

# **CLOSURE METHODS FOR THE SINGLE-SERVER RETRIAL QUEUE**

by

**DAVID WARREN SABO**

B. Sc. (Honours), The University of Alberta, 1972  
Ph. D., The University of British Columbia, 1977

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE**

in

**THE FACULTY OF GRADUATE STUDIES  
(Department of Mathematics)**

**We accept this thesis as conforming  
to the required standard:**

**THE UNIVERSITY OF BRITISH COLUMBIA**

**June, 1987**

**©David Warren Sabo, 1987**

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Mathematics

The University of British Columbia  
1956 Main Mall  
Vancouver, Canada  
V6T 1Y3

Date June 19/87

## **ABSTRACT**

This work focusses on the development and evaluation of so-called "closure methods" for solving the equations governing the time-dependent behaviour of single-server retrial queues. These methods involve assuming that particular known algebraic relationships between various characteristics of the corresponding steady-state queue also apply approximately when the queue is not at steady-state. The objective is to replace a problem requiring the solution of dozens or hundreds of simultaneous linear differential equations with a system of a few differential equations that has a solution that approximates those queue characteristics of immediate interest. The viability of such closure methods is assessed by examining the results of a series of test calculations.

The methods described in this thesis apply to a retrial queue in which inter-arrival times for new customers, inter-retrial times, and service times are all assumed to be exponentially distributed. The steady-state solution for such a queue is described in some detail. A survey of the literature indicates that the description of this steady-state retrial queue has become quite sophisticated, whereas only very tentative steps have been taken in the study of the time-dependent behaviour of such queues. On the other hand, the time-dependent behaviour of the simple M/M/s queues have been studied to a much greater extent. The apparent value of closure methods in computing approximations to various basic time-dependent M/M/s queue characteristics motivated this examination of the extension of such methods to the single-server retrial queue.

After discussing the basic approach to be used in devising and testing prospective closure methods for the single-server retrial queue, a variety of such methods is presented, with each being tested in considerable detail. It is found that three of the methods devised give results of comparable or better accuracy than those closure methods for the simple M/M/s queues which motivated this study. All recommended closure methods developed here involve systems of either two or three differential equations and permit the calculation of good approximations to four of the characteristics of greatest interest for non-stationary queues: the probability that the server is idle, the mean queue length, the variance of the queue length, and the conditional mean number of customers in the system given that the server is idle. Each of the methods presented is tested for queues with constant mean arrival, retrial and service rates, as well as for queues in which arrival and retrial rates vary sinusoidally with time.

## TABLE OF CONTENTS

Abstract . . . . .	ii
List of Tables . . . . .	vi
List of Figures . . . . .	vii
Acknowledgements . . . . .	x
 CHAPTER 1 Review of the Retrieval Queue, Time Dependence and the Notion of Closure Methods . . . . .	 1
1.1 The Single Server Retrieval Queue . . . . .	2
Description of the Retrieval Queue . . . . .	2
Equations Governing Queue Behaviour . . . . .	4
The Steady State Solution . . . . .	10
Evaluation of Steady State Moments . . . . .	13
Related Queueing Models . . . . .	15
1.2 Retrieval Queues to Date . . . . .	18
Cohen . . . . .	18
Keilson, Cozzolini and Young . . . . .	19
Alexsandrov . . . . .	21
Falin . . . . .	21
Hashida and Kodaira . . . . .	22
Choo and Conolly . . . . .	22
Kulkarni . . . . .	24
Conclusions . . . . .	25
1.3 Time-Dependence of Queues . . . . .	27
Why Bother With Explicit Time Dependence? . . . . .	27
Koopman's Work . . . . .	28
Solution by Power Series Expansion . . . . .	29
Determination of Time-Dependent Probability Generating Functions . . . . .	32
Transform Methods . . . . .	34
Previous Work on Time-Dependent Queues . . . . .	35
Numerical Approaches . . . . .	37
Truncation of Systems of Differential Equations . . . . .	41
1.4 Closure Approximations for M/M/s Queues . . . . .	46
Details of the Rothkopf and Oren Closure Method . . . . .	46
Numerical Results . . . . .	49
Period-Mean Values and Period-Mean Errors . . . . .	54
Extensions and Other Approaches . . . . .	57

CHAPTER 2	Closure Techniques for the Time-Dependent Single-Server Retrial Queue . . . . .	66
2.1	The Basic Equations . . . . .	67
	Retrial Queue Characteristics of Immediate Interest . . . . .	67
	Derivation of the Basic Equations . . . . .	68
	Additional Criteria . . . . .	70
	Notation . . . . .	75
2.2	One-Equation Closure Methods . . . . .	76
	Method 1A ( $\Pi_0(t)$ ) . . . . .	76
	Method 1B ( $m(t)$ ) . . . . .	80
2.3	Two-Equation Closure Methods: $\Pi_0$ and $m$ . . . . .	87
	Methods 2A and 2B ( $\Pi_0(t)$ and $m(t)$ ) . . . . .	87
	Numerical Performance of Methods 2A and 2B . . . . .	89
	Method 2C ( $\Pi_0(t)$ and $m(t)$ ) . . . . .	90
	Constant Parameter Calculations Using Method 2C . . . . .	91
	Oscillatory Parameter Calculations Using Method 2C . . . . .	93
	Computation of Period-Averages of $\Pi_0(t)$ and $m(t)$ . . . . .	100
2.4	Two-Equation Closure Methods: $\Pi_0$ , $m$ and $v$ . . . . .	114
	Method 2D ( $\Pi_0$ , $m$ , $v$ ) . . . . .	114
	Numerical Performance of Method 2D . . . . .	116
	Method 2E ( $\Pi_0$ , $m$ , $v$ ) . . . . .	118
	Constant Parameter Calculations Using Method 2E . . . . .	119
	Oscillatory Parameter Calculations Using Method 2E . . . . .	122
	Computation of Period-Averages of $m(t)$ and $v(t)$ . . . . .	123
	Estimation of $m_0(t)$ Using Method 2E . . . . .	127
2.5	Closure Methods With Three Or More Equations . . . . .	145
	Three Equation Methods . . . . .	145
	Closure Methods Involving Four or More Equations . . . . .	149
2.6	Summary and Conclusions . . . . .	153
Appendix A	Details of Calculation of Exact Queue Properties. . . . .	157
Bibliography	. . . . .	159

## LIST OF TABLES

TABLE 1	Period-Mean Values of $p_0(t)$ , $m(t)$ , and $v(t)$ and Their Relative Errors for the M/M/1 Queue When the Rothkopf and Oren Closure Method is Used . . . . .	59
TABLE 2	Formulas for Moment Derivatives for the Single Server Retrial Queue . . . . .	71
TABLE 3a	Exact Period-Mean Values of $\Pi_0(t)$ and $m(t)$ and Mean Absolute and Relative Mean Absolute Errors When These Quantities are Computed Using Closure Method 2C for Various Values of $\rho_0$ . . . . .	102
TABLE 3b	Exact Period-Mean Values of $\Pi_0(t)$ and $m(t)$ and Mean Absolute and Relative Mean Absolute Errors When These Quantities are Computed Using Closure Method 2C for Various Values of $c_0$ or $v_0$ . . . . .	103
TABLE 4a	Period Averages of $\Pi_0(t)$ and $m(t)$ for Various New Customer Arrival Rates, Based on Method 2C . . . . .	104
TABLE 4b	Period Averages of $\Pi_0(t)$ and $m(t)$ for Various Retrial Rates for the Single-Server Retrial Queue, Based on Method 2C . . . . .	105
TABLE 5a	Period Averages and Mean Absolute Errors in $\Pi_0(t)$ , $m(t)$ , and $v(t)$ for Various New Customer Arrival Rates Using Method 2E . . . . .	129
TABLE 5b	Period Averages and Mean Absolute Errors in $\Pi_0(t)$ , $m(t)$ , and $v(t)$ for Various Retrial Rates Using Method 2E . . . . .	131
TABLE 6	Period Averages and Relative Mean Absolute Errors in $\Pi_0(t)$ , $m(t)$ and $v(t)$ Using Method 2E With Various Amounts of Mixing, $\eta$ , in the Closure Formula for $m_0(t)$ . . . . .	133
TABLE 7	Summary of Serviceable Closure Methods for Computing the Time-Dependent Behaviour of the Single-Server Retrial Queue . . . . .	155

## LIST OF FIGURES

FIGURE 1.1	Schematic Representation of a Single-Server Retrial Queue . . . . .	3
FIGURE 1.2	Possible State Transitions and Differential Transition Rates for the Single-Server Retrial Queue . . . . .	8
FIGURE 1.3	The Runge-Kutta-Felberg-45 Algorithm . . . . .	39
FIGURE 1.4	Relative Error in $p_Q(t)$ Computed Using Rothkopf and Oren's Closure Method for the Simple M/M/1 Queue With Constant Arrival and Service Rates . . . . .	60
FIGURE 1.5	Relative Error in $m(t)$ Computed Using Rothkopf and Oren's Closure Method for the Simple M/M/1 Queue With Constant Arrival and Service Rates . . . . .	61
FIGURE 1.6	Relative Error in $v(t)$ Computed Using Rothkopf and Oren's Closure Method for the Simple M/M/1 Queue With Constant Arrival and Service Rates . . . . .	62
FIGURE 1.7	Comparison of Exact $p_Q(t)$ and $p_Q(t)$ Computed Using Rothkopf and Oren's Closure Method for the M/M/1 Queue in the Oscillatory Parameter Case . . . . .	63
FIGURE 1.8	Comparison of Exact $m(t)$ and $m(t)$ Computed Using Rothkopf and Oren's Closure Method for the M/M/1 Queue in the Oscillatory Parameter Case . . . . .	64
FIGURE 1.9	Comparison of Exact $v(t)$ and $v(t)$ Computed Using Rothkopf and Oren's Closure Method for the M/M/1 Queue in the Oscillatory Parameter Case . . . . .	65
FIGURE 2.1	$\Pi_Q(t)$ Computed Using Method 1A in the Constant Parameter Case . . . . .	83
FIGURE 2.2	Comparison of $\Pi_Q(t)$ Computed Approximately Using Method 1A With Its Exact Values in the Oscillatory Parameter Case . . . . .	84
FIGURE 2.3	Mean Number of Customer, $m(t)$ , in the Retrial Queue Computed Using Method 1B in the Constant Parameter Case . . . . .	85

FIGURE 2.4	Mean Number of Customers, $m(t)$ , in the Retrial Queue Computed Using Method 1B in the Oscillatory Parameter Case . . . . .	86
FIGURE 2.5	Relative Errors in $\Pi_0(t)$ and $m(t)$ When Computed Using a Two-Equation Closure Approximation with Basic Methods 'A' and 'B' in the Constant Parameter Case . . .	106
FIGURE 2.6	Relative Errors in $\Pi_0(t)$ and $m(t)$ When Computed Using a Two-Equation Closure Approximation with Basic Methods 'A' and 'B' in the Oscillatory Parameter Case . . . . .	107
FIGURE 2.7	Relative Errors in $\Pi_0(t)$ Using the Two-Equation Closure Method 'C' in the Constant Parameter Case With Various Arrival and Retrial Rates . . . . .	108
FIGURE 2.8	Relative Errors in $m(t)$ Using the Two-Equation Closure Method 'C' in the Constant Parameter Case With Various Arrival and Retrial Rates . . . . .	109
FIGURE 2.9	$\Pi_0(t)$ Computed Using the Two-Equation Closure Method 'C' in the Oscillatory Parameter Case for Various Values of $\rho_0$ . . . . .	110
FIGURE 2.10	$m(t)$ Computed Using the Two-Equation Closure Method 'C' in the Oscillatory Parameter Case for Various Values of $\rho_0$ . . . . .	111
FIGURE 2.11	$\Pi_0(t)$ Computed Using the Two-Equation Closure Method 'C' in the Oscillatory Parameter Case for Various Values of $c_0$ . . . . .	112
FIGURE 2.12	$m(t)$ Computed Using the Two-Equation Closure Method 'C' in the Oscillatory Parameter Case for Various Values of $c_0$ . . . . .	113
FIGURE 2.13	Relative Errors in $\Pi_0(t)$ , $m(t)$ and $v(t)$ Using Closure Method 2D for Various Values of $\rho$ in the Constant Parameter Retrial Queue . . . . .	137
FIGURE 2.14	Relative Errors in $\Pi_0(t)$ , $m(t)$ and $v(t)$ Using Closure Method 2D for Various Values of $c_0$ in the Constant Parameter Retrial Queue . . . . .	138
FIGURE 2.15	Comparison of Exact and Approximate Quantities for the Retrial Queue Using Method 2D . . . . .	139

FIGURE 2.16	Comparison of Exact and Closure Method 2D Approximation Values of $\Pi_0(t)$ and $m(t)$ for Oscillatory Parameter Retrial Queues . . . . .	140
FIGURE 2.17	Relative Errors in $v(t)$ Computed Using the Closure Method 2E in the Constant Parameter Case . . . . .	141
FIGURE 2.18	Performance of the Closure Method 2E When the Retrial Rate is Much Smaller Than the New Customer Arrival Rate . . . . .	142
FIGURE 2.19	Comparison of Exact and Method 2E Approximation $v(t)$ for Various Values of $\eta$ in Equation (2.4.9) in the Oscillatory Parameter Case . . . . .	143
FIGURE 2.20	Comparison of Exact $m_0(t)$ with $m_0(t)$ Given by Equation (2.4.9) for Selected Values of $\eta$ in an Oscillatory Parameter Retrial Queue . . . . .	144
FIGURE 2.21	Relative Errors and Comparison of Exact Quantities With the Results of the Closure Method 3A for a Constant Parameter Retrial Queue . . . . .	150
FIGURE 2.22	Comparison of Exact Quantities With the Results of the Closure Method 3A for an Oscillatory Parameter Retrial Queue . . . . .	151
FIGURE 2.23	Comparison of Exact Quantities With the Results of the Closure Method 3B for an Oscillatory Parameter Retrial Queue . . . . .	152

## ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to Dr. P. E. Greenwood for her guidance and many helpful suggestions regarding the work which lead to the completion of this thesis. I would also like to acknowledge the assistance of Dr. G. Hooghiemstra, who introduced me to the retrieval queue. Dr. R. A. Restrepo and Dr. M. Sion have had significant influence on the progress of my studies for this degree from the beginning.

My studies for this degree have involved nearly five years as a part-time student in the program, and one year before that as an unclassified student. I would like to thank my full-time employer, the British Columbia Institute of Technology, for allowing me to rearrange my work schedule to permit me to pursue these studies, and for providing financial support for some aspects of the program.

Most of all, I wish to acknowledge the tolerance displayed by my wife, Marlene, and by my children, Marlis and Robert, of a part-time program which all too often dominated evenings, weekends, and summers that could have been used very well in activities of greater interest to them.

## CHAPTER 1

### Review of the Retrial Queue, Time Dependence, and the Notion of Closure Methods

*'Consider your verdict,' the King said to the jury.*

*'Not yet, not yet!'*, the Rabbit hastily interrupted.

*'There's a great deal to come before that!'*

(Lewis Carroll, Alice in Wonderland)

## 1.1 The Single Server Retrial Queue

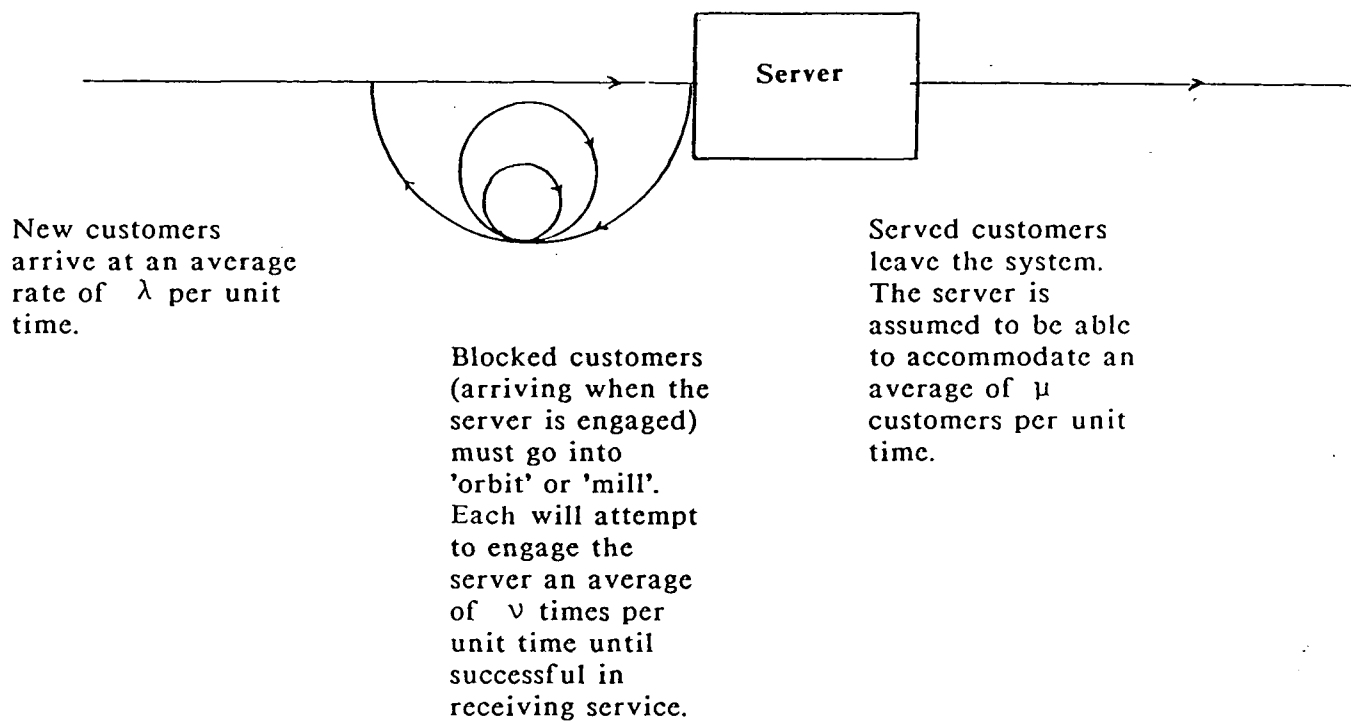
*I have seen something under the sun: The race is not to the swift or the battle to the strong, nor does food come to the wise or wealth to the brilliant or favour to the learned; but time and chance happen to them all.*

(Ecclesiastes 9:11, NIV)

### Description of the Retrial Queue

This thesis deals with an approach to computing the time dependent behaviour of busy telephones and similar systems. The main results apply to the simplest such system, denoted technically as the M/M/1/1 retrial queue. The situation to be dealt with can be described schematically as in Figure 1.1.

Here, the box labelled S represents the 'server', which at any given time can be idle ( $\xi = 0$ ) or busy ( $\xi = 1$ ). If the server is idle when a customer approaches, the customer receives immediate service. On the other hand, if the customer finds the server busy, he must abandon his attempt to engage the server, and goes into 'orbit'. Blocked (or orbiting or milling) customers are presumed to continue to approach the server repeatedly until they find the server idle, receive service, and exit the system, though it is possible to formulate more sophisticated models in which an orbiting customer eventually tires and abandons further attempts at service [Cohen, 1957]. The characteristics of this system which we will be interested in calculating are quantities like the number or mean number of customers in the system and its variance as a function of time, the fraction of the time that the server is idle, the length of time a customer can expect spend in the system to receive service, and so on.



**FIGURE 1.1** Schematic Representation of a Single-Server Retrial Queue.

What makes the consideration of this problem nontrivial is that in actual systems of this type, the exact

- times at which new customers enter the system
- times at which orbiting customers will 'retry' for service
- length of time the server will spend with each customer

are random variables. This means that rather than being able to state precisely when a particular customer (the first or the tenth, etc.) will first approach the server, or precisely when the server will be approached by a new or orbiting customer, or precisely how long the server will spend with a particular customer, we will only be able to speak of the probability of such events occurring in time intervals of some length or of the expectation of these quantities. The degree of complexity this indeterminacy brings to the analysis of the problem depends on the nature of the probability distributions of the random variables involved.

