J)) .GE. 1.0E-7 ) THEN
           DO 275 JJ = 1 , N
           DO 275 II = 1 , N
             R(II,JJ) = Z(II,JJ)
 275         CONTINUE
           GO TO 300
         ENDIF
 280    CONTINUE
C
C       *********** CONVERGES *************************
       DO 285 J = 1 , N
       DO 285 I = 1 , N
         R(I,J) = Z(I,J)
 285    CONTINUE
C
C       WRITE(6,286) ITER
C286     FORMAT(/,5X,'R CONVERGES IN ',I4,' ITERATIONS',//,
C   &   5X,'THE RATE MATRIX R IS :')
C       DO 290 I = 1 , N
C         WRITE(6,287) (R(I,J),J=1,N)
 287      FORMAT(25X,5F12.6)
C290    CONTINUE
       GO TO 400
 300  CONTINUE
C
C    ************ DOES NOT CONVERGE *****************
    WRITE(8,310)
 310  FORMAT(//,3X,'ACHTUNG !!!!!!!!!!!!!!!!',/,
   &       3X,'R DOES NOT CONVERGE IN 500 ITERATIONS;',/,
   &       3X,'BUT CALCULATIONS CONTINUE WITH R :=  ',//)
    DO 320 I = 1 , N
      WRITE(8,287) (R(I,J),J=1,N)
 320  CONTINUE
C    ****************************************************
C          END OF ITERATION FOR R ....................
C    ****************************************************
C
 400  CONTINUE
C
C ******** COMPUTE B[R]. THIS IS ALSO AN INTERNAL ACCURACY ******
C ******** CHECK ON R; B[R] SHOULD BE STOCHASTIC ......... ******
C
    DO 410 J = 1 , N
    DO 410 I = 1 , N
      Z(I,J) = BB(I,J,ITRUNC)
 410  CONTINUE
C
    DO 450 L = ITRUNC-1, 1 , -1
      CALL MATMUL (N,R,Z,WKMAT)
```

```
      DO 440 J = 1 , N
      DO 440 I = 1 , N
        Z(I,J) = BB(I,J,L) + WKMAT(I,J)
440     CONTINUE
450   CONTINUE
C
C     WRITE(6,456)
C456  FORMAT(///,5X,'INTERNAL ACCURACY CHECK:',/,
C  &          5X,'========================',//,
C  &     15X,'MATRIX B[R] (THE LAST COLUMN IS THE ROW SUM) :',/)
C     DO 470 I = 1 , N
C       SUM = 0.0
C       DO 460 J = 1 , N
C         SUM = SUM + Z(I,J)
C460     CONTINUE
C       WRITE(6,466) (Z(I,J),J=1,N), SUM
C466     FORMAT(25X,5F12.6)
C470     CONTINUE
C
C     ********************************************************
C     COMPUTE THE EIGENVALUES OF THE RATE MATRIX R
C     ********************************************************
C
    DO 481 J = 1 , N
      DO 480 I = 1 , N
        WKMAT(I,J) = R(I,J)
480     CONTINUE
481   CONTINUE
    CALL EVLRG ( N, WKMAT, LDA, EIGR )
```

```
C    *** FIND ETA, THE LARGEST POSITIVE REAL EIGENVALUE OF R  ***
C
     ETA = -99999.9
     NN = N + N
     DO 490 I = 1 , NN, 2
       IF ( EIGR(I) .GT. ETA ) ETA = EIGR(I)
 490 CONTINUE
C
C ************ PRINT THE EIGENVALUES *****************
C
C    WRITE(6,493) (EIGR(I),I=1,NN)
C493 FORMAT(//,5X,'THE EIGENVALUES OF R:',/,
C   &       5(20X,F10.5,' + ',F10.5,' I',/),/)
C    WRITE(6,496) ETA
C496 FORMAT(/,'THE PRINCIPAL EIGENVALUE : ETA = ',F12.5)
C
C    ******* COMPUTE THE IDLE PROBABILITY: PI*(I-R) *********
C
     DO 498 J = 1 , N
       DO 497 I = 1 ,N
         WKMAT(I,J) = - R(I,J)
 497   CONTINUE
       WKMAT(J,J) = 1.0 + WKMAT(J,J)
 498 CONTINUE
C
     PIDLE = 0.0
     DO 503 J = 1 , N
       DO 502 I = 1 , N
         PIDLE = PIDLE + PI(I) * WKMAT(I,J)
 502   CONTINUE
 503 CONTINUE
C
C    ******* COMPUTE  (I-R) INVERSE ........ ***************
     CALL LINRG ( N, WKMAT, LDA, RINV, LDAINV )
C
C ******* COMPUTE THE MEAN AND VARIANCE OF THE MARGINAL ******
C ******* QUEUE LENGTH DISTRIB. EMBEDDED AT ARRIVAL TIMES ******
C
     CALL MATMUL (N, RINV, R, WKMAT)
     ENA = 0.0
     DO 510 J = 1 , N
       DO 500 I = 1 , N
         ENA = ENA + PI(I) * WKMAT(I,J)
 500   CONTINUE
 510 CONTINUE
     CALL MATMUL (N, RINV, WKMAT, Z)
C
     DO 530 J = 1 , N
       DO 520 I = 1 , N
         X(I,J) = R(I,J)
 520   CONTINUE
       X(J,J) = X(J,J) + 1.0
 530 CONTINUE
```

```
C
      CALL MATMUL (N, Z, X, WKMAT)
      VARNA = 0.0
      DO 550 J = 1 , N
        DO 540 I = 1 , N
          VARNA = VARNA + PI(I) * WKMAT(I,J)
 540    CONTINUE
 550  CONTINUE
C
      VARNA = VARNA - ENA * ENA
      STDEVA = SQRT (VARNA)
C
C ******* COMPUTE THE MEAN AND VARIANCE OF THE MARGINAL ******
C ******* QUEUE LENGTH DISTRIB. AT ARBRITRARY TIMES ..... ******
C
      ENT = 0.0
      DO 561 J = 1 , N
        DO 560 I = 1 , N
          ENT = ENT + PI(I) * RINV (I,J)
 560    CONTINUE
 561  CONTINUE
      ENT = RHO * ENT
C
      CALL MATMUL (N, RINV, RINV, Z)
      CALL MATMUL (N, Z, X, WKMAT )
C
      VARNT = 0.0
      DO 580 J = 1 , N
        DO 570 I = 1 , N
          VARNT = VARNT + PI(I) * WKMAT(I,J)
 570    CONTINUE
 580  CONTINUE
      VARNT = RHO * VARNT
      VARNT = VARNT - ENT * ENT
      STDEVT = SQRT (VARNT)
      RETURN
      END
```

SUBROUTINE MATMUL (N,A,B,C)

```
   REAL A(5,5),B(5,5),C(5,5)
C *************************************************************************
C MATRIX MULTIPLICATION ROUTINE THAT MINIMIZES STRIDES                    *
C AND MAKE USE OF VECTOR CAPABILITY OF VECTOR PROCESSOR                   *
C IF AVAILABLE ......................................                     *
C                                                                         *
C *************************************************************************
C
   DO 20 J = 1 , N
     DO 10 I = 1 , N
       C(I,J) = 0.0
10   CONTINUE
     DO 20 K = 1 , N
     DO 20 I = 1 , N
       C(I,J) = C(I,J) + A(I,K) * B(K,J)
20 CONTINUE
   RETURN
   END
```

The vita has been removed from the scanned document