#### Equations Governing Queue Behaviour

We assume that the arrival times of new and retrying customers are statistically independent and that the probability of more than one arrival of each type occurring in some time interval approaches zero as the length of the time interval approaches zero. While it is possible to envision unusual circumstances under which, say, telephone calls placed to a particular receiver (server) do not

conform to these assumptions<sup>1</sup>, they do reflect a sufficiently wide range of experience to justify adopting. These assumptions are consistent with adopting the exponential probability distribution for interarrival times for new as well as retrying customers. A similar, though weaker, case can be made for service times having the exponential distribution.<sup>2</sup> An immediate consequence of assuming that arrival, retrial, and service rates are exponentially distributed is that the state of the system is completely specified by the pair of numbers  $(\xi, n)$ , with  $\xi = 0$  or  $1$  denoting the server as idle or busy respectively, and  $n$  indicating the number of customers in orbit.<sup>3</sup>

Let  $\lambda$  be the mean number of new customers entering the system per unit time,  $\nu$  be the mean number of retrials a customer makes per unit time and  $\mu$  be the mean number of customers the server can handle per unit time. For the

---

<sup>1</sup>The telephone calls of a Wall Street stock broker on the afternoon of October 24, 1929 comes to mind as a potential example of a of a retrial queue in which both assumptions are violated.

<sup>2</sup>In his first paper on the subject, Cohen (1957) refers to experimental results which 'fully justify' the adoption of the exponential distribution for interarrival times as well as service times. Because the exponential distribution does reflect the behaviour of a considerable variety of systems which can be modelled as queues, and because of its mathematical simplicity, it is usual to adopt this distribution for interarrival times and service times as a first approximation at least. Since the initial goal here is to develop computational techniques for a model already appearing in the literature, nothing further will be said in attempting to justify the model (though the issue is not entirely irrelevant).

<sup>3</sup>No other continuous probability distributions share the 'lack of memory' or Markov property of the exponential distribution (see, for example, Ross (1983), pages 24 -25). If interarrival times or service times were assumed to follow some other such distribution, then in addition to the values of  $\xi$  and  $n$ , a complete description of the state of the system might also require the time at which the most recent customer arrived or the time at which the currently served customer entered service.

moment, we can think of  $\lambda$ ,  $\mu$ , and  $\nu$ , as being constants, though the end result of the derivation below does not change materially if they are sufficiently smooth functions of time. Using the properties of the negative exponential distribution, we can state the following for some small time interval  $\Delta t$ :

$$\Pr(\text{a new customer arrives in the interval } \Delta t) = \lambda \Delta t + o(\Delta t), \quad (1a)$$

$$\Pr(\text{service of a customer is completed in the interval } \Delta t | \text{ the system is not empty}) = \mu \Delta t + o(\Delta t), \quad (1b)$$

$$\Pr(\text{a customer retries for service in the interval } \Delta t | \text{ there are } n \text{ customers in orbit}) = \nu_n \Delta t + o(\Delta t), \quad (1c)$$

$$\Pr(\text{more than one new customer arrives in } \Delta t) = o(\Delta t), \quad (1d)$$

$$\Pr(\text{more than one customer completes service in } \Delta t) = o(\Delta t), \quad (1e)$$

$$\Pr(\text{more than one customer retries for service in } \Delta t) = o(\Delta t), \quad (1f)$$

$$\Pr(\text{more than one event of any type in } \Delta t) = o(\Delta t). \quad (1g)$$

The notation  $o(\Delta t)$  indicates a quantity that becomes negligible compared to  $\Delta t$  as  $\Delta t$  approaches zero (in an abuse of notation, we can say that  $o(\Delta t)$  is any 'quantity' satisfying  $\lim_{\Delta t \rightarrow 0} o(\Delta t)/\Delta t = 0$ .)

It is now a simple matter to derive a system of differential equations for the probabilities  $\{p_{\xi n}(t), \xi = 0, 1; n = 0, 1, 2, \dots\}$  where

$p_{0n}(t) = \text{Pr}(\text{server is idle and there are } n \text{ customers in orbit at time } t),$

$$n = 0, 1, 2, \dots$$

and

$p_{1n}(t) = \text{Pr}(\text{server is busy and there are } n \text{ customers in orbit at time } t),$

$$n = 0, 1, 2, \dots$$

To do this, we simply use the so-called law of total probability, which can be written here as

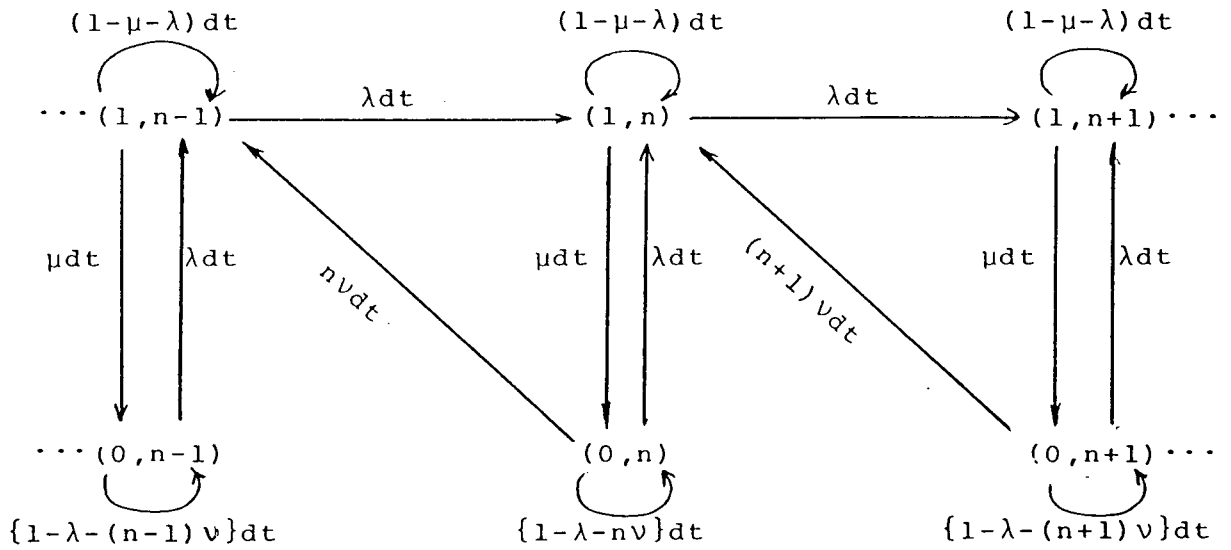
$$p_{\xi n}(t + \Delta t) = \sum_{k=0}^{\infty} \sum_{\omega=0}^1 p_{k\omega}(t) \Pi_{\omega k, \xi n}(\Delta t), \quad \begin{matrix} \xi = 0, 1 \\ n = 0, 1, 2, \dots \end{matrix} \quad (2)$$

where  $\Pi_{\omega k, \xi n}(\Delta t)$  is the conditional probability that the system will be in state  $(\xi, n)$  at time  $t + \Delta t$  given that it was in state  $(\omega, k)$  at time  $t$ .<sup>4</sup>

If we are concerned only with retaining terms to first order in  $\Delta t$  in equations (2), then the conditions given in equations (1) drastically reduce the number of nonzero terms in the summation above. The solid arrows in Figure 1.2 represent the  $\Pi_{\omega k, \xi n}(\Delta t)$  which are first order in  $\Delta t$  or larger as  $\Delta t$  becomes small. Using that diagram, for example, we have for any  $n = 0, 1, 2, \dots$ :

---

<sup>4</sup>Here,  $\Pi_{\omega k, \xi n}(\Delta t)$  depends only on  $\Delta t$  because all the interarrival and service times are assumed exponentially distributed. If that were not true, the  $\Pi_{\omega k, \xi n}$  would also depend on the value of  $t$ .



**FIGURE 1.2** Possible State Transitions and Differential Transition Rates for the Single-Server Retrial Queue

The six pairs  $(\xi, n)$  above represent possible states of the retrial queue, with  $\xi = 0$  or  $1$  corresponding to the server being idle or busy respectively, and 'n' being the number of customers in orbit. The arrows represent possible system state transitions for this single-server retrial queue with the corresponding differential transition rates as indicated.

$$\begin{aligned}
p_{0n}(t + \Delta t) &= p_{0n}(t) \cdot \Pi_{0n,0n}(\Delta t) + p_{1n}(t) \cdot \Pi_{1n,0n}(\Delta t) \\
&= p_{0n}(t) \cdot \text{Pr}(\text{the system is in state } (0,n) \text{ at time } t \text{ and no} \\
&\quad \text{arrivals of either type occur in the} \\
&\quad \text{interval } \Delta t) \\
&\quad + p_{1n}(t) \cdot \text{Pr}(\text{the system is in state } (1,n) \text{ at time } t \text{ and there} \\
&\quad \text{is one service completion in } \Delta t \text{ but no} \\
&\quad \text{arrivals of either type}) \\
&= p_{0n}(t) \{1 - \lambda \Delta t + o(\Delta t)\} \{1 - n\nu \Delta t + o(\Delta t)\} \\
&\quad + p_{1n}(t) \{\mu \Delta t + o(\Delta t)\} \{1 - \lambda \Delta t + o(\Delta t)\} \{1 - n\nu \Delta t + o(\Delta t)\} \\
&= p_{0n}(t) - \lambda p_{0n}(t) \Delta t - n\nu p_{0n}(t) \Delta t + \mu p_{1n}(t) \Delta t \\
&\quad + o(\Delta t) .
\end{aligned}$$

Subtracting  $p_{0n}(t)$  from either side, and dividing this last equation through by  $\Delta t$  gives

$$\frac{p_{0n}(t + \Delta t) - p_{0n}(t)}{\Delta t} = -(\lambda + n\nu) p_{0n}(t) + \mu p_{1n}(t) + \frac{o(\Delta t)}{\Delta t}$$

In the limit as  $\Delta t \rightarrow 0$ , this reduces to

$$\begin{aligned}
\frac{d}{dt} p_{0n}(t) &= -(\lambda + n\nu) p_{0n}(t) + \mu p_{1n}(t), \\
n &= 0, 1, 2, \dots .
\end{aligned} \tag{3a}$$

In a similar fashion we find that

$$\begin{aligned} \frac{d}{dt} p_{1n}(t) = & -(\lambda + \mu) p_{1n}(t) + \lambda p_{1, n-1}(t) + \lambda p_{0n}(t) \\ & + (n+1) \nu p_{0, n+1}(t), \quad n=0, 1, 2, \dots, \end{aligned} \quad (3b)$$

with the understanding that  $p_{\xi n}(t)$  is identically zero whenever  $n < 0$ .

The characteristics of the system which are of practical interest are all calculable from the  $p_{\xi n}(t)$  ( $\xi = 0, 1$ ;  $n = 0, 1, 2, \dots$ ) in a straightforward manner. However, these  $p_{\xi n}(t)$  can be obtained only by solving the combined infinite sets of differential equations in (3a) and (3b).

### The Steady State Solution

An exact solution of the combined system of equations above has been obtained only when  $\lambda$ ,  $\mu$ , and  $\nu$  are constants,  $\lambda < \mu$ , and in the limit that  $t$  becomes infinite. In this case, the  $p_{\xi n}(t)$  become independent of the value of  $t$ , and the values of the constants  $p_{\xi n} = p_{\xi n}(\infty)$  are unaffected by the initial state of the system.

This so-called steady-state solution can be obtained in one of several ways [Gross & Harris, 1974, pages 44-51]<sup>5</sup>. While we will be primarily interested in

---

<sup>5</sup>That such a steady-state solution of equations (3a) and (3b) exists is inferred from the construction of it which follows. This is a common approach (see, for example, Gross and Harris (1974), pages 38 - 40 and Cooper (1981), section 2.3). It is recognized that the existence of such steady state solutions involves the concept of ergodicity, for which some rather general results are

examining techniques for approximating the time dependent behaviour of the retrial queue, the major technique studied makes extensive use of this steady state solution. Our intent is to use the characteristics of the calculable steady state solution to develop approximation techniques which will allow bypassing the full systems of equations in (3a) and (3b).

Perhaps the simplest approach to obtaining the steady state solution to this single server retrial queue problem is to assume the solution exists, replace the derivatives in (3a) and (3b) by zeros, and solve the resultant difference equations recursively. With the derivatives replaced by zeros, the systems (3a) and (3b) can be rearranged to give

$$p_{1n} = \frac{\rho}{c}(c + n)p_{0n}, \quad (4a)$$

$$p_{0,n+1} = \frac{c(\rho + 1)}{(n + 1)\rho} \cdot p_{1n} - \frac{c}{n + 1}(p_{1,n-1} + p_{0n}), \quad (4b)$$

$$n = 0, 1, 2, \dots,$$

where

$$\rho = \frac{\lambda}{\mu} \quad \text{and} \quad c = \frac{\lambda}{\nu}. \quad (5)$$

---

known (see Shiryayev (1984), section I.12 and Chapter V). To quote Gross and Harris (1974, page 40), "It is generally not necessary to appeal to these [ergodic] theorems to determine the conditions under which queueing processes are ergodic, since they usually fall out from other considerations." We can apply the remarks of Cooper (1981, page 19 and following) for the M/M/s queue, to indicate here that if limiting values of  $p_{0n}(t)$  and  $p_{1n}(t)$  exist as  $t \rightarrow \infty$ , then equations (3a) and (3b) imply that their derivatives  $p'_{0n}(t)$  and  $p'_{1n}(t)$  must also have limiting values as  $t \rightarrow \infty$ . But the existence of limiting values for the probabilities is possible only if the limiting values of their derivatives are zeros. If on replacing the derivatives in equations (3a) and (3b) by zeros, we obtain a system of difference equations with no solution or with no acceptable solution (as when  $\lambda > \mu$ , when the probabilities diverge with increasing  $n$  to  $\infty$ ), we will presumably have encountered an example of "other conditions" indicating that the queue is not ergodic.

Alternating between (4a) and (4b) for increasing values of  $n$ , we get successively

$$p_{10} = \rho p_{00}$$

$$p_{01} = c \rho p_{00}$$

$$p_{11} = \rho^2 (c + 1) p_{00}$$

$$p_{02} = \frac{\rho^2 c (c + 1)}{2} p_{00}$$

$$p_{12} = \frac{\rho^3 (c + 1) (c + 2)}{2} p_{00}$$

$$p_{03} = \frac{\rho^3 c (c + 1) (c + 2)}{2 \cdot 3} p_{00}$$

$$\begin{array}{ccc} \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \end{array}$$

and so on. A reasonably clear pattern in the formulas for these steady state probabilities has already emerged, leading us to hypothesize that for general  $n$ ,

$$p_{on} = \frac{\rho^n c (c+1) (c+2) \cdots (c+n-1)}{n!} p_{00} = \binom{c+n-1}{n} \rho^n p_{00}, \quad (6a)$$

$$n = 1, 2, 3, \dots,$$

and

$$p_{ln} = \frac{\rho^{n+1} (c+1) (c+2) \cdots (c+n)}{n!} p_{00} = \binom{c+n}{n} \rho^{n+1} p_{00}, \quad (6b)$$

$$n = 0, 1, 2, \dots$$

Substitution of (6) into (4a) and (4b) confirm that the formulas (6) do indeed satisfy the system of difference equations. Finally, since the sum of all of the steady state probabilities must be 1, a bit of routine algebra allows determination

of  $p_{00}$  as

$$p_{00} = (1 - \rho)^{c+1} \quad (7)$$

which together with the formulas in (6) completes the specification of the steady state probabilities.

We note in passing that  $p_n$ , the probability that at steady state there are a total of  $n$  customers in the system, either in service or in orbit, is given by

$$p_n = p_{0n} + p_{1,n-1} = \binom{c+n}{n} \rho^n (1 - \rho)^{c+1}. \quad (8)$$

These probabilities form the so-called negative binomial distribution with parameters  $c+1$  and  $1 - \rho$ .

### Evaluation of Steady-State Moments

In what follows, we shall make considerable use of formulas for steady state moments for the system. The evaluation of such moments is greatly facilitated by first obtaining corresponding probability generating functions

$$P_0(z) = \sum_{n=0}^{\infty} p_{0n} z^n = \frac{(1 - \rho)^{c+1}}{(1 - \rho z)^c} \quad (9a)$$

$$P_1(z) = \sum_{n=0}^{\infty} p_{1n} z^n = \frac{(1 - \rho)^{c+1}}{(1 - \rho z)^{c+1}} \quad (9b)$$

and

$$P(z) = \sum_{n=0}^{\infty} p_n z^n = P_0(z) + zP_1(z) = \frac{(1-\rho)^{c+1}}{(1-\rho z)^{c+1}}. \quad (9c)$$

The steady state quantities of particular importance are

$$\Pi_0 = \text{Pr}(\text{server idle}) = \sum_{n=0}^{\infty} p_{0n} = P_0(1) = 1 - \rho, \quad (10)$$

$$\Pi_1 = \text{Pr}(\text{server busy}) = \sum_{n=0}^{\infty} p_{1n} = 1 - \Pi_0 = \rho, \quad (11)$$

$$m_0 = \sum_{n=0}^{\infty} n \cdot p_n = \left. \frac{d}{dz} P_0(z) \right|_{z=1} = c\rho, \quad (12)$$

$$m_1 = \sum_{n=0}^{\infty} (n+1) p_n = \left. \frac{d}{dz} P_1(z) \right|_{z=1} = \frac{\rho(c\rho + 1)}{1 - \rho}, \quad (13)$$

$m$  = mean number of customers in the system

$$= m_0 + m_1$$

$$= \frac{\rho(c + 1)}{1 - \rho} \quad (14)$$

$$\zeta_0 = \sum_{n=0}^{\infty} n^2 p_{0n} = \left. \frac{d^2}{dz^2} P_0(z) \right|_{z=1} + m_0 = \frac{\rho c(\rho c + 1)}{1 - \rho} \quad (15)$$

$$\begin{aligned} \zeta_1 &= \sum_{n=0}^{\infty} (n+1)^2 p_{1n} = \left. \frac{d^2}{dz^2} \{zP_1(z)\} \right|_{z=1} + m_1 \\ &= \frac{\rho\{(c\rho + 1)(c\rho + 2) - (1 - \rho)\}}{(1 - \rho)^2} \end{aligned} \quad (16)$$

$$\begin{aligned}
\zeta &= \text{expectation of the square of the number of customers in the system} \\
&= \zeta_0 + \zeta_1 \\
&= \frac{(c+1)(c\rho^2 + \rho^2 + \rho)}{(1-\rho)^2} \quad (17) \\
\eta_0 &= \sum_{n=0}^{\infty} n^3 p_{0n} = \frac{d^3}{dz^3} P_0(z) \Big|_{z=1} + 3\zeta_0 + 2m_0 \\
&= \frac{\rho c}{(1-\rho)^2} \{(\rho c + 1)(\rho c + 2) - \rho(1-\rho)\} \quad (18)
\end{aligned}$$

We also note that the variance,  $v$ , of the number of customers in the system is given by

$$v = \zeta - m^2 = \frac{\rho(c+1)}{(1-\rho)^2} \quad (19)$$

### Related Queueing Models

Before leaving this rather condensed description of the basic system we will be examining in greater detail shortly, it is worth recognizing that the problem is in a sense intermediate between two of the very small number of queueing models for which not only the steady state problem is easily solved, but analytic solutions have been found for all positive values of  $t$ .

If we regard the orbiting customers as forming a sort of balky waiting line, the single server retrial queue is seen to be a generalization of the well-known  $M/M/1/\infty$  single server queue with service in random order (SIRO).<sup>6</sup> The big difference between the two is that whereas in the ordinary  $M/M/1/\infty$  queue, the server will never be idle as long as there are any customers in the system, in the retrial queue, it is quite possible for the server to be idle for a time even though quite a large number of customers are orbiting unserved. We would expect that as the mean retrial rate,  $\nu$ , becomes larger and larger, the retrial queue will look more and more like the  $M/M/1/\infty$ /SIRO queue. We note that the system of equations governing the time-dependent behaviour of this queue have been solved formally only in the case that the arrival and service rates are constants.

Alternatively we can regard the orbiting customers as similar to customers who are not waiting in the system at all, giving the impression that the retrial queue is a bit like an ordinary single server queue with no waiting room at all (ie. the ordinary  $M/M/1/1$  queue). This queueing model is perhaps too simple to provide much useful insight into the characteristics or behaviour of the single server retrial queue.

---

<sup>6</sup>Queues are usually described with a notation developed by Kendall (1953) and others, making use of a collection of symbols separated by slashes:  $A/B/C/D/E$ . Here 'A' indicates the interarrival time distribution, 'B' indicates the service time distribution, 'C' indicates the number of parallel service channels, 'D' indicates the capacity of the system including customers in service, and 'E' indicates the waiting line discipline. An extensive list of the symbols commonly used can be found in most basic books on queueing theory (for example, Gross and Harris (1974), page 9 and pages 464-471). The symbol 'M' is used to indicate the exponential distribution, so the notation  $M/M/1/\infty$  denotes a queue with exponentially distributed interarrival and service times, a single server, and infinite capacity. To indicate that customers in the system are served in random order, we could expand the notation used to read  $M/M/1/\infty$ /SIRO.

The similarity between the single server retrial queue and the ordinary  $M/M/1/\infty$ /SIRO queue is what prompted the investigations described in Chapter 2 below regarding the adaptation to the retrial queue of so-called closure methods which have been applied successfully to computing the approximate time dependent behaviour of the simpler  $M/M/1/\infty$  queues.

It should be noted also that the retrial queue is distinct from so-called 'feedback queues' in which customers may return to the queue to await service again after they have completed service. For example, while a particular customer might have been released from service, the service provided may have been incomplete or unsatisfactory in some way. In the retrial queue, customers return to the queue after an unsuccessful attempt to engage the server, but once served, they are viewed mathematically as leaving the system forever.

## 1.2 Retrial Queues to Date

*Miller's Law: 'You can't tell how deep a puddle is until you step into it.'*

(The Official Rules, Paul Dickson, ed.)

### Cohen

The literature on the subject seems to agree that Cohen [1957] was the first to consider the retrial queue problem in detail. In this lengthy paper, he deals both with the underlying assumptions of the formalism as they apply to telephony as well as with developing the mathematical formalism in great detail.

Cohen actually treats the multiple-server retrial queue in the paper. He uses an approach similar to that outlined in the section 1.1 of this thesis leading to equations (3a) and (3b) to derive the underlying system of differential equations which can be solved for the probabilities of finding the system in any of its possible states. . However, to cope with the considerably greater complexity of the resultant equations whose solutions are the steady state probabilities, he first transforms the equations for the probabilities into a system of a smaller number of equations for probability generating functions.

Despite this stratagem, obtaining the formal solution of the problem is no mean feat. The expression Cohen gets for the mean number of 'congested calls' at steady state (what we would refer to here as the mean number of customers in orbit) is a fraction with both the numerator and denominator containing double index summations over products of generalized Laguerre polynomials and integrals

involving components of probability generating functions for the system. Cohen then goes on to demonstrate how solutions for a variety of special subcases of the problem can be obtained from his general solution. His results are so detailed and comprehensive that no attempt will be made to itemize or summarize them here.

### Keilson, Cozzolini and Young

The next appearance of the retrial queue in the literature seems to be in Keilson, Cozzolini and Young [1968]. These authors consider both the single-server and the two-server retrial queue with exponentially distributed interarrival times and general service time distributions. They consider the time-dependent behaviour of the queues but obtain formal and numerical results for both types of system only at steady state.

Note that when service times have a general (non-exponential) distribution, the single-server retrial queue is characterized by something like a pair  $(n, \xi)$ , where  $n$  is a non-negative integer giving the number of customers in orbit,  $\xi = 0$  indicates the server is idle, and  $\xi > 0$  indicates that the server has been engaged by the current customer for  $\xi$  time units already. Instead of the two discrete values of  $\xi$  (0 and 1 above) required to characterize this system when service times are exponentially distributed,  $\xi$  is now needed to be a non-negative real quantity.

For the steady-state single server retrial queue, Keilson, Cozzolini and Young

derive formulas for the probability that the server is idle ( $P_E = 1 - \lambda \bar{T}$ ), the mean number of customers waiting in orbit ( $\bar{m} = [\frac{1}{2} \lambda^2 \bar{T}^2 + (\frac{\lambda}{\mu}) \lambda \bar{T}] / [1 - \lambda \bar{T}]$ ), the mean waiting time in orbit of a customer ( $\bar{\tau} = \bar{m} / \lambda$ ), and the mean number of attempts at service per customer for each successful call ( $\psi = 1 + \nu \bar{\tau}$ ). All of these quantities are determined by the mean service time,  $\bar{T}$ , which is simply  $1/\mu$  when service times are exponentially distributed.

For the multiple-server case, Keilson, Cozzolini and Young assume exponentially distributed service times as well as exponential arrival and retrial rates. This permits them to describe the system by a pair of numbers,  $(n, \xi)$ , where  $n$  is the number of customers in orbit and  $\xi$  is the number of servers engaged. Thus,  $\xi$  can only assume values from a finite set of non-negative integers.

They obtain formal results for the steady-state of a two-server retrial system using an approach borrowed from the study of current flows in networks of resistors in electrical engineering. Their analysis gives formulas for quantities  $P_m$ ,  $Q_m$ , and  $R_m$ , which are the probabilities that there are  $m$  customers in orbit and zero, one, or two servers busy respectively. The characteristics of the retrial queue mentioned in the previous paragraph for the single-server case can all be computed for this two-server case as sums of multiples of the  $P_m$ ,  $Q_m$ , and  $R_m$ . While use of a computer is necessary because these sums cannot be rearranged into a simple closed algebraic form, the necessary formulas are not difficult to program.

### Alexsandrov

Chronologically, the next treatment of the retrial queue appears in Alexsandrov [1974], who derives a number of results for the M/G/1 system considered earlier by Keilson et. al. [1968]. This short paper concentrates on obtaining expressions for the first two moments of the system at steady state. He also derives expressions for the expected time of sojourn of a customer in the system, the effective arrival rate of retrying customers, and the expected number of retrials each customer will make before succeeding in obtaining service. Alexsandrov's paper provides a number of references to earlier papers on the subject appearing in the Russian literature.

### Falin

The Russian torch, as far as the retrial queue is concerned, is then taken up by G. I. Falin, culminating in a lengthy paper (Falin, [1979]) dealing with, in his words, the 'non-stationary regime of operation of the system', and concentrating on characterizing a 'system busy period'.

Falin defines the busy period for the system as the time interval from when a customer enters service in an otherwise empty system until the system becomes totally free of customers either in service or in orbit. While for non-retrial queues, the system busy period defined in this way would correspond to the server busy period, for the retrial queue there may be many intervals during the system busy period when the server is idle. Expressions are obtained for the

length of the system busy period and the number of service completions in such a busy period. It is noteworthy that this difficult paper makes reference only to the Russian literature.

### Hashida and Kodaira

Meanwhile, Hashida and Kodaira [1975] published a treatment of the retrial queue in which new customer interarrival times were assumed exponentially distributed, but inter-retrial times and service times had general distributions. They were interested in the problem of the inefficient use of circuit switching networks because of line time wasted by customers during repeated unsuccessful attempts to obtain service. Their approach is similar to that of Keilson et. al. [1968], and they produce some numerical results for the probability that incoming calls will find the server busy.

### Choo and Conolly

Choo and Conolly [1979] present some results for the system time of customers, service idle time, and system busy period for the steady state  $M/G/1$  retrial queue. It is in this paper that the term 'orbit' is first used in reference to the situation in which those customers unsuccessful in obtaining service on their first attempt find themselves. Choo and Conolly attempt to extend Aleksandrov's work by obtaining formulas for the customer's system time (waiting time in orbit plus time in eventual service), the server idle time and the system busy period. Their derivations assume service times are exponentially distributed

as are new customer arrival times and retrial times, though some of the results are valid when service times have a general distribution.

Choo and Conolly derive an infinite system of equations for probability distribution functions of the first passage times between various states of the queue. This is motivated by the observation that the system busy period is the first passage time from the state  $(\xi, n) = (1, 0)$  to the state  $(\xi, n) = (0, 0)$ . Unfortunately, it does not appear possible to find an exact time-dependent solution to this system of equations. The authors present computational algorithms for calculating the first and higher moments for these first passage times, from which moments of the system busy period can be obtained. For example, they find that the mean system busy period is given by  $\hat{B}_s = \frac{1}{\lambda} \left[ \frac{1}{p_{00}} - 1 \right]$ , where  $p_{00}$  is the probability that the system is empty (see equation (7) in section 1.1 above).

Choo and Connolly apparently have greater success in obtaining simple algebraic formulas for the mean system time of a customer (the time from when the customer first attempts to obtain service until service is completed for him) and its variance. Unfortunately, Kulkarni [1982] and Falin<sup>1</sup> point out an apparent error in the approach of Choo and Conolly to the analysis of a customer's waiting

---

<sup>1</sup>Private Communication with B. W. Conolly. In a short letter to the editor of the Journal of Applied Probability, appearing immediately after Kulkarni [1982], Conolly acknowledges the 'serious difference of opinion' expressed by Kulkarni [1982] as an error in the waiting time analysis appearing in Choo and Conolly [1979] and laments the apparent lack of collaboration between people in Britain, the United States, and the USSR, interested in the retrial problem.

time in the system. Kulkarni presents a corrected analysis of the problem and finds it impossible to obtain a simple closed form expression for the waiting time. Instead, his result is in the form of a numerically well-behaved convergent infinite series whenever  $\rho < 1$ , with terms defined by a set of recursion relations.

Choo [1978] considered the retrial queue in his Ph. D. thesis as well.

### Kulkarni

The most recently appearing paper on the retrial queue seems to be Kulkarni [1983]. Kulkarni develops a simple formula,  $\lambda R = \lambda_s R_s$ , where  $\lambda$  is the mean arrival rate for new customers,  $\lambda_s$  is the mean service rate,  $R$  is the average number of unsuccessful retrials made by a customer, and  $R_s$  is the average number of unsuccessful retrials made during one service period. This formula is reminiscent of the so-called Little's Formula ( $L = \lambda W$ )<sup>2</sup>, though its intuitive

---

<sup>2</sup>The intuitive basis of Kulkarni's formula comes from recognizing that each unsuccessful retrial belongs to some customer and takes place during some service period (according to the model, an unsuccessful retrial cannot take place when the server is idle). Little's formula as stated, with  $L$  being the mean total number of customers in the system at steady state, including customers in service,  $\lambda$  the mean arrival rate, and  $W$  the mean time a customer spends in the system including time in service, applies with great generality to a wide variety of steady state queues and is one of the most celebrated and discussed results in queueing theory. It too has a simple intuitive base (the mean number of customers in the system must be the mean number of customers that arrive during the time a single customer spends in the queue and in service). Kulkarni's proof of his formula is modelled on a proof developed by Stidham [1972, 1974] for Little's formula, and Kulkarni claims the same sort of generality for his formula relative to retrial systems as is the case for Little's formula for non-retrial queues.

plausibility belies the intricacy of the analysis required in a rigorous proof. This is another feature shared by the two formulas.

Kulkarni applies his formula to the steady state analysis of a retrial queue involving two types of customers having exponentially distributed arrival, retrial and service times, but with differing mean values. He obtains expressions for the expected waiting time for each type of customer in the system and the expected number of each type of customer in the system.

One of the interesting special cases to which Kulkarni's results can be applied is a situation in which one customer decides to conduct retrials at a different average rate than all other customers in the system. Kulkarni presents a formula for the expected waiting time (before service) of such a customer. One would expect that a detailed understanding of the characteristics of such a result would have some very useful applications.<sup>3</sup>

### Conclusions

Though a number of the works cited in this section contain derivations of equations governing the time-dependent behaviour of single-server or multiple-server retrial queues (for example, Keilson, Cozzolini and Young [1968]), the resultant equations invariably do not have simple closed form solutions. Because

---

<sup>3</sup>This result of Kulkarni's would seem to be a useful starting point for developing strategies for a customer who finds himself in orbit, and would like to minimize his expected waiting time in orbit. It would provide some way of estimating the value of 'keen-ness' in such a customer.

of this, major results are presented only for the much more tractable steady-state case for such queues.

This twenty-five or more years of work on the steady-state retrial queue has led to some very sophisticated results. However, it is not difficult to think of actual systems corresponding to retrial queues to which these steady-state results do not apply because arrival, retrial, and/or service rates vary in time to the extent that a steady-state regime cannot be established.

Because the main objective in this thesis is to examine some viable approaches to determining the time-dependent behaviour of a single-server retrial queue, it is unnecessary to discuss any further the results described briefly in this section for the steady-state retrial queue. However, we note that these steady-state results could bear a more detailed examination, since there appear to be some discrepancies outstanding between the forms of results obtained by different authors. One example is the relatively simple expression Keilson, Cozzolini and Young [1968] obtain for the mean waiting time in orbit of a customer, which contrasts markedly with, for example, the complexity of the corresponding result obtained by Kulkarni [1982].

### 1.3 Time-Dependence of Queues

*Anderson's Law: 'I have yet to see any problem, however complicated, which, when you look at it in the right way, did not become still more complicated.'*

(The Official Rules, Paul Dickson, ed.)

#### Why Bother With Explicit Time Dependence?

The main part of this thesis is concerned with formulating and testing a particular set of approximation methods for studying the time-dependence of the single server retrial queue. In this section, the motivation for these studies is developed by discussing the importance of considering time-dependent behaviour of non-stationary queues, by noting the inapplicability of various traditional approaches to solving the relevant differential equations or systems of differential equations, and finally, by noting the computational problems encountered using standard numerical approaches to the problem.

For some years it has been recognized that there were many applications of queueing theory for which the steady state solutions were inappropriate in principle if not in practice. The problem was not so much a need to take account of the transient time-varying regime as systems evolve from some initial state to a steady state, but rather that many systems have intrinsic non-stationarity because arrival rates and/or service rates vary with time. Further, this variation in time may not (in fact, likely will not) be as simple low order polynomials in  $t$ , but more likely as a periodic function involving trigonometric functions.

### Koopman's Work

Koopman [1972] pointed out that the queue formed by aircraft wishing to take-off or land at an airport had a definite 24-hour periodicity. Within each 24-hour period, demand for access to runways varied from zero during the early morning hours to a level during rush hours which could for some airports be well in excess of the allowable runway capacity. To model such a system as a steady state queue is clearly inappropriate.

To deal with such a significant time variation in the characteristics of the queue, Koopman first noted that the smallest time interval over which a change in the state of the queue could occur was strictly positive. Since an aircraft takes of the order of one or two minutes to enter the runway and leave it, either during take-off or landing, changes in the state of the queue can only occur on that sort of time scale, and so the whole problem is discrete.

Secondly, he noted that for even the largest airports, the number of aircraft that can be waiting on the ground to take-off, or the number which can be waiting in the vicinity to land is necessarily quite small. Thus one can realistically assume that there is a specific finite number (maybe a few dozen, and certainly not more than a few hundred), with the probability of the number of aircraft in the system ever exceeding that number being zero in practice. As a result, the infinite system of linear differential equations analogous to (1.1.3a) and (1.1.3b) above which needs to be solved to describe the time-dependent behaviour of the system can be truncated to a finite system. The solution of the equations

using numerical methods then becomes feasible.

Though Koopman was concerned primarily with the repercussions his computational results had for airport design and operation, and though he considered a number of other issues<sup>1</sup> in the application of queueing models to systems like this, the main impact of his paper seems to have been to establish the necessity of taking time dependence of queues into account explicitly in modelling many real systems. His quantitative results showed that quantities like the expected queue length and its standard deviation, the probabilities of finding the queue empty or finding it full, the cumulative number of customers turned away from a full queue, and the expected waiting time for a customer entering the queue, varied appreciably with variations in arrival and service rates.

#### Solution by Power Series Expansion

There are several straightforward approaches to determining analytic solutions for systems of linear differential equations like (1.1.3a) and (1.1.3b). Power series

---

<sup>1</sup>Koopman also considered the issue of choosing an appropriate probability distribution for the service times. He carried out a number of calculations using first the exponential distribution (corresponding to minimum information regarding the time of the next service completion) and then assuming constant service times (corresponding to maximum information about the time of the next service completion). He found that the quantitative results were relatively insensitive to which of these service time distributions was used. Considering these alternative models to represent extremes, he concluded that it was an acceptable approximation to adopt the exponential distribution to take advantage of simplifications in the formalism, in place of a more realistic but likely much more complicated distribution. Koopman's devotion to realism is further demonstrated by his use of actual airport records to evaluate his computed results.

solutions are sometimes useful. In this case, we can write<sup>2</sup>

$$p_{on}(t) = \sum_{j=0}^{\infty} a_j^{(n)} t^j, \quad p_{ln}(t) = \sum_{j=0}^{\infty} b_j^{(n)} t^j, \quad (1)$$

$$n = 0, 1, 2, \dots$$

Substituting these two expansions into equations (1.1.3a) and (1.1.3b) and requiring that the coefficients of a given power of  $t$  on either side of each equation be identical, we obtain

$$a_{j+1}^{(n)} = \frac{\mu b_j^{(n)} - (\lambda + n\nu) a_j^{(n)}}{j+1} \quad (3a)$$

and

$$b_{j+1}^{(n)} = \frac{\lambda b_j^{(n-1)} + \lambda a_j^{(n)} + (n+1)\nu a_j^{(n+1)} - (\lambda + \mu) b_j^{(n)}}{j+1} \quad (3b)$$

When  $n = 0$ , the first term in the numerator of (3b) must be dropped. These formulas express coefficients with lower index ' $j+1$ ' completely in terms of

---

<sup>2</sup>Carter and Cooper [1972] employ a power series approach in a study of the stationary waiting time distributions for various queues. They factor a factorial out of the coefficients in the series. In our case, that approach would make formulas (1) look like

$$p_{on}(t) = \sum_{j=0}^{\infty} \frac{c_j^{(n)}}{j!} t^j, \quad p_{ln}(t) = \sum_{j=0}^{\infty} \frac{d_j^{(n)}}{j!} t^j. \quad (2)$$

The coefficients  $c_j^{(n)}$  and  $d_j^{(n)}$  are then given by equations (2a) and (2b) respectively, but with the denominators ' $j+1$ ' dropped. One suspects that this approach will lead to difficulties in calculating factorial-sized quantities explicitly when large values of  $j$  are handled.

coefficients with lower indices 'j', and so form a feasible basis for computing these coefficients recursively. The above formulas apply when  $\lambda$ ,  $\mu$ , and  $\nu$  are all constants, and the situation would certainly become more complicated if any or all of these quantities were time-dependent. One can readily imagine the additional complexity involved in separating powers of  $t$  if, for example, arrival rates were sinusoidal.

Of course, the real problem with this approach is that even formally convergent power series expansions form hazardous bases for computation except when the value of the independent variable is very small. In this case, the numerical sum of these series will always be a number between zero and one (since the sum represents a probability), and yet, for values of  $t$  much greater than one, the small sum may be constructed out of terms having magnitudes much greater than the final sum, and so serious round-off error can occur in summing the series as a result of the finite precision with which all computing machinery stores numbers.

For example, when  $\lambda = 0.5$ ,  $\mu = 0.7$ , and  $\nu = 0.3$  and all constant, the probability  $p_{00}(20)$  is very near its steady state value of 0.035. Yet in evaluating this quantity using the power series expansions above, one finds that more than twenty terms in the sum including those involving  $t^{40}$  through  $t^{60}$  are all greater in magnitude than  $10^{-12}$ , and alternate signs in an irregular fashion. This means that something like the fifteen lowest order significant figures of the computed series sum will be garbage. Since the highest precision (in the computerese sense) normally used in scientific calculations is around 15 or 16 significant

figures, the inappropriateness of this approach to computing these probabilities is clear.<sup>3</sup>

### Determination of Time-Dependent Probability Generating Functions

An alternative approach to solving equations (1.1.3a) and (1.1.3b) exactly is to transform each set of equations into a single partial differential equation for the corresponding probability generating function. All probabilities and moments of the system can be computed straightforwardly from the probability generating functions.

For the retrial queue, we define the two generating functions,

$$\underline{P}_0(z, t) = \sum_{n=0}^{\infty} p_{0n}(t) z^n, \quad \underline{P}_1(z, t) = \sum_{n=0}^{\infty} p_{1n}(t) z^n. \quad (4)$$

Multiplying equations (1.1.3a) and (1.1.3b) each as written by  $z^n$ , summing the resultant equations over  $n$ , and rearranging terms appropriately, we obtain

$$\begin{aligned} \frac{\partial}{\partial t} \underline{P}_0(z, t) &= -\lambda \underline{P}_0(z, t) + \mu \underline{P}_1(z, t) - \nu z \frac{\partial}{\partial z} \underline{P}_0(z, t), \\ \frac{\partial}{\partial t} \underline{P}_1(z, t) &= -(\lambda + \mu) \underline{P}_1(z, t) + \lambda z \underline{P}_1(z, t) + \lambda \underline{P}_0(z, t) + \frac{\partial}{\partial z} \underline{P}_0(z, t). \end{aligned} \quad (5)$$

---

<sup>3</sup>The basis of this rough quantitative estimate of the effect of numerical roundoff error on the accuracy of the computed sum of these power series is somewhat simplistic. More sophisticated analyses of the problem are available (see, for example, Lyashenko and Nikulin [1986]). Such analyses appear to indicate that the true cumulative round-off error in computed sums can be significant even when there are not such large differences in the magnitudes of the individual summands.

So, the doubly infinite system of ordinary differential equations has been replaced by a system of two partial differential equations. These equations can be solved directly (see Zachmanoglou [1976], chapter 10), but that approach seems to result in a system of two integral equations which are solvable only by iteration, and the complexity of the expressions arising out of the iterative approach may make attainment of adequate accuracy infeasible.<sup>4</sup>

---

<sup>4</sup>Following the method of characteristics as outlined by Zachmanoglou [1976], we obtain:

$$v_0(z, t) = \frac{e^{-\nu t}}{z} - \int_0^t \left\{ \left[ -\nu + \lambda - \frac{\mu}{ze^{-\nu(\bar{t}-t)}} \right] v_0(z_0(\bar{t}), \bar{t}) + \frac{\mu}{ze^{-\nu(\bar{t}-t)}} v_1(z_0(\bar{t}), \bar{t}) \right\} d\bar{t},$$

$$v_1(z, t) = \frac{1}{z} - \int_0^t \left\{ \left[ 2\lambda - \frac{\mu}{z} + \mu \right] v_0(z_1(\bar{t}), \bar{t}) + \left[ \frac{\mu}{z} - (\lambda + \mu - \lambda z) \right] v_1(z_1(\bar{t}), \bar{t}) \right\} d\bar{t}.$$

where

$$z_0(\bar{t}) = ze^{-\nu(\bar{t}-t)}, \quad z_1(\bar{t}) = z$$

are the equations of the characteristics of equations (5) passing through the point

### Transform Methods

Alternatively, one could attempt to solve equations (5) using Laplace transforms,

$$\bar{P}_0(z, s) = \int_0^\infty \underline{P}_0(z, t) e^{-st} dt, \quad \bar{P}_1(z, s) = \int_0^\infty \underline{P}_1(z, t) e^{-st} dt, \quad (6)$$

which, on substitution into (5), gives the two equations

$$s\bar{P}_0(z, s) - \underline{P}_0(z, 0) = -\lambda\bar{P}_0(z, s) + \mu\bar{P}_1(z, s) - \nu z \frac{\partial}{\partial z} \bar{P}_0(z, s) \quad (7)$$

and

$$s\bar{P}_1(z, s) - \underline{P}_1(z, 0) = -(\lambda + \mu - \lambda z)\bar{P}_1(z, s) + \lambda\bar{P}_0(z, s) + \nu \frac{\partial}{\partial z} \bar{P}_0(z, s),$$

if we assume  $\lambda$ ,  $\mu$ , and  $\nu$ , are independent of time.

In principle, we could solve the first equation for  $\bar{P}_1(z, s)$ , substitute this into the second equation, and obtain a single first order partial differential equation for  $\bar{P}_0(z, s)$ . At best the resultant equation is of daunting complexity. This

---

(z,t). Here,

$$\underline{P}_0(z, t) = z v_0(z, t) \quad \text{and} \quad \underline{P}_1(z, t) = v_1(z, t) - v_0(z, t).$$

It would appear possible to evaluate the necessary integrals analytically if an iterative method was employed to solve these equations, even if some or all of  $\lambda$ ,  $\mu$ , and  $\nu$  are non-constant, though the amount of work required may increase rapidly with successive iterations. An as yet unresolved problem here is that in employing the successive approximation approach described in Zachmanoglou [1976], we seem to get an expression for  $\underline{P}_1(z, t)$  involving negative powers of  $z$ , which contradicts equation (4).

approach is not likely to be feasible when any or all of  $\lambda$ ,  $\mu$ , or  $\nu$ , are functions of time, since then equations (7) would become even more complicated, and potentially involve higher order derivatives of the Laplace transforms.

### Previous Work on Time-Dependent Queues

Even though comparatively little attention has been given to the time-dependent behaviour of queues in the literature, the number of papers which have appeared on the subject is quite large. No attempt will be made here to give a comprehensive review of the entire history of the subject. However, it is worthwhile to note the types of problems which have been solved successfully, and the sorts of approaches used.

Among the early works on the subject was that of Morse [1958], who was able to obtain the time-dependent probabilities for the M/M/1 queue with constant service and arrival rates using the probability generating function approach. These analytic solutions were very complicated, involving sums over terms containing Bessel functions. Later, Saaty [1961] succeeded in obtaining the Laplace transforms of the time-dependent probabilities for the M/M/s queue with constant arrival and service rates, but was able to invert the transforms only for the two-server case. In both works, the time-variation of the probabilities was a transient effect as the systems 'decayed' to their steady state. Similar approaches would not have produced corresponding solutions had the arrival and/or service rates themselves been functions of time.

With the failure of frontal attack for more complicated queue models, an alternative approach was developed by Newell [1968a, b, c], with extensions more recently by Keller [1982] and Massey [1985]. Newell was interested in modelling the behaviour of a traffic system going through a rush hour during which the 'arrival rate' significantly exceeded the capacity of the system. Since the length of a queue experiencing more arrivals than departures on average will grow without limit, a steady state cannot be established. Newell recognized that in such a saturated queue, to a good approximation one could consider arrivals and departures to be a continuous process because the time intervals between successive such events are very small. This approach led to a kind of diffusion equation for the time-dependent distribution function for queue length. That equation could not be solved generally, but its form facilitated the drawing of various conclusions about the behaviour of the system.

Keller [1982] employed a formal asymptotic analysis amounting to adopting a time scale over which changes in arrival and service rates are very small. Keller found that his analysis reproduced many of Newell's results. Massey has apparently put Keller's work on a more solid theoretical foundation, particularly in proving that several of Keller's results were in fact asymptotic to the exact solutions. These analyses are quite complicated, and neither Keller nor Massey attempt to compare the results of their approach with exact numerical solutions. All three authors deal with some variant of the M/M/1 queue, though Keller states that his method of analysis applies to other cases as well. To extend this approximation approach to the retrial system would be quite a major undertaking.

Kelton and Law [1985] have restructured the M/M/s problem in terms of the probabilities:

$$P_k(n,i) = \Pr\{X_n = i \mid k \text{ customers were already present at } t = 0\},$$

where  $X_n$  is the number of customers in the system at the instant the  $n^{th}$  'new' customer arrives, including the new arrival. Thus we are looking at the system at a monotonely increasing sequence of random, discrete times, rather than in a continuous time framework. Using this approach, Kelton and Law are able to compute the expected delay in the queue for each arriving customer exactly. This approach is worth examining for the retrial queue, but that will not be done here because our immediate goal is to produce a viable approach to describing the continuous time behaviour of the system.

### Numerical Approaches

As far as being able to predict the continuous-time behaviour of actual single server retrial queues, there seems to be little hope of finding computationally feasible exact analytic solutions. Turning attention to strictly numerical methods, we find that while they introduce difficulties of their own, these can be minimized at a price. None of the factors of the retrial queueing problem which dash hopes of an exact analytic solution are of any great consequence as far as implementation of numerical methods are concerned. If we adopt a specific

numerical approach to solve the systems (1.1.3a,b), the algorithm will simply require us to be able to supply numerical values of  $\lambda$ ,  $\mu$ , and  $\nu$ , for any specified value of  $t$ . Whether these parameters are constants throughout or are generated by complicated formulas involving transcendental functions or worse, is effectively irrelevant to the structure of the numerical algorithm.<sup>5</sup>

The available algorithms for the numerical solution of ordinary differential equations easily number in the hundreds if not thousands. For the work discussed following this section, a modification of the Runge-Kutta-Felberg-45 or RKF45 method [Felberg, 1970] based on Algorithm 6.3 of Burden, Faries, and Reynolds [1978] was used exclusively (see Figure 1.3). This method generates both a so-called fourth order estimate and a fifth-order estimate of the solution efficiently at the end of each interval in  $t$ . There are good reasons for considering the difference between these two solution estimates as an acceptable estimate of the numerical error in the fourth-order estimate (see Burden, Faries, and Reynolds [1978], section 6.5). Since the fifth order estimate is the one retained, this difference between the two estimates is generally a very pessimistic estimate of the actual numerical error in the adopted solution.

The main reason for choosing this method was that it not only produces an estimate of the numerical error being incurred, but it can be implemented so that

---

<sup>5</sup>Most numerical methods for integrating systems of differential equations are based on some combination of polynomial extrapolation and interpolation. Changing the functional form of various coefficients in the equations would not require modifying the structure of the algorithm, though it would likely affect the detailed performance of the algorithm. As these coefficients, and so the solution, become less like a smooth low-order polynomial, the polynomial extrapolation/interpolation approach may become less effective or efficient.

### FIGURE 1.3 The Runge-Kutta-Fehlberg-45 Algorithm

The following is the implementation of a Runge-Kutta method as developed by Fehlberg (see Burton, Fairies, and Reynolds [1978], section 6.3) as used exclusively to solve systems of differential equations for the calculations presented in this thesis. The algorithm produces a fourth order and a fifth order estimate of the solution, requiring a total of only six derivative evaluations per internal step. The difference between the two estimates is taken as an upper limit on the error in the fifth order estimate, and internal step lengths are adjusted automatically to meet a preset tolerance on this error limit.

#### INITIALIZATION:

The algorithm computes an approximation to  $y(b) = \{y_1(b), y_2(b), \dots, y_n(b)\}$  with  $n$  an integer greater than or equal to 1, given a function subprogram which computes

$$y'(t) = \{y_1'(t), y_2'(t), \dots, y_n'(t)\} = \{f_1(t), f_2(t), \dots, f_n(t)\} \\ = f(t)$$

and the initial values

$$y(a) = \{y_1(a), y_2(a), \dots, y_n(a)\}.$$

It is also necessary to specify

$\epsilon$  = the desired maximum absolute difference ( $\epsilon > 0$ ) between the generated fourth and fifth order estimates of  $y_i(t)$ ,  $i = 1, 2, \dots, n$ , over the range  $a \leq t \leq b$ .

STEP 1: Estimate the initial internal step length:  $h = (\epsilon)^{1/4}$   
Set  $t_0 = a$ .  
Set  $(w_0)_i = y_i(a)$ ,  $i = 1, 2, \dots, n$ .

STEP 2: Set  $t = t_0 + h$ .  
If  $t > b$ , reset  $t = b$ ,  $h = b - t_0$ .

STEP 3a: Set  $(k_1)_i = h \cdot f_i(t_0, w_0)$ ,  $i = 1, 2, \dots, n$ .

STEP 3b: Set  $(k_2)_i = h \cdot f_i(t_0 + \frac{1}{4}h, w_0 + \frac{1}{4}k_1)$ ,  $i = 1, 2, \dots, n$ .

STEP 3c: Set  $(k_3)_i = h \cdot f_i(t_0 + \frac{3}{8}h, w_0 + \frac{3}{32}k_1 + \frac{9}{32}k_2)$ ,  $i = 1, 2, \dots, n$ .

STEP 3d: Set  $(k_4)_i = h \cdot f_i(t_0 + \frac{12}{13}h, w_0 + \frac{1932}{2197}k_1 - \frac{7200}{2197}k_2 + \frac{7296}{2197}k_3)$ ,

$i = 1, 2, \dots, n$ .

Figure 1.3 continued.....

$$\text{STEP 3e: Set } (k_5)_i = h \cdot f_i(t_o + h, w_o + \frac{439}{216}k_1 - 8k_2 + \frac{3680}{513}k_3 - \frac{845}{4104}k_4),$$

$$i = 1, 2, \dots, n.$$

$$\text{STEP 3f: Set } (k_6)_i = h \cdot f_i(t_o + \frac{1}{2}h, w_o - \frac{8}{27}k_1 + 2k_2 - \frac{3544}{2565}k_3 + \frac{1859}{4104}k_4 - \frac{11}{40}k_5), \quad i = 1, 2, \dots, n.$$

$$\text{STEP 3g: Set } w_4 = w_o + \frac{25}{216}k_1 + \frac{1408}{2565}k_2 + \frac{2197}{4104}k_4 - \frac{1}{5}k_5.$$

$$\text{STEP 3h: Set } w_5 = w_o + \frac{16}{135}k_1 + \frac{6656}{12825}k_3 + \frac{28561}{56430}k_4 - \frac{9}{50}k_5 + \frac{2}{55}k_6.$$

$$\text{STEP 3i: Compute } r = \max_i \{ |(w_5)_i - (w_4)_i|, i = 1, 2, \dots, n \} / h$$

STEP 4: Update internal step length:

$$\text{Compute } \delta = 0.84 (\epsilon / r)^{1/4}$$

If  $r$  is effectively zero (say  $r < 10^{-14}$ ), go to step 6.

If  $\delta < 0.1$ , replace  $h$  by  $0.1h$ .

If  $\delta > 4.0$ , replace  $h$  by  $4h$ .

If  $0.1 < \delta < 4.0$ , replace  $h$  by  $\delta h$ .

STEP 5: If  $r > \epsilon$ , return to step 2 (the fifth order estimate  $w_5$  has been rejected).

STEP 6: We have  $r \leq \epsilon$ , therefore the current fifth order estimate at  $t$  is accepted.

If  $t = b$ , go to step 7.

Copy  $w_5$  into  $w_o$ .

Set  $t_o$  equal to  $t$ .

Return to step 2.

STEP 7: The numbers  $(w_5)_i$ ,  $i = 1, 2, \dots, n$ , estimate  $y_i(b)$ ,  $i = 1, 2, \dots, n$ , with local error estimated to be less than  $\epsilon$  (the sum of the values of  $r$  computed for each internal step gives a rough estimate of the global error over the interval  $[a, b]$ ).

internal step sizes are adjusted automatically to keep this numerical error below a pre-set threshold. This is a very important feature since we wish to be able to control numerical errors entering computed results so that discrepancies between results obtained from the exact equations (1.1.3a,b) and our approximations can be reliably attributed to the nature of the approximation rather than to carelessness in handling the equations.<sup>6</sup>

Questions of efficiency of implementation and speed of computation are important, but secondary to the objectives here, so no studies have been done to determine how the RKF45 algorithm performs compared to other types of algorithms.

### Truncation of Systems of Differential Equations

No numerical algorithm can cope with an infinite system of differential equations, and so the system (1.1.3a,b) had to be truncated. In practice, this truncation is done fairly arbitrarily. As calculations reported in this thesis were carried out, the sum of the computed probabilities was subtracted from one at each point where the solution was computed to give an estimate of the residual probability 'lost' due to truncation. If that residual probability became greater than  $1.0 \times 10^{-6}$ , the calculation was repeated with the system of equations

---

<sup>6</sup>In the calculations reported in this thesis, this error threshold was taken to be 0.00001 throughout on the probabilities  $p_{\xi_n}(t)$ . Except in rare cases when the expected number of customers in the system can become very large, this ensures that the computed first and second moments at least should be accurate to three or four significant digits. Of course, it must be remembered that this error estimate threshold is an upper limit, and the actual errors incurred may be very much smaller than this.

truncated at a larger dimension. It would not be difficult to implement the algorithm to permit the dimension of the truncated system to be adjusted from step to step automatically in order to keep the residual probability below a pre-set value, but this was not done.

Also, at each step, the residual probability was added into the highest index probability computed (or pro rata into the two highest index 'n' probabilities for the retrial queue) to reduce system truncation errors in computed moments, and the buildup of these errors from step to step as the complete specified range of time was covered.

Truncation of the system to a finite size is justified on the basis that the equations dropped are effectively all identities,  $0 = 0$ . In order for this to be true, we must retain equations for all states of the system which have a probability greater than some criterion of negligibility.

A problem with the retrial system is that for a given utilization of the queue, the number of equations is roughly twice or more the number which would have to be included for the simple M/M/1-type queue, since the probabilities are indexed both by the number of customers in the system and by whether the server is idle or busy. Also, we can have a relatively lightly utilized retrial system ( $\rho = \lambda/\mu$  significantly smaller than 1) but if the retrial rate,  $\nu$ , is also small, there can be significant probabilities of finding larger numbers of customers in the system and so the dimension at which (1.1.3a,b) can be truncated has to be correspondingly greater.

Steady-state probabilities illustrate this problem. The  $p_n$  for the M/M/1 queue and the  $p_{0n}$  and  $p_{1n}$  for the M/M/1/1 retrial queue normally exhibit a maximum value for some value of the index  $n$  and then decrease with increasing  $n$ , eventually to zero. A relative indication of the necessary dimensions of truncated systems of the differential equations can be obtained by noting the number of elements of these sequences of steady-state probabilities with value at or above a specific threshold  $\epsilon$ .

For example, take the "negligibility" threshold  $\epsilon$  as  $1.0 \times 10^{-6}$ . When  $\rho = 0.5$  and  $c = 1$ , we find that  $p_n < \epsilon$  for  $n > 19$ , whereas  $p_{0n} < \epsilon$  for  $n > 26$  and  $p_{1n} < \epsilon$  for  $n > 20$ . So, while only the first 18 steady-state probabilities for the M/M/1 queue are non-negligible by this criterion, there are  $25 + 19 = 44$  non-negligible steady-state probabilities for the corresponding retrial queue. Using  $\rho = 0.5$  and  $c = 10$ , the number of non-negligible steady-state probabilities for the retrial queue is 86 by this criterion (and is still 18 for the M/M/1 queue, since the  $p_n$  depend only on  $\rho$ ).

Alternatively, as  $c$  gets nearer to zero (indicating retrial rates are much greater than new customer arrival rates), the number of non-negligible  $p_{0n}$ 's for the retrial queue becomes similar to the number of non-negligible  $p_n$ 's for the simple M/M/1 queue, and the number of non-negligible  $p_{1n}$ 's approaches zero. For example, when  $\rho = 0.5$  and  $c = 0.1$ , there are 18 non-negligible  $p_{0n}$  and 10 non-negligible  $p_{1n}$  for a total of 28 non-negligible steady-state probabilities for this retrial queue (compared to the total of 18 for the corresponding simple

M/M/1 queue).

Of course, for time dependent calculations, the truncated systems may have to have higher dimensions than indicated here to ensure that at times before a steady-state is approached, no non-negligible probabilities are excluded from the calculation. However, this rough analysis gives an indication of the differences in dimension of the system of differential equations which must be solved in computing the time-dependent behaviour of the two queues starting with systems (1.1.3a) and (1.1.3b).

This problem is of significance in computing moments of the distribution, since errors in high index probabilities (including the 'error' of neglecting them) are magnified. Thus, for retrial queues, we will find that it is necessary to truncate the system of differential equations to be solved numerically at dimensions two or three or more times the size as would be necessary for a simple M/M/1 queue. In the calculations reported in this thesis, necessary dimensions for truncations of the systems (1.1.3a,b) were typically in the range of 50-100 though a few calculations required handling systems of 130 equations to attain the desired accuracy.<sup>7</sup> Details are given in Appendix A.

---

<sup>7</sup>Again, the actual computed results were considered more important than how fast or efficiently they were computed. Programs were written in FORTRAN and calculations were carried out on an IBM mainframe using the WATFIV compiler for convenience. With the numerical error threshold set at 0.00001 over a step of length  $\Delta t = 0.5$ , of the order of 0.1 to 1 second of cpu time was required for integration of systems of equations with dimensions in the 50 - 100 range. On an IBM PC/XT without a numerical coprocessor and using the Microsoft FORTRAN compiler, execution times for calculations like these would require of the order of one to ten minutes per step of length  $\Delta t = 0.5$ . These numbers are of order of magnitude only since no attempt was made to write FORTRAN code optimized for execution speed, and since the algorithm will break

It is noted that there are multi-step methods for the numerical solutions of systems of differential equations which also permit dynamic step length/error control (see for example, Algorithm 6.5 in Burden, Faries and Reynolds [1978], based on a fourth order predictor-corrector approach). These algorithms require a single-step routine like the Runge-Kutta methods to start them off, and they would be worthwhile evaluating if a search was being made for the most efficient algorithms. Similarly there are higher order single-step methods (see, for example, Shanks [1966], who gives several eighth-order Runge-Kutta methods). These higher order formulas in principle give greater accuracy for a given step length, or, more importantly, they allow longer step lengths for given error tolerances. Higher order formulas are mainly used to avoid unacceptable round-off error due to the finite precision with which a digital computer stores numbers,<sup>8</sup> which grows as the step length gets shorter.

---

an interval of length  $\Delta t = 0.5$  up into the required number of smaller subintervals to meet the error tolerance. The number of such subintervals required may vary widely with the value of  $t$  for certain functional forms of  $\lambda(t)$ ,  $\mu(t)$ , and  $\nu(t)$ .

<sup>8</sup>As distinguished from what we have been calling the numerical error, which is due to the approximation of the function of interest by a low order polynomial over small intervals.

#### 1.4 Closure Approximations for M/M/s Queues

*Snafu Law #1: Given any problem containing  $n$  equations, there will be  $n+1$  unknowns.*

(1001 Logical Laws..., Peers & Bennett)

##### Details of the Rothkopf and Oren Closure Method

The rationale behind the use of so-called closure techniques to the approximation of solutions of infinite systems of differential equations is well described in Rothkopf and Oren [1979]. The essential feature in applying such a technique in a queueing context or in any other context is to make one or more assumptions about the form of the functional relationship between various quantities with the result that an infinite system of equations can be reduced to a small number of equations determining a few quantities of direct interest.

We illustrate this approach as applied by Rothkopf and Oren to the M/M/1 queue (with contributions by earlier authors as cited in their paper). Defining  $p_n(t)$  as the probability that there will be  $n$  customers in the M/M/1/ $\infty$  system at time  $t$ , including any customer in service at that time, we can proceed much as in section 1.1 above to obtain the equations

$$\frac{d}{dt} p_0(t) = -\lambda p_0(t) + \mu p_1(t) \quad (1)$$

$$\frac{d}{dt} p_n(t) = -(\lambda + \mu) p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t) \\ n = 1, 2, 3, \dots$$

Here,  $\lambda$  is the average arrival rate and  $\mu$  the average service rate. In order to compute any characteristic of the system, such as the mean or the variance of

the number of customers at time  $t$ , it is necessary in principle to solve the infinite system (1). The difficulties to be encountered have been described in detail in the preceding section of this thesis.

What Rothkopf and Oren propose is the following. If each of the equations in (1) is multiplied by  $n$ , then all equations summed over  $n$ , we get first

$$\sum_{n=0}^{\infty} \frac{d}{dt} \{n \cdot p_n(t)\} = -(\lambda + \mu) \sum_{n=0}^{\infty} n \cdot p_n(t) + \lambda \sum_{n=0}^{\infty} n \cdot p_{n-1}(t) + \mu \sum_{n=0}^{\infty} n \cdot p_{n+1}(t)$$

With some rearrangement, this can be written as

$$\frac{d}{dt} m(t) = \lambda - \mu \{1 - p_0(t)\} \quad (2a)$$

where  $m(t) = \sum_{n=0}^{\infty} n p_n(t)$  is the mean number of customers in the system at time  $t$ . In a similar way, the equations can be multiplied by  $n^2$ , summed, and using the fact that the variance of the number of customers in the system  $= v(t) = E(n^2) - m^2$ , so that

$$\frac{d}{dt} v(t) = \frac{d}{dt} E(n^2) - 2m \frac{dm}{dt},$$

we get

$$\frac{d}{dt} v(t) = \lambda + \mu - \mu p_0(t) \{2m(t) + 1\} \quad (2b)$$

Equations (2a) and (2b) contain three unknown functions:  $p_0(t)$ ,  $m(t)$  and  $v(t)$ .

There is no obvious way to eliminate one of the three exactly.

Rothkopf and Oren then proposed an approximation to eliminate  $p_0(t)$  in terms

of  $m(t)$  and  $v(t)$ , so that equations (2a) and (2b) would become a system of two differential equations in these two unknown functions only. This involves introducing a so-called "closure assumption", in the form of an approximate relation (or "closure formula") between the three quantities. We will use the terms closure assumption and closure formula interchangeably.

Rothkopf and Oren noted that the steady-state probabilities for the  $M/M/1/\infty$  system were a geometric distribution, which was a special case of the negative binomial or Pascal distribution. They noted further that for the Pascal distribution, one could write

$$p_0 = \left( \frac{m}{v} \right)^{m^2 / (v - m)} \quad (3)$$

They then adopt this functional relationship to eliminate  $p_0(t)$  from equations (2a) and (2b).<sup>1</sup> The choice of this functional relationship is prompted and justified in part by the resultant equations correctly reproducing the steady-state values of  $m$  and  $v$  when the derivatives in (2a,b) are replaced by zeros. However, as Rothkopf and Oren note, "the key to the success of a closure method is its closure assumption, while the test of any such assumption is the empirical usefulness of the resulting equation."

The success of the approximation will depend on how close to a negative binomial distribution the  $p_n(t)$  are over all finite values of  $t$ . Intuitively, we

---

<sup>1</sup>A modification of this formula is given for use in the case that  $v=m$  -- this usually only occurs at the beginning of a calculation when the system is assumed to be empty so that  $m(0) = v(0) = 0$ .

would expect an approach like this to perform most poorly when at least some of the probabilities,  $p_n(t)$ , are changing rapidly with time, since then it is plausible that the deviation from steady-state relationships between these quantities is likely to be greatest.

### Numerical Results

The largest part of Rothkopf and Oren [1979] consists of the presentation and discussion of various computed results. Their assessment of this closure approximation as giving an acceptably accurate practical computational method is generally favorable. However, the results they present are mainly graphical and so cannot be subjected to detailed quantitative analysis.

A set of standard calculations were performed here, to which corresponding results for retrial queues will later be compared. For one set of calculations, the mean arrival and service rates were assumed constant, so that over time, the closure solutions obtained should describe the transient decay of the system to steady state. At the other extreme, arrival rates were assumed to have the functional form

$$\lambda(t) = \lambda_0 \left\{ 1.1 + \sin\left(\frac{2\pi t}{P}\right) \right\} \quad . \quad (4)$$

where  $\lambda_0$  is a constant and  $P$  represents the period of oscillation of the arrival rate (taken to be 10 time units throughout this thesis). We will refer to the first type of calculations as constant parameter calculations, and to the second type as

oscillatory parameter calculations. The number '1.1' was used in (4) to avoid  $\lambda(t)$  becoming zero. This seemed to cause numerical algorithm problems. For both types of calculations, the service rates were held constant in time.

The RKF45 algorithm was used to solve all systems of differential equations here to ensure that the discrepancies found between the quantities computed from the effectively exact solutions of the system (1) and those obtained by solving (2a,b) with (3) are due to errors in the closure approximation, and not due in any significant way to the numerical approximation of the solutions of the equations.

Rothkopf and Oren used the simple forward Euler method for the work reported in their paper. While the adequacy of this simplest of all algorithms may not be strenuously questioned for very well-behaved systems of only two differential equations at the level of accuracy required, it is not necessary to adopt an algorithm so prone to numerical error, given the power of commonly accessible computing machinery these days.

For the constant parameter case, graphs depicting the relative errors<sup>2</sup> in  $p_0(t)$ ,  $m(t)$  and  $v(t)$  respectively as calculated by the closure approximation are given in Figures 1.4, 1.5, and 1.6. Calculations were done for  $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$ , and 0.95 (where  $\rho = \lambda/\mu$  is a measure of the mean utilization of the server) to compare the accuracy of the closure approximation for queues approaching steady-state with a range of characteristics. Only the results for the three largest

---

<sup>2</sup>Throughout this thesis, errors are always computed as 'approximate - exact', and relative errors are that quantity divided by the exact value of the quantity being considered.

values of  $\rho$  are graphed to avoid clutter. In all calculations, it was assumed that the system was empty at  $t=0$  (that is,  $p_0(0)=1$ ,  $m(0)=0$ , and  $v(0)=0$ ) and in all cases, the closure approximation solution converged to the steady-state solution

$$p_0(\infty) = 1 - \rho, \quad m(\infty) = \frac{\rho}{1 - \rho}, \quad v(\infty) = \frac{\rho}{(1 - \rho)^2}, \quad (5)$$

monotonically or with small oscillations about a monotonically converging path.

As can be seen from Figures 1.4 - 1.6, the relative errors in each of the three quantities computed with this closure approximation start out at zero, increase in magnitude to a maximum, and then decrease slowly and in a somewhat oscillatory fashion to zero. On the scales adopted in the figures, the relative error in  $v(t)$  when  $\rho = 0.1$  shows up only as a barely noticeable blip near  $t = 1.5$ . Relative errors increase in magnitude with  $\rho$ . For  $p_0(t)$  and  $m(t)$ , the maximum relative errors when  $\rho = 0.95$  would appear to be under 4%, whereas for  $v(t)$ , this maximum is nearer to 15%. When  $\rho$  is significantly less than 0.9 or 0.95, these maximum relative errors are much smaller.

These constant parameter calculations are not difficult. Even when  $\rho = 0.95$ , the RKF45 algorithm was able to produce a solution to a truncated system of 80 equations as in (1) with error estimates in the range  $10^{-8}$  to  $10^{-12}$  using step lengths between 0.5 and 1.0. Nevertheless, some fairly significant changes in values are occurring (for example, when  $\rho = 0.95$ ,  $p_0(1.5) = 0.251$ , a drop of 75% in value in 1.5 time units,  $m(10) = 5.62$  against  $m(0) = 0$ , and  $v(25) = 53.4$  against  $v(0)=0$ , so that over this range, the derivative of  $v(t)$  averages more than 2.0). It

is encouraging to see that the much simpler closure approximation does reproduce these changes rather accurately.

It is more difficult to get an overall impression of the "pattern" of accuracy of the closure approximation for the oscillatory parameter calculations. Figures 1.7a, 1.8a, and 1.9a show the relative errors in  $p_0(t)$ ,  $m(t)$  and  $v(t)$  respectively for  $\rho_0 = \lambda_0/\mu$  values of 0.5, 0.7, and 0.9. Now the relative errors as a function of  $t$  are oscillatory with rather complicated shapes.

It is somewhat disquieting to note that the relative errors for larger values of  $\rho_0$  attain rather large magnitudes for certain ranges of values of  $t$ . For  $\rho_0 = 0.9$  and in the range  $t = 0$  to  $t = 25$ , the relative error in  $p_0(t)$  approaches 0.9 three times (always near the mid-period where  $\lambda(t)$  is greatest and so  $p_0(t)$  is smallest), and the error seems to be exhibiting an increasing sequence of maxima like 0.5, 0.7, .... at the beginning of each period (where the value of  $\lambda(t)$  is midway between its greatest and its smallest values).

Similar types of behaviour of the relative errors in  $m(t)$  and  $v(t)$  are also observed, but with one important qualification. The extreme values in the relative error plots are not greatest for the largest values of  $\rho_0$  over the range of values of  $\rho_0$  represented in Figures 1.8a and 1.9a. For smaller values of  $\rho_0$  like 0.1 and 0.3, the relative error extremes for  $m(t)$  and  $v(t)$  are smaller than they are for  $\rho_0 = 0.5$ . For the values of  $\rho_0$  for which calculations were done,  $\rho_0 = 0.5$  leads to the greatest extremes in the relative error in  $m(t)$  and  $\rho_0 = 0.7$  leads to the greatest extremes in the relative error in  $v(t)$ .

The reason for this shape of the relative error plots can be seen from Figures 1.7b-1.7d for  $p_0(t)$ , Figures 1.8b-1.8d for  $m(t)$ , and Figures 1.9b-1.9d for  $v(t)$ . In these figures, exact and approximate values are plotted for each of  $\rho_0 = 0.5, 0.7$ , and  $0.9$ . It is seen that the correspondence between the approximate and exact curves is not all that bad. Though the gaps between the two are readily observable in each instance, in most of these plots the approximate curve tracks the exact curve rather well.

These quantities are all oscillatory, reflecting the oscillatory mean arrival rate. Near the minimum values in each curve, where the graphs reverse direction, the gaps between the exact and approximate curves widen. As a result, absolute errors tend to be largest where the values of these quantities are smallest, leading to larger relative errors in those regions.

What is happening is that the approximation is poorest where the queue is not only changing rapidly, but where the rates of change of the queue are changing rapidly (which, as remarked earlier, is precisely where we would expect these closure techniques to break down since they are based on steady state relationships). This happens to coincide at times with regions where the quantities being computed attain their minimal absolute values. The instantaneous relative error is thus not a very good indication of the overall accuracy of the approximation, since its graph will be dominated visually by small regions of relatively large errors in small quantities.

### Period-Mean Values and Period-Mean Errors

Perhaps a better approach is to get some estimate of the 'average' relative error over a single period of oscillation. Using the known values of  $p_0(t)$ ,  $m(t)$  and  $v(t)$  in steps of  $\Delta t = 0.5$  forming the basis for the plots in Figures 1.7 - 1.9, we can apply Simpson's Rule to estimate the integral of each of these three quantities and their absolute errors over 'periods' of length 10 time units, and so obtain their mean value over that time interval.<sup>3</sup> Dividing the mean absolute error over such a period by the mean value of the quantity itself then gives a relative mean absolute error, which will be more representative (it is hoped) of the overall discrepancy between the exact and approximate solutions than are the point-by-point relative errors. This has been done in Table 1 for each of  $\rho_0 = 0.1, 0.3, 0.5, 0.7, \text{ and } 0.9$ , and in each case, for the two ranges  $t = 0 \text{ to } t = 10$  and  $t = 10 \text{ to } t = 20$ .

It is worth taking note of several features of the information in Table 1. First, the relative mean absolute errors do seem to reflect the degree of agreement between the graphs of approximate and exact values of these quantities. Except for  $m(t)$ , these relative errors increase with the value of  $\rho$  as happens in the constant parameter case.

---

<sup>3</sup>In effect, we would be using a 21-point Simpson's formula on an interval of length 10. When the function  $\sin(2\pi t/10)$  is integrated on the interval  $[0,5]$  using points spaced at intervals of 0.5 in width, the error in the integral obtained is 0.000174 on a value of 3.1830989 or about 0.0055%. While the quantities we are trying to integrate above are not simple sine functions, their shape is not so different that we would expect the Simpson formula to perform drastically poorer. It would appear that this approach will produce integrals of adequate accuracy for our purposes here.

A good example of the detailed correspondence between the Figures 1.7 - 1.9 and Table 1 occurs for the variance when  $\rho_0 = 0.9$ . In Figure 1.9d, it is clear that the two curves are moving apart steadily as  $t$  increases. From Table 1, we see that the relative mean absolute error over the interval  $[0,10]$  is 0.1953, whereas over the interval  $[10,20]$  it is 0.2470.

One does not get the same sense from the corresponding plot of the point-by-point relative error in Figure 1.9a, which consists of a sequence of peaks at about  $t = 9.5, 19.5, \dots$ . In fact, from Figure 1.9a, one probably gets the opposite impression since the peak at  $t = 19.5$  has the value 0.457 against a value of 0.493 at  $t = 9.5$ , and we tend to perceive peak height as most significant in graphs of errors. A closer look indicates that while the peak at  $t = 9.5$  is slightly higher than the peak at  $t = 19.5$ , the former peak is also sharper than the latter, and so the average error over an interval containing the first peak could well be smaller than the average error over an interval containing the second peak. In order that Figure 1.9a appear to agree with Table 1, we have to focus on both the heights and the widths of the peaks in Figure 1.9a. In work dealing with the oscillating parameter case for the retrial queue below, we will rely heavily on this relative mean absolute error to indicate the goodness of a particular closure approximation.

What is most striking about the relative mean absolute errors listed in Table 1 is that they are so small in general. The mean queue length,  $m(t)$ , seems particularly well determined by this closure approximation. Errors in  $v(t)$  and

$p_0(t)$  only become disquietingly large when  $\rho_0$  is 0.7 or larger. Because of the factor '1.1' in equations (4), this corresponds to a time average ratio of interarrival to service times of 0.77 or greater.

Also included in Table 1 are values of the relative errors found in computing mean values of  $p_0(t)$ ,  $m(t)$ , and  $v(t)$  over 10-unit time periods. These relative errors cannot be greater than the relative mean absolute errors discussed above, and in fact, will be equal to those only when the approximation produces uniformly positive or uniformly negative errors over the period in question. When actual approximation errors change sign, we find that errors in the computed mean value of these queue characteristics can become very small. In the case of  $p_0(t)$ , the relative mean absolute errors over the time intervals  $[0,10]$  and  $[10,20]$  are 23% and 38% respectively when  $\rho_0 = 0.9$ , whereas the relative error in the mean value of  $p_0(t)$  over each of these intervals are only 5% and 7% respectively. The difference between these two types of error for  $p_0(t)$  are even greater when  $\rho_0 = 0.5$ .

On the other hand, for  $v(t)$  with  $\rho_0 = 0.5, 0.7$ , and  $0.9$ , both types of relative error are identical because the Rothkopf and Oren closure method consistently overestimates  $v(t)$ . Relative errors in computed period-mean values are useful if the primary interest is to compute such mean values of queue properties. However, the relative mean absolute error is a more sensitive single number measure of the overall accuracy of an approximation method here, since it takes into account all errors with no cancellation between positive and negative errors in different subintervals of time.

The influence of the initial state of the system shows up in the difference between the averages computed in Table 1 applying to the time interval [0,10] compared to the time interval [10,20]. It is expected that after a sufficiently long time has elapsed, all of the properties of the queue will attain a sort of periodic steady state. Koopman [1972] refers to this periodic solution, with the statement that it is a consequence of classical theorems that only one solution of the basic system of differential equations will have the periodicity of the mean arrival rate and all other solutions will approach this periodic one exponentially with increase of time. He notes that this periodic solution is analogous to the steady state solution approached exponentially with increasing time in the constant parameter case [see Odoni and Roth, 1983]. For  $\rho_0 = 0.1$ , this periodicity is easily observed to six figures or better in the numerical results by the time  $t = 15$ . For other values of  $\rho_0$ , the graphs demonstrate a tendency towards establishing such a periodicity, but it has not occurred to similar precision over the time ranges for which solutions were generated.

### Extensions and Other Approaches

Finally, we note that Rothkopf and Oren [1979] also present a closure method for use with simple multiserver queues. In the case of  $s > 1$  servers, they require expressions for  $p_1(t)$ ,  $p_2(t)$ , ...,  $p_{s-1}(t)$  in terms of  $m(t)$  and  $v(t)$  in order to 'close' their pair of equations. They compute these probabilities using the negative binomial distribution, which is no longer quite correct. This means that the closure equations no longer produce the correct steady state solution in the

constant parameter case. Calculations indicate that the error due to this problem is small, and Rothkopf and Oren include a table of corrections to be applied (which depend on the number of servers and the value of  $\lambda/\mu$ ). The errors they observed in the corrected results obtained using their approximation were apparently of an acceptable magnitude.

Clark [1981] has also considered closure approaches to the multiple server problem, using what he calls a 'surrogate distribution'. He found that his approach generally was more accurate than the method of Rothkopf and Oren, though it required the solution of a system of five differential equations as opposed to their two. Since this thesis will not deal with the multiple server retrieval queue, it is not necessary here to examine these works in greater detail.

TABLE 1: Period-Mean Values of  $p_0(t)$ ,  $m(t)$  and  $v(t)$  and Their Relative Errors for the M/M/1 Queue When the Rothkopf and Oren Closure Method is Used.

The first column in each triple gives exact mean values over the time intervals indicated. The second column in each triple gives the relative error in the period-mean computed using values of the quantity given by the Rothkopf and Oren closure method. The third column in each triple gives the relative mean absolute error in the Rothkopf and Oren approximation in each case.

$\rho_0$	time interval	$p_0(t)$			$m(t)$			$v(t)$		
0.1	[0,10]	.8927	$-7.246 \times 10^{-7}$	$1.640 \times 10^{-5}$	.1271	$3.934 \times 10^{-7}$	.004351	.1058	$7.231 \times 10^{-4}$	.002532
	[10,20]	.8900	$1.231 \times 10^{-7}$	$1.494 \times 10^{-5}$	.1304	$4.858 \times 10^{-6}$	.004146	.1549	$6.350 \times 10^{-4}$	.002411
0.3	[0,10]	.6790	$-3.194 \times 10^{-5}$	.001772	.5965	-.002217	.008117	1.1086	.02674	.03398
	[10,20]	.6700	$-1.526 \times 10^{-6}$	.001755	.6162	-.002126	.007849	1.1518	.02564	.03292
0.5	[0,10]	.4677	.003103	.02301	1.7427	-.01787	.03886	4.9362	.1547	.1547
	[10,20]	.4501	.0004911	.02684	1.8373	-.01126	.03665	5.3220	.1607	.1607
0.7	[0,10]	.2722	.03590	.1246	4.1337	-.04107	.05356	14.4653	.2638	.2638
	[10,20]	.2357	.01114	.1496	4.7874	.03815	.05179	18.9005	.3799	.3799
0.9	[0,10]	.1304	-.05368	.2253	7.9358	-.03627	.03627	28.5147	.1953	.1953
	[10,20]	.07323	-.06970	.3817	11.3146	-.01401	.01536	55.2801	.2470	.2470

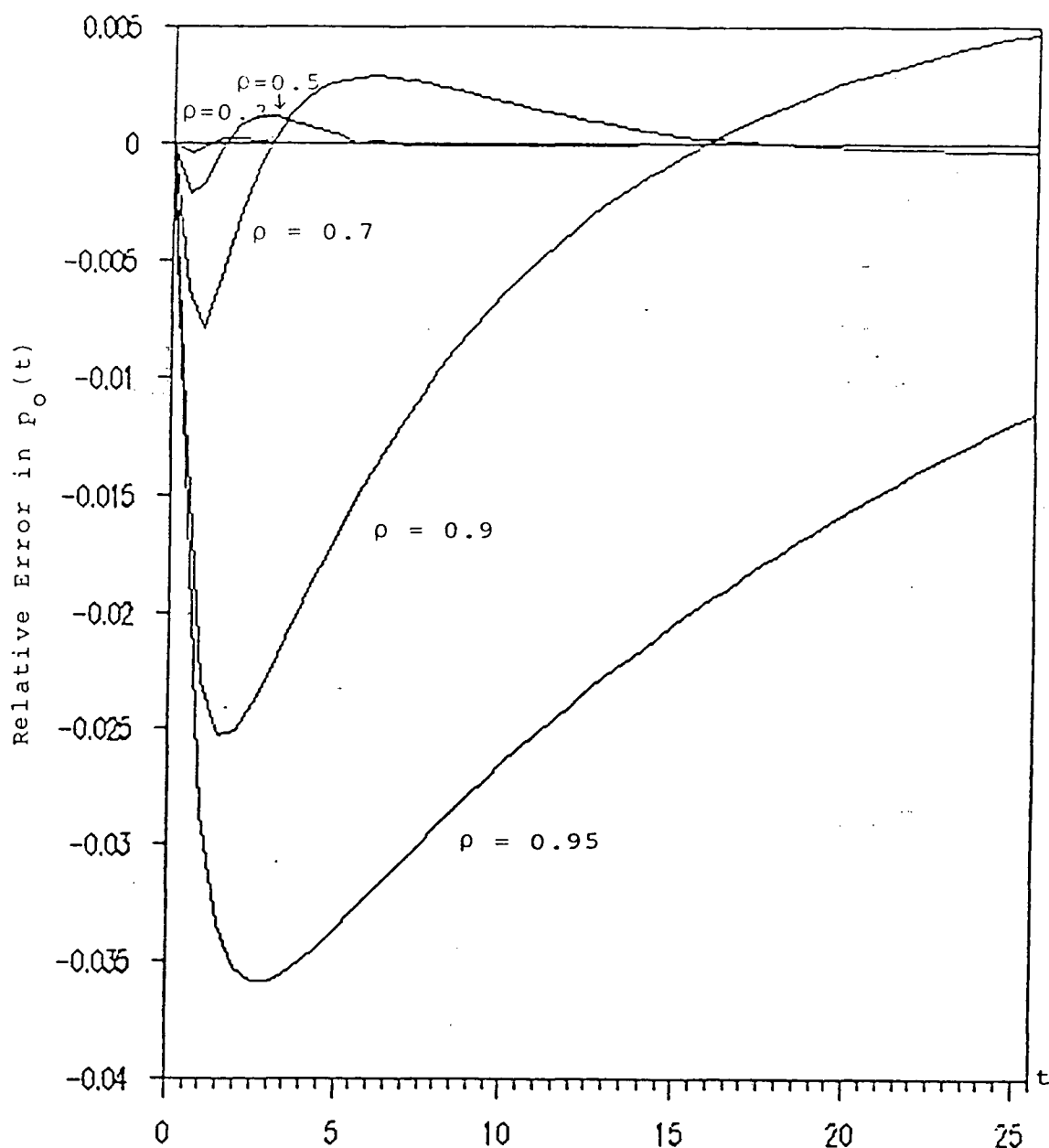


FIGURE 1.4 Relative Error in  $p_0(t)$  Computed Using Rothkopf and Oren's Closure Method for the Simple M/M/1 Queue with Constant Arrival and Service Rates.

The relative error is computed here as  $(\text{approximate value} - \text{exact value})/(\text{exact value})$ , plotted here for various values of  $\rho$ . The relative error when  $\rho = 0.3$  is discernable on the scale above only for small values of  $t$ , and that when  $\rho = 0.1$  is not distinguishable from the  $t$ -axis on this plot. The service rate  $\mu$  is fixed at 4.0/unit time in each calculation and the system is assumed empty at  $t = 0$ . Details of calculations to obtain 'exact values' are given in Appendix A.

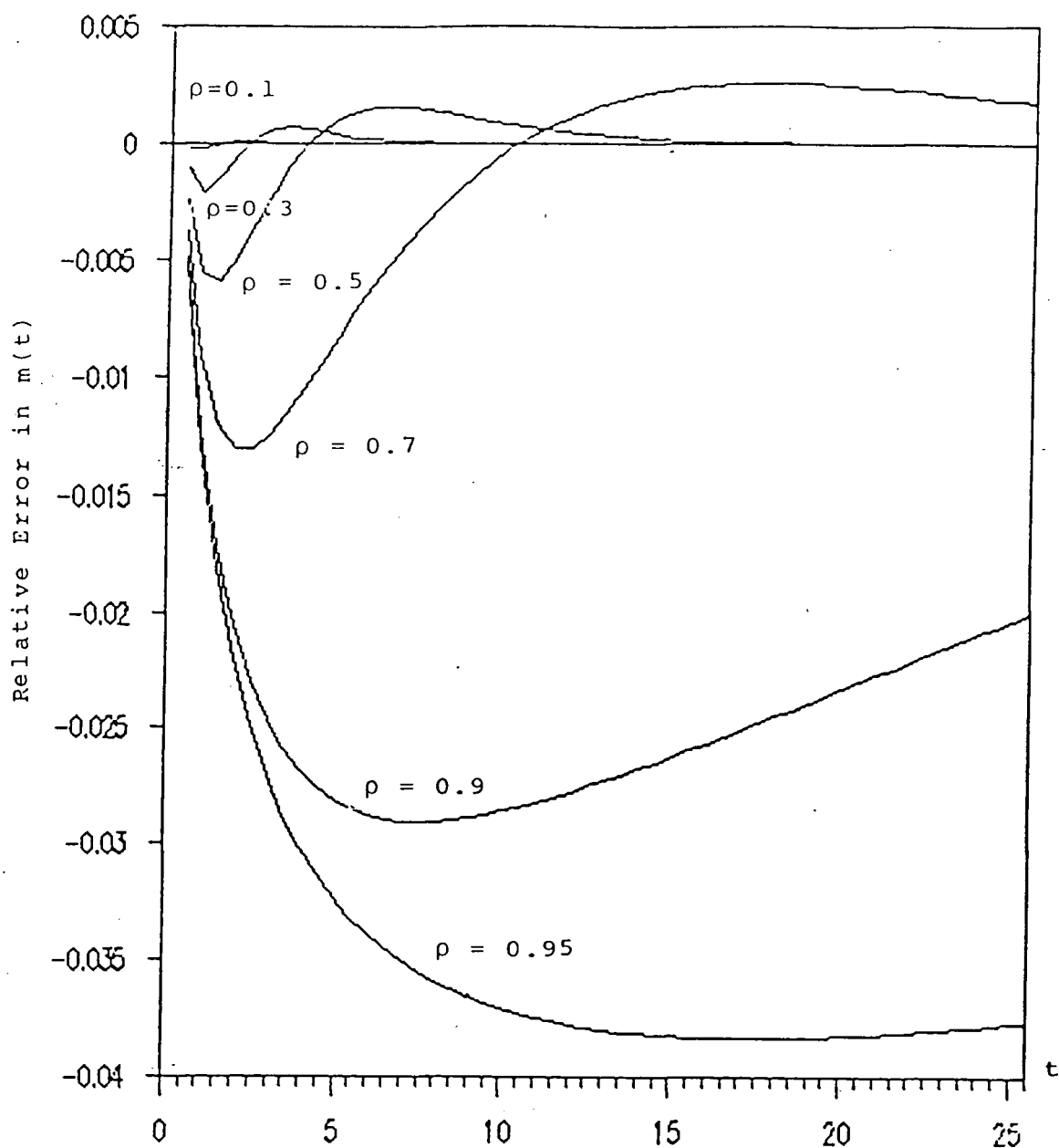
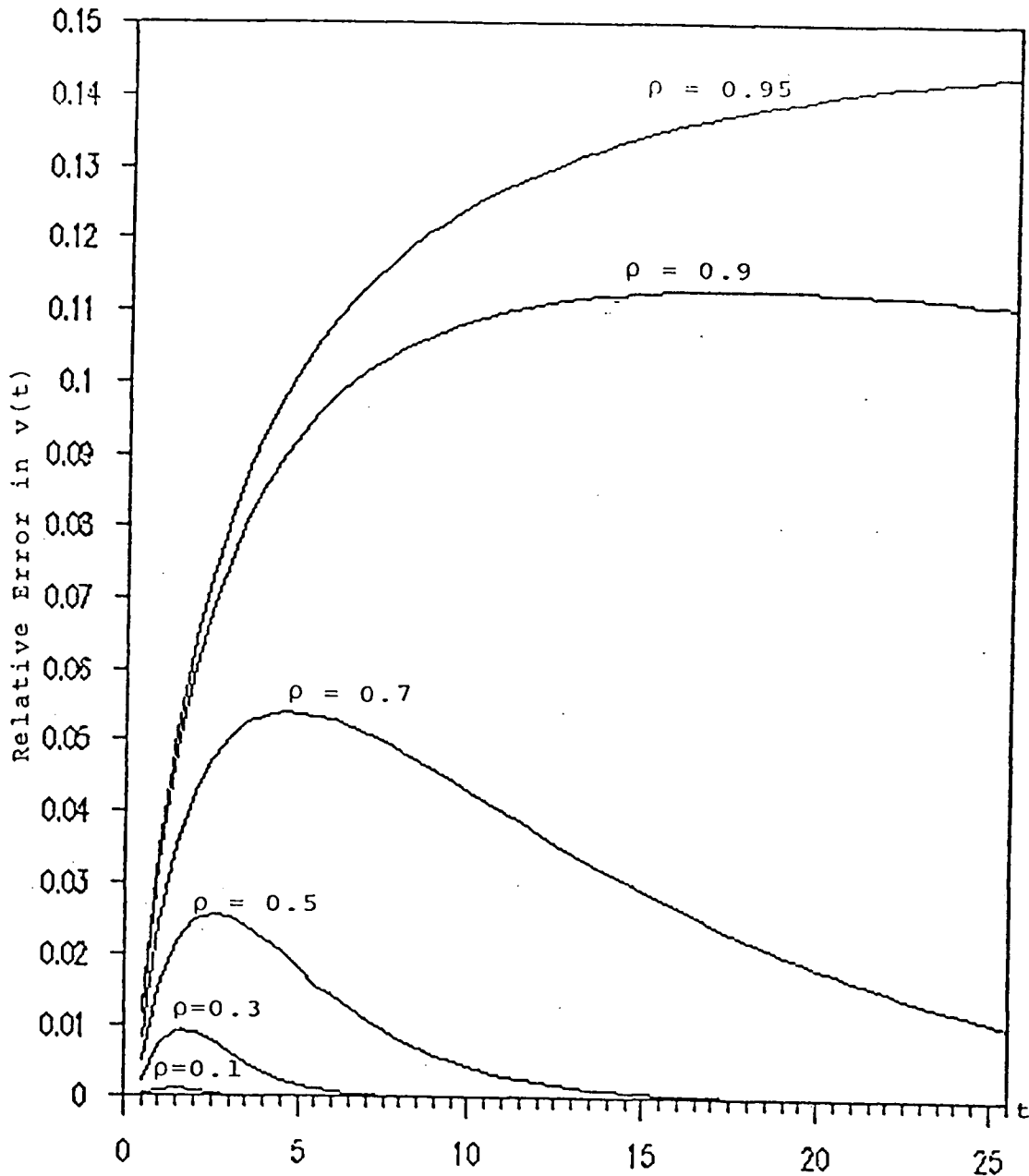


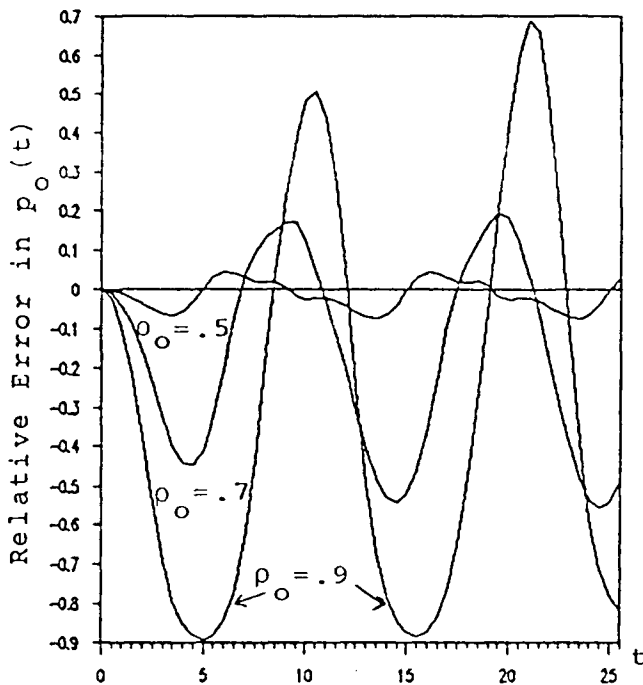
FIGURE 1.5 Relative Error in  $m(t)$  Computed Using Rothkopf and Oren's Closure Method for the Simple M/M/1 Queue with Constant Arrival and Service Rates.

The relative error is computed here as  $(\text{approximate value} - \text{exact value})/(\text{exact value})$ , plotted here for various values of the utilization  $\rho$ . The relative error when  $\rho = 0.1$  shows up only as a slight thickening of the  $t$ -axis for small values of  $t$  on this plot. The service rate  $\mu$  is fixed at 4.0/unit time in each calculation and the system is assumed empty at  $t = 0$ . Because  $m(0) = 0$ , the relative error is not defined at  $t = 0$  (but at  $t = 0$ , the absolute error in  $m$  is zero by definition here). Details of calculations to obtain 'exact values' are given in Appendix A.

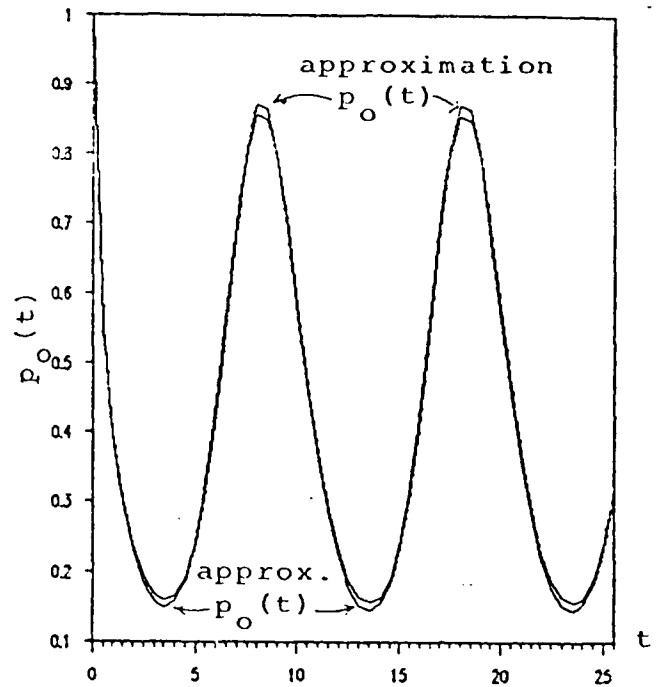
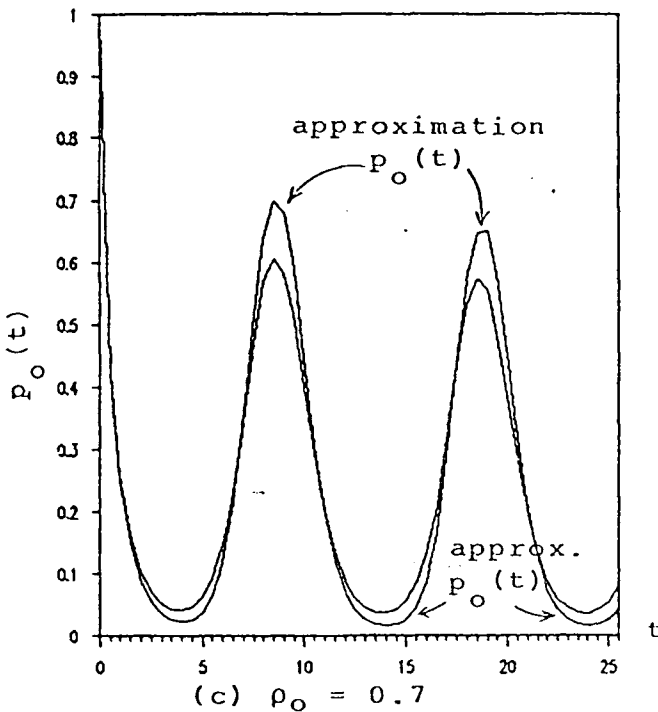
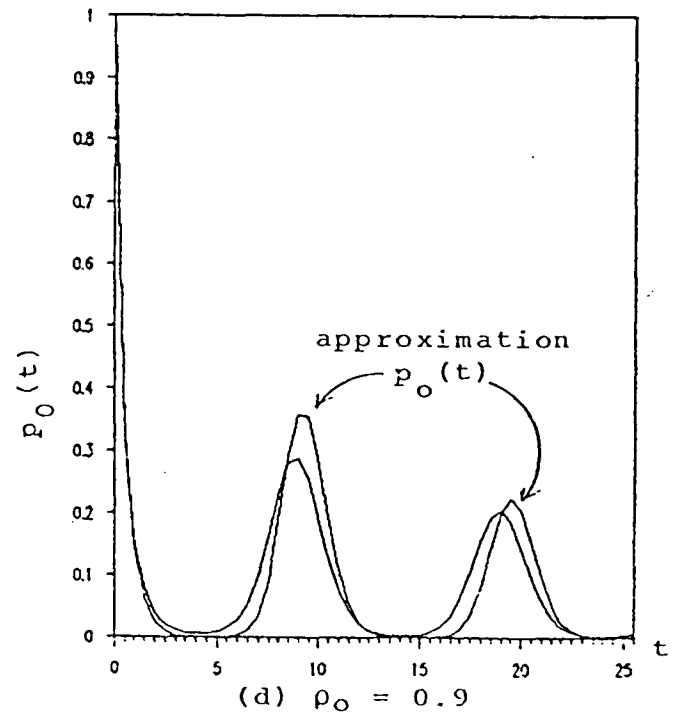


**FIGURE 1.6** Relative Error in  $v(t)$  Computed Using Rothkopf and Oren's Closure Method for the Simple M/M/1 Queue with Constant Arrival and Service Rates.

The relative error is computed here as  $(\text{approximate value} - \text{exact value})/(\text{exact value})$ , plotted here for various values of the utilization  $\rho$ . The relative error when  $\rho = 0.1$  is barely discernable along the  $t$ -axis for small values of  $t$  on this plot. The service rate  $\mu$  is fixed at 4.0/unit time in each calculation and the system is assumed empty at  $t = 0$ . Because  $v(0) = 0$ , the relative error is not defined at  $t = 0$  (but at  $t = 0$ , the absolute error in  $v$  is zero by definition here). Details of calculations to obtain 'exact values' are given in Appendix A.

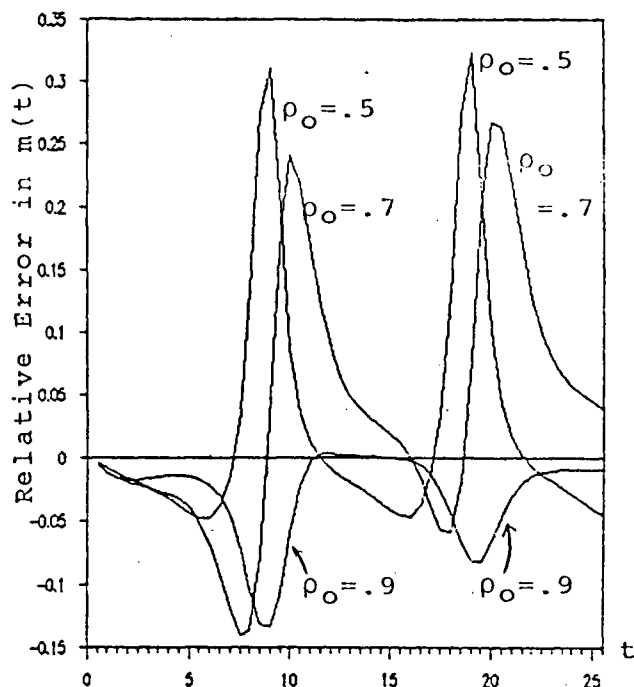


(a)

(b)  $\rho_0 = 0.5$ (c)  $\rho_0 = 0.7$ (d)  $\rho_0 = 0.9$ 

**FIGURE 1.7** Comparison of Exact  $p_0(t)$  and  $p_0(t)$  Computed Using Rothkopf & Oren's Closure Method of the M/M/1 Queue in the Oscillatory Parameter Case.

Figure 1.7a gives the relative error in the closure approximation  $p_0(t)$  for  $\rho_0 = 0.5$ , 0.7, and 0.9. Figures 1.7b,c,d plot the exact  $p_0(t)$  along with the closure approximation  $p_0(t)$  for  $\rho_0 = 0.5$ , 0.7, and 0.9, respectively. In each case, the plot of the approximate  $p_0(t)$  tends to swing wider at the maxima or minima relative to the plot of the exact  $p_0(t)$ . Here the period of oscillation is 10 time units and the system is assumed to be empty at  $t = 0$  (i.e.  $p_0(0) = 1.0$ ).



(a)

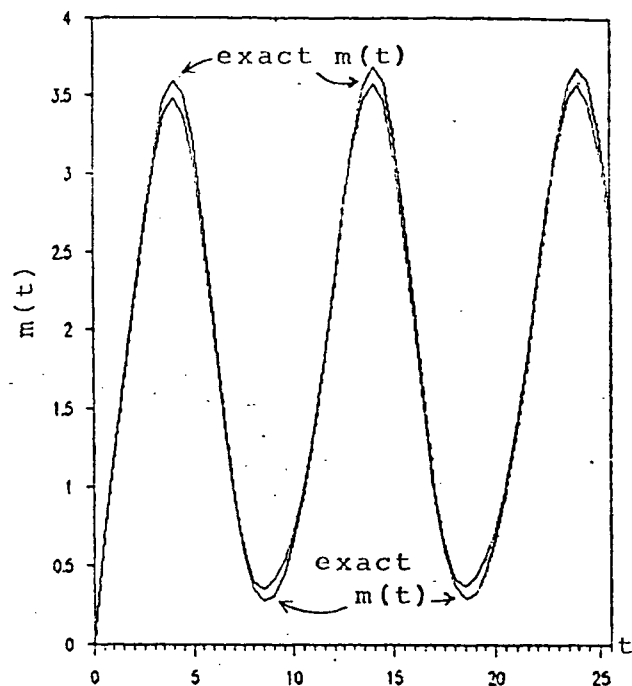
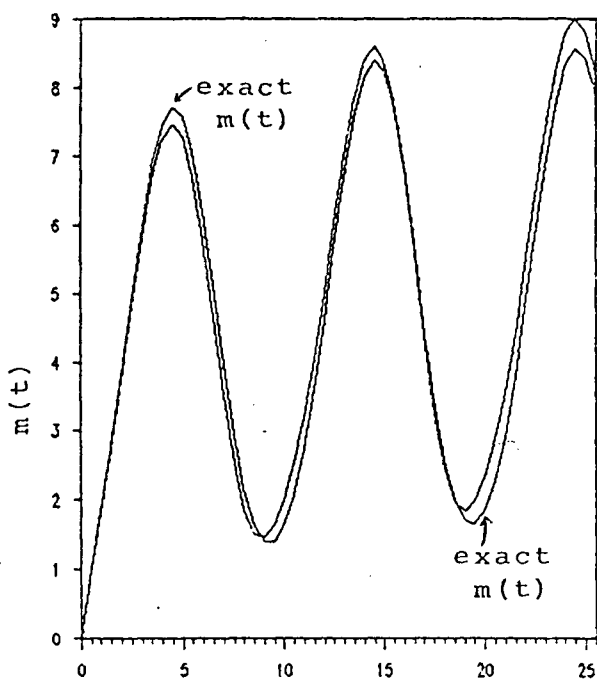
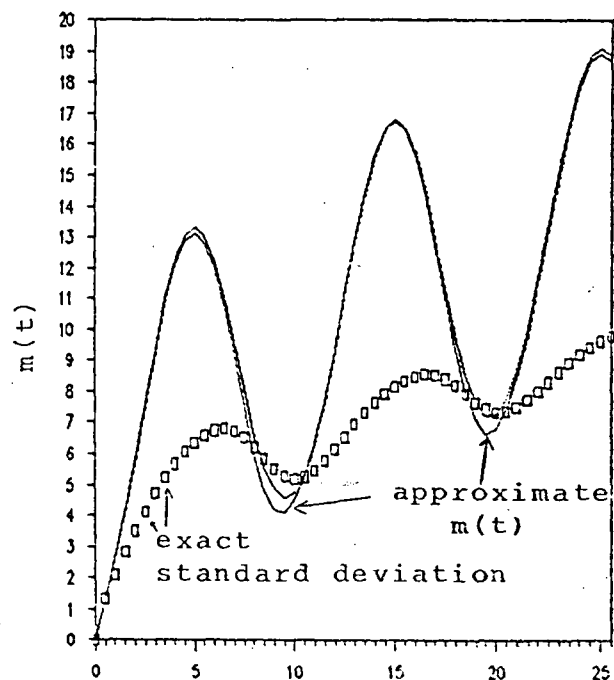
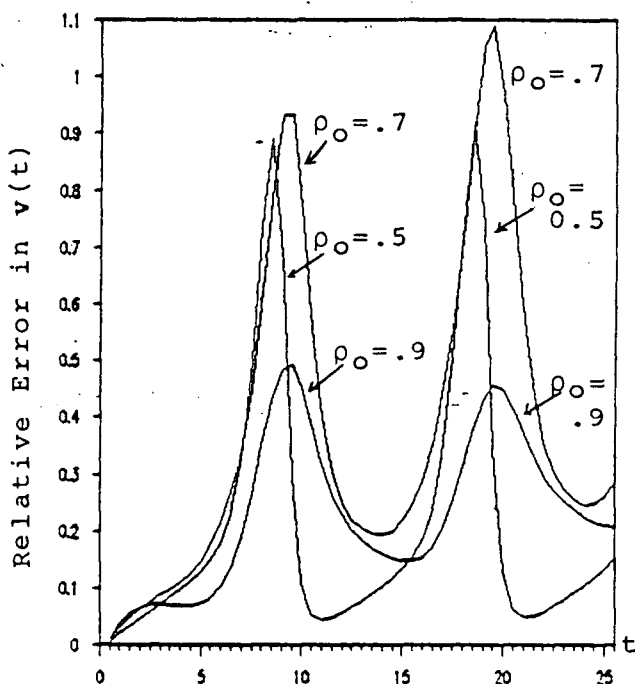
(b)  $\rho_O = 0.5$ (c)  $\rho_O = 0.7$ (d)  $\rho_O = 0.9$ 

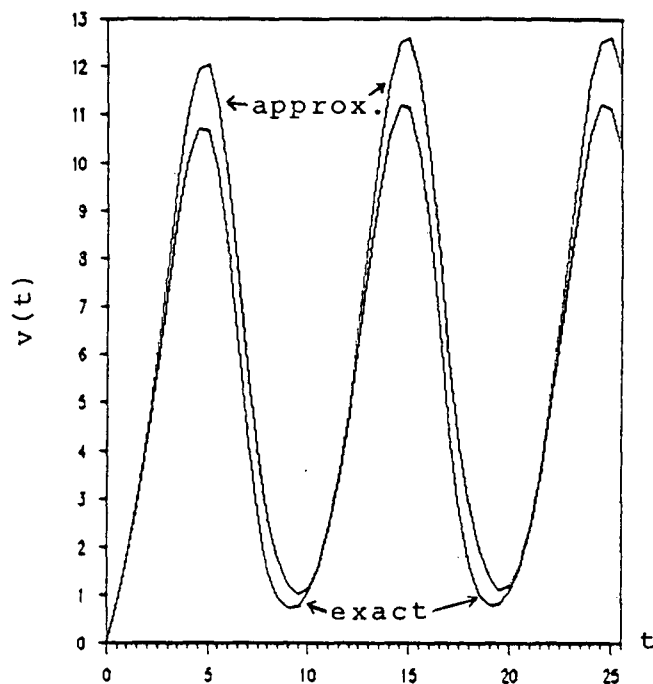
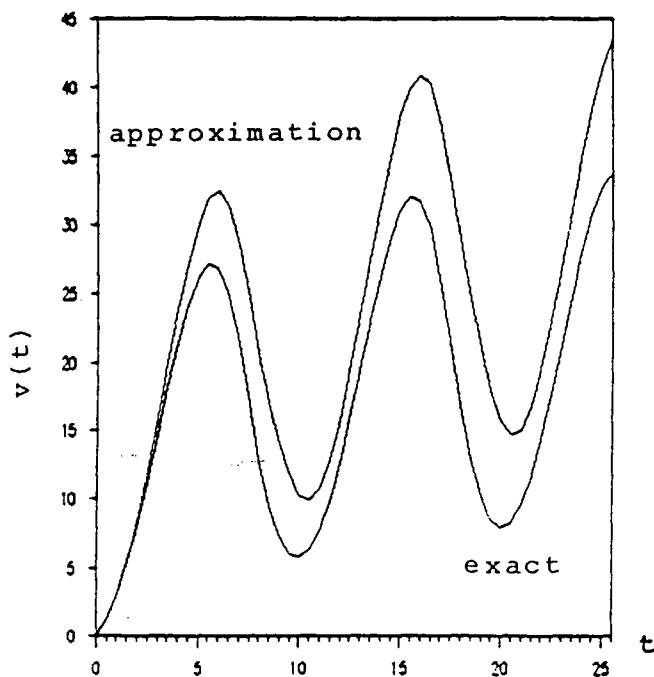
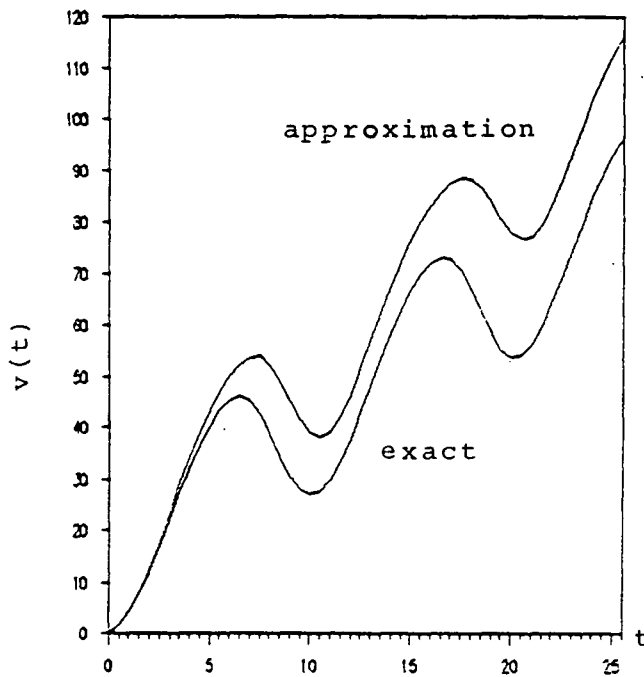
FIGURE 1.8

Comparison of Exact  $m(t)$  and  $m(t)$  Computed Using Rothkopf & Oren's Closure Method of the M/M/1 Queue in the Oscillatory Parameter Case.

Figure 1.8a gives the relative error in the closure approximation  $m(t)$  for  $\rho_O = 0.5$ , 0.7, and 0.9. Figures 1.8b,c,d plot the exact  $m(t)$  along with the closure approximation  $m(t)$  for  $\rho_O = 0.5$ , 0.7, and 0.9, respectively. At least in part because  $m(t)$  generally increases with  $\rho_O$ , the magnitude of relative errors generally decreases with  $\rho_O$  as shown in Figure 1.8a. Here the period of oscillation is 10 time units and the system is assumed to be empty at  $t = 0$  (i.e.  $m(0) = 0.0$ ).



(a)

(b)  $\rho_O = 0.5$ (c)  $\rho_O = 0.7$ (d)  $\rho_O = 0.9$ 

**FIGURE 1.9** Comparison of Exact  $v(t)$  and  $v(t)$  Computed Using Rothkopf & Oren's Closure Method of the M/M/1 Queue in the Oscillatory Parameter Case.

Figure 1.9a gives the relative error in the closure approximation  $v(t)$  for  $\rho_O = 0.5$ , 0.7, and 0.9. Figures 1.9b,c,d plot the exact  $v(t)$  along with the closure approximation  $v(t)$  for  $\rho_O = 0.5$ , 0.7, and 0.9, respectively. The approximate  $v(t)$  produces the upper curves throughout (note that the relative errors over the time range shown in Figure 1.9a are positive). Here the period of oscillation is 10 time units and the system is assumed to be empty at  $t = 0$  (i.e.  $v(0) = 0.0$ ).

## CHAPTER 2

### Closure Techniques for the Time-Dependent Single-Server Retrial Queue

*'Write that down', the King said to the jury, and the jury eagerly wrote down all three dates on their slates, then added them up, and reduced the answer to shillings and pence.*

*(Alice in Wonderland, Lewis Carroll)*

## 2.1 The Basic Equations

*Our little party of travellers awakened next morning refreshed and full of hope,....*

*(The Wizard of Oz, L. Frank Baum)*

### Retrial Queue Characteristics of Immediate Interest

We wish to do now for the M/M/1/1 retrial queue what was done for the simple M/M/1 queue in section 1.4. The retrial queue is more complicated and so the system of differential equations (1.1.3a,b) contains more detail than does the corresponding system (1.4.1) for the M/M/1 queue. As a result, the variety of potential closure techniques is much greater for the retrial queue.

The characteristics of the retrial queue of greatest likely interest are:

$$(1) \Pi_0(t) = \Pr(\text{server idle at time } t) = \sum_{n=0}^{\infty} p_{0n}(t)$$

(2)  $m(t) =$  mean number of customers in the system at time  $t$ , including any customer in service

$$\begin{aligned} &= \sum_{n=0}^{\infty} n \cdot p_{0n}(t) + \sum_{n=0}^{\infty} (n+1) p_{1n}(t) \\ &= m_0(t) + m_1(t) \end{aligned}$$

(3)  $v(t) =$  variance from the mean number of customers in the queue at time  $t$

$$\begin{aligned}
&= \sum_{n=0}^{\infty} n^2 p_{0n}(t) + \sum_{n=0}^{\infty} (n+1)^2 p_{1n}(t) - \{m(t)\}^2 \\
&= \zeta_0(t) + \zeta_1(t) - \{m(t)\}^2
\end{aligned}$$

(4)  $Y_0(t) =$  conditional mean number of customers in the system at time  $t$   
given that the server is idle at that time

$$= \frac{m_0(t)}{\pi_0(t)} .$$

The quantities  $m_0(t)$ ,  $m_1(t)$ ,  $\zeta_0(t)$ ,  $\zeta_1(t)$ , and so on are defined in section 1.1. The four quantities above are all determined by second or lower order moments of the probability distribution for the state of the retrial queue.

### Derivation of the Basic Equations

The first step in developing these closure techniques is to obtain equations for the derivatives of the quantities of interest. This is easily done by summing appropriate multiples of equations (1.1.3a) or (1.1.3b). For example, we have

$$p'_{0n}(t) = -(\lambda + n\nu) p_{0n}(t) + \mu p_{1n}(t), \quad n = 0, 1, 2, \dots \quad (1.1.3)$$

Summing both sides of this equation over all values of  $n$ ,  $n = 0, 1, 2, \dots$ , we get

$$\underbrace{\sum_{n=0}^{\infty} p'_{0n}(t)}_{\frac{d}{dt} \left[ \sum_{n=0}^{\infty} p_{0n}(t) \right]} = -\lambda \underbrace{\sum_{n=0}^{\infty} p_{0n}(t)}_{\Pi_0(t)} - \nu \underbrace{\sum_{n=0}^{\infty} n \cdot p_{0n}(t)}_{m_0(t)} + \mu \underbrace{\sum_{n=0}^{\infty} p_{1n}(t)}_{\Pi_1(t)}$$

or

$$\Pi'_0(t) = -\lambda \Pi_0(t) - \nu m_0(t) + \mu \Pi_1(t) . \quad (1)$$

Similarly, we can first multiply equation (1.1.3a) by  $n$  on both sides before summing, to get

$$\sum_{n=0}^{\infty} n \cdot p'_{0n}(t) = -\lambda \sum_{n=0}^{\infty} n \cdot p_{0n}(t) - \nu \sum_{n=0}^{\infty} n^2 p_{0n}(t) + \mu \sum_{n=0}^{\infty} n \cdot p_{1n}(t) .$$

We notice that

$$\begin{aligned} \sum_{n=0}^{\infty} n \cdot p_{1n}(t) &= \sum_{n=0}^{\infty} \{(n+1) - 1\} p_{1n}(t) = \sum_{n=0}^{\infty} (n+1) p_{1n}(t) - \sum_{n=0}^{\infty} p_{1n}(t) \\ &= m_1(t) - \Pi_1(t) , \end{aligned}$$

and, since  $n$  is independent of  $t$  on the left hand side,

$$\sum_{n=0}^{\infty} n \cdot p'_{0n}(t) = \frac{d}{dt} \left[ \sum_{n=0}^{\infty} n \cdot p_{0n}(t) \right] = m'_0(t) .$$

Thus, we obtain

$$m'_0(t) = -\lambda m_0(t) - \nu \zeta_0(t) + \mu m_1(t) - \mu \Pi_1(t) . \quad (2)$$

We can proceed in the same way to derive equations for  $m'_1(t)$ ,  $m''(t)$ , and so

on. The derivative of  $\gamma_0(t) = m_0(t)/\Pi_0(t)$  can be obtained by applying the basic formula for differentiating a quotient and using (1) and (2). Formulas for the derivatives of moments through second order are listed in Table 2.

### Additional Criteria

The next step is to take sets of one or more equations from Table 2 and eliminate enough of the quantities they contain using relationships between those quantities suggested by steady-state formulas so that the resultant system can be solved uniquely. These relationships ("closure formulas") can often be constructed in a number of different ways, and the assessment of their goodness or badness will ultimately depend on the outcome of test calculations. However, it seems useful to impose two ground rules on the process to eliminate approximations which are very likely to fail the computational tests.

*(1) Whenever possible, exact relationships are to be used to eliminate quantities from equations in preference to presumed approximate relationships based on inspection of the steady state solution.*

Thus, to eliminate  $\Pi_1(t)$  in terms of either  $\Pi_0(t)$  or  $m_0(t)$  from equation (1), we would always use  $\Pi_1(t) = 1 - \Pi_0(t)$ , since this relation must always hold for valid probability distributions. At steady state, it is true that  $\Pi_1 = \rho = m_0/c$ . However, we would not consider eliminating  $\Pi_1(t)$  from equation (1) by writing

**TABLE 2. Formulas for Moment Derivatives for the Single Server Retrial Queue.**

$$(1) \quad \Pi'_0 = -\lambda \Pi_0 + \mu \Pi_1 - \nu m_0$$

$$(2) \quad m'_0 = -\lambda m_0 - \nu \zeta_0 + \mu m_1 - \mu \Pi_1$$

$$(3) \quad m'_1 = \lambda m_0 + \nu \zeta_0 + \lambda - \mu m_1$$

$$(4) \quad m' = \lambda - \mu(1 - \Pi_0)$$

$$(5) \quad \zeta'_0 = -\lambda \zeta_0 - \nu \eta_0 + \mu \zeta_1 - 2\mu m_1 + \mu \Pi_1$$

$$(6) \quad \zeta'_1 = \lambda \zeta_0 + \nu \eta_0 - \mu \zeta_1 + 2\mu m + \lambda$$

$$(7) \quad \zeta' = 2(\lambda - \mu)m + \lambda + \mu \Pi_1 + 2\mu m_0$$

$$(8) \quad \nu' = \lambda + \mu - \mu \Pi_0(2m + 1) + 2\mu m_0$$

$\Pi_1(t) = m_0(t)/c$  because this would allow the resulting closure equations to have a solution violating the fundamental relation

$$\Pi_0(t) + \Pi_1(t) = 1, \quad (3)$$

whereas using  $\Pi_1(t) = 1 - \Pi_0(t)$  would eliminate that possibility. Other inviolable relations here are

$$m_0(t) + m_1(t) = m(t), \quad (4)$$

and

$$\zeta_0(t) + \zeta_1(t) = \zeta(t). \quad (5)$$

*(2) We will insist that proposed closure methods correctly produce steady state values as  $t$  becomes large in constant parameter calculations.*

This amounts to requiring that when derivatives in the closure equations are replaced by zeros, and  $\lambda$ ,  $\mu$ , and  $\nu$  are constants, the resulting system of algebraic equations give uniquely the correct steady state expressions for the quantities whose derivatives were originally present. For example, in section 2.3 below, we propose a closure technique for computing  $\Pi_0(t)$  and  $m(t)$  based on the two-equation system

$$\begin{aligned} \Pi_0'(t) &= -\lambda \Pi_0(t) + \mu \Pi_1(t) - \nu m(t) + \nu m_1(t) \\ m'(t) &= \lambda - \mu \Pi_1(t) \end{aligned} \quad (6)$$

The steady state solution is obtained correctly by replacing the derivatives by zeros here to get

$$0 = -\lambda \Pi_0 + \mu \Pi_1 - \nu m + \nu m_1 ,$$

$$0 = \lambda - \mu \Pi_1 .$$

The second equation gives immediately

$$\Pi_1 = \frac{\lambda}{\mu} = \rho ,$$

or, using  $1 - \Pi_0 = \Pi_1$ , we get  $1 - \Pi_0 = \rho$ , so that the solution can be rewritten as  $\Pi_0 = 1 - \rho$ , which is the correct expression for the constant parameter steady state case. Then from the first equation, we have

$$\begin{aligned} m &= \frac{1}{\nu} \{-\lambda \Pi_0 + \mu \Pi_1\} + m_1 \\ &= \frac{1}{\nu} \{-\lambda(1 - \rho) + \mu \rho\} + m_1 \\ &= \frac{\lambda}{\nu} \rho + m_1 \\ &= c\rho + \frac{\rho(c\rho + 1)}{1 - \rho} = \frac{\rho(c + 1)}{1 - \rho} , \end{aligned}$$

as required. The substitution of the steady state expression for  $m_1$  in the last line above is justified by the fact that we will eliminate  $m_1(t)$  from the system (6) using steady state relationships to obtain a viable closure method for computing  $\Pi_0(t)$  and  $m(t)$ .

An example of a system of equations which does not satisfy this criterion is the three-equation system

$$\begin{aligned}\Pi'_0(t) &= -\lambda\Pi_0(t) + \mu\Pi_1(t) - \nu m(t) + \nu m_1(t), \\ m'(t) &= \lambda - \mu\{1 - \Pi_0(t)\}, \\ v'(t) &= \lambda + \mu - \mu\Pi_0(t)\{2m(t) + 1\} + 2\mu m_0(t),\end{aligned}\tag{7}$$

which might be used as a basis for a closure method for computing  $\Pi_0(t)$ ,  $m(t)$  and  $v(t)$ . When the derivatives on the left hand side are replaced by zeros, we obtain a system of equations which contains no reference to the variance  $v$ .

The resultant system can be solved for the correct steady state expressions for say  $\Pi_0$ ,  $m_0$ , and  $m$  once  $\Pi_1 = 1 - \Pi_0$  and  $m_1 = m - m_0$  are eliminated (no approximations are used here). But, since the steady state expression for the variance cannot be determined in this fashion, we would reject this system of equations as a basis for a closure method for computing  $\Pi_0(t)$ ,  $m(t)$  and  $v(t)$ .

In fact, if one proceeds to eliminate say  $\Pi_0(t)$ ,  $m_1(t)$ , and  $m_0(t)$  from equations (7) in terms of  $\Pi_0(t)$  and  $m(t)$ , the transient time-dependent solution of the system for the constant parameter case is not too bad for  $\Pi_0(t)$  and  $m(t)$ . The solution for  $v(t)$  is not at all good, however, and as  $t$  increases,  $v(t)$  does not apparently converge to a value anywhere near its actual steady state value.<sup>1</sup>

---

<sup>1</sup>As mentioned earlier, Rothkopf and Oren [1979] did propose a closure method for multiserver queues which did not converge in the constant parameter case to the correct steady state values. However, they were able to devise a table of approximate numerical corrections, which, when applied to the solutions produced by their method, did result in correct steady state values. These

In a small number of test calculations done for five or six methods which violated either or both of these ground rules, the numerical results were invariably unrealistic, with moments negative at times, probabilities significantly exceeding unity at times, and clear convergence to incorrect steady state values for increasing values of  $t$  in constant parameter calculations. The first rule above amounts to avoiding use of inexact relationships when viable exact relationships are available. The second rule amounts to requiring that the methods proposed be capable of handling the simplest aspect of the simplest type of time-dependent queue behaviour if we are to consider it seriously for describing time-dependent queue behaviours in more complicated situations.

### Notation

The closure methods examined in this thesis are described in detail in the following four sections. We categorize them according to the number of equations involved. For example, methods 1A and 1B are alternative approaches based on a single equation. Similarly methods 2A, 2B, 2C, ..., would be alternative approaches based on systems of two equations.

---

numerical corrections were then carried over to variable parameter calculations. In their case, Rothkopf and Oren were able to trace the source of the errors to specific assumptions in the underlying probability distribution on which they based their approach. Here, the error in  $v(t)$  at steady state is being approached is an artifact of the numerical algorithm employed to solve the system (7), and so general correction factors do not exist.

## 2.2 One-Equation Closure Methods

*Committees of twenty deliberate plenty,  
Committees of ten act now and then,  
But most jobs are done by committees of one.*

(old rhyme quoted by Laurence Peter in  
Why Things Go Wrong)

It is possible to devise relatively simple single-equation closure methods for calculation of either  $\Pi_0(t)$  or  $m(t)$ . Though neither method produces highly accurate results, it is interesting to note how close their solutions are to the exact solution when so little detail is included in the computation.

### Method 1A ( $\Pi_0(t)$ )

To obtain a single equation for approximating  $\Pi_0(t)$ , we begin with the first equation in Table 2,

$$\Pi_0'(t) = -\lambda \Pi_0(t) + \mu \Pi_1(t) - \nu m_0(t) \quad . \quad (1)$$

$\Pi_1(t)$  is replaced here by the exact equivalent,  $1 - \Pi_0(t)$ . At steady state,

$$m_0 = c\rho = c(1 - \Pi_0) \quad , \quad (2)$$

as seen from equation (1.1.12). Thus, as an approximation, we can write equation (1) as

$$\Pi_0'(t) = -\lambda \Pi_0(t) + \mu \{1 - \Pi_0(t)\} - \nu c \{1 - \Pi_0(t)\} \quad ,$$

or, simplifying, as

$$\Pi_0'(t) = -\mu \Pi_0(t) + \mu - \lambda \quad . \quad (3)$$

There is a certain degree of plausibility to this formula. When the server is idle,  $\Pi_0(t) = 1$ , then  $\Pi_0'(t) = -\lambda$ , indicating that the probability of the server being idle will decrease with increasing new customer arrival rate. When the server is busy,  $\Pi_0(t) = 0$ , then  $\Pi_0'(t) = \mu - \lambda$ , indicating that  $\Pi_0(t)$  will tend to increase if the service completion rate is greater than the new customer arrival rate. The term  $-\mu \cdot \Pi_0(t)$  in equation (3) simply indicates that one cannot expect the difference in service and arrival rates to lead to high rates of change in  $\Pi_0(t)$  if the system is largely empty ( $\Pi_0(t)$  near 1).

One significant deficiency in this formula is that it is independent of the retrial rate,  $\nu(t)$ . In a way, this is what you would expect of a formula based on steady-state relationships. When the number of customers in the system is no longer changing, then over any given time interval, the number of customers served (whether they are new customers served the first time they request service, or old customers who have been attempting to obtain service over a very long time) must equal the number of new customers arriving in that time interval.

It is difficult to speculate about the effect of this deficiency other than to say that for very low or for very high retrial rates, the value of  $\Pi_0$  will be primarily determined by  $\lambda$  and  $\mu$  when the server is not saturated. Under these conditions, when  $\nu$  is very small, the arrival of an orbiting customer is a rare event and most of the customers seen by the server are first-trial customers. The system would look much like an M/M/1 queue with no waiting room.

Similarly, when  $\nu$  is very large compared to  $\lambda$ , orbiting customers would tend to be served before the next new customer arrives, again giving something like the M/M/1 queue with no waiting room and a random delay of onset of service after the customer arrives. On this basis, we would expect the lack of dependence of (3) on  $\nu(t)$  to be most critical at intermediate values of  $\nu(t)$ , when compared with  $\lambda$  and  $\mu$ . It would be quite straightforward to test this suggestion by a series of calculations, but that has not been done.

When  $\lambda$  and  $\mu$  are constants, equation (3) can be solved easily in closed form to get

$$\Pi_0(t) = (1 - \rho) + \{\Pi_0(0) - (1 - \rho)\}e^{-\mu t} \quad (4)$$

In particular, if  $\Pi_0(0) = 1$  (that is, if the server is initially idle), then this becomes

$$\Pi_0(t) = (1 - \rho) + \rho e^{-\mu t} \quad (5)$$

Figure 2.1a shows plots of  $\Pi_0(t)$  against time for  $\Pi_0(t)$  computed exactly and then using equation (5), for the constant parameter case with  $\lambda = 2$  and  $\mu = 4$ . It is seen that equation (5) gives a solution here that decays to the steady state value of  $\Pi_0$  much too fast with time. Figure 2.1b gives a plot of the relative error in  $\Pi_0(t)$  computed using (5). While this plot indicates a maximum relative error of only 18% at  $t = 1$ , it should be noted that for these values of  $\lambda$  and  $\mu$ , equation (5) does not represent a solution significantly more accurate numerically than simply defining

$$\Pi_0(0) = 1$$

$$\Pi_0(t) = 1 - \rho, \quad t > 0 \quad .$$

It will be recalled from section 1.4 that maximum relative errors in  $\Pi_0(t)$  in the constant parameter case using Rothkopf and Oren's closure method and  $\rho = 0.5$  was significantly below 1%, so equation (5) cannot seriously be considered as an acceptably accurate approximation.

When  $\lambda$  or  $\mu$  vary with time, equation (3) must be solved numerically. Figure 2.2 is a plot of  $\Pi_0(t)$  computed exactly and  $\Pi_0(t)$  computed using (3) with  $\mu(t) = 4$  throughout and  $\lambda(t)$  given by equation (1.4.4) with  $\lambda_0 = 2.0$  and  $P = 10.0$ . The retrial rate is also taken as oscillatory now, using the formula

$$v(t) = v_0 \left[ 1.1 + \sin \left( \frac{2\pi(t-1)}{P} \right) \right]. \quad (6)$$

The 't-1' was included in the oscillatory part to give the effect of retrial rate changes lagging new customer arrival rate changes.<sup>1</sup>

Equation (3) does lead to a solution which reproduces the periodic behaviour of

---

<sup>1</sup>One might attribute this kind of variation in retrial rates to unsuccessful customers (of which there will be more when the new customer arrival rate is higher) initially increasing their rate of reapplication for service. Then, eventually the ones who still don't succeed in obtaining service will decrease their rate of retrial out of frustration or exhaustion. Of course, the retrial rate,  $v(t)$ , is applied to all customers in orbit -- those who have just entered the system and those who have been orbiting for some time, and so it represents a kind of average collective 'enthusiasm' of all customers in orbit at a particular time. This is not the place to discuss in greater detail the psychology of unsatisfied customers, but these remarks underline again the fact that the usefulness in actual application of all of the computational techniques discussed or used in this thesis may well be overshadowed by non-mathematical considerations. The main goal of the work in this thesis is to find viable ways of circumventing the solution of large systems of first order linear differential equations. This is only a sub-goal of the much larger goal of being able to predict the behaviour of real-life single server retrial queues.

the exact solution rather well, but the approximate solution overshoots badly at the relative extrema of  $\Pi_0(t)$ . This overshoot causes the approximate  $\Pi_0(t)$  to become slightly negative in this case in the vicinity of the relative minima of  $\Pi_0(t)$ . The mean relative error over the time interval  $[0,10]$  is 0.336 and over the time interval  $[10,20]$  is 0.347, which is also not very good.

#### Method 1B ( $m(t)$ )

A second single-variable closure approximation can be obtained from the fourth equation in Table 2,

$$m'(t) = \lambda - \mu\{1 - \Pi_0(t)\} \quad , \quad (7)$$

to determine  $m(t)$ . We note that at steady state,

$$1 - \Pi_0 = \Pi_1 = \rho$$

and

$$m = \frac{\rho(c+1)}{1-\rho} = \frac{\Pi_1(c+1)}{\Pi_0} \quad . \quad (8)$$

Solving (8) for  $\Pi_1 = 1 - \Pi_0$ , in terms of  $m$ , we get

$$1 - \Pi_0 = \frac{m}{(c+1+m)} \quad , \quad (9)$$

which, on substitution into (7) gives the equation

$$m'(t) = \lambda - \frac{\mu m(t)}{c+1+m(t)} \quad (10)$$

As with equation (5) above for  $\Pi_0(t)$ , equation (10) leads to a transient time-

dependent solution in the constant parameter case which converges much faster than the exact solution to the steady state value of  $m(t)$ . This can be seen in Figure 2.3a. Figure 2.3b shows that the maximum relative error is again much greater than for Rothkopf and Oren's method applied to the M/M/1 queue.

The results of a calculation based on (10) when the arrival rate is oscillatory are shown in Figure 2.4. The tracking of the exact graph of  $m(t)$  by the solution of (10) is remarkable, though again the closure approximation overshoots on the bends. The average relative error on the time interval  $[0,10]$  is 0.269 and on the time interval  $[10,20]$  is 0.196. These compare with values of 0.0389 and 0.0367 respectively for the Rothkopf and Oren method applied to the simple M/M/1 queue under similar loads.

For what it's worth, equation (10) is separable when  $\lambda$ ,  $\mu$ , and  $\nu$  are constants, and can be rewritten as

$$\frac{c + 1 + m}{\lambda(c + 1) + (\lambda - \mu)m} dm = dt \quad . \quad (11)$$

Integration of both sides gives

$$\frac{m}{\lambda - \mu} - \frac{\mu(c + 1)}{(\lambda - \mu)^2} \log \left[ \frac{\lambda(c + 1) + (\lambda - \mu)m}{\lambda(c + 1)} \right] - m(0) = t \quad (12)$$

For example, when  $\lambda = 2$ ,  $\mu = 4$ ,  $\nu = 2$ , and  $m(0) = 0$ , this reduces to

$$\frac{-m}{2} - 2 \cdot \log(1 - \frac{m}{2}) = t \quad . \quad (13)$$

We could regard this as a closure method for computing approximate times at which  $m(t)$  attains particular allowed values in the constant parameter case. Since it is not a particularly difficult nonlinear equation to solve numerically, it can also be used as an alternative to equation (10) in computing values of  $m(t)$  for specific numerical values of  $t$ . In fact, formula (12) is more efficient to use when  $m(t)$  is desired for only a few isolated values of  $t$ , since the integration of equation (10) requires determining  $m(t)$  for a sequence of values of  $t$  between the initial time and the time of interest.

Neither of these closure approximations would be suitable for serious study of the time-dependent behaviour of a single-server retrial queue. However, it is noteworthy that the solutions of the approximating single equations mimic the exact behaviour of the corresponding characteristics of the queue so well, even when nothing resembling a steady state is occurring.

**FIGURE 2.1**  $\Pi_O(t)$  Computed Using Method 1A in the Constant Parameter Case.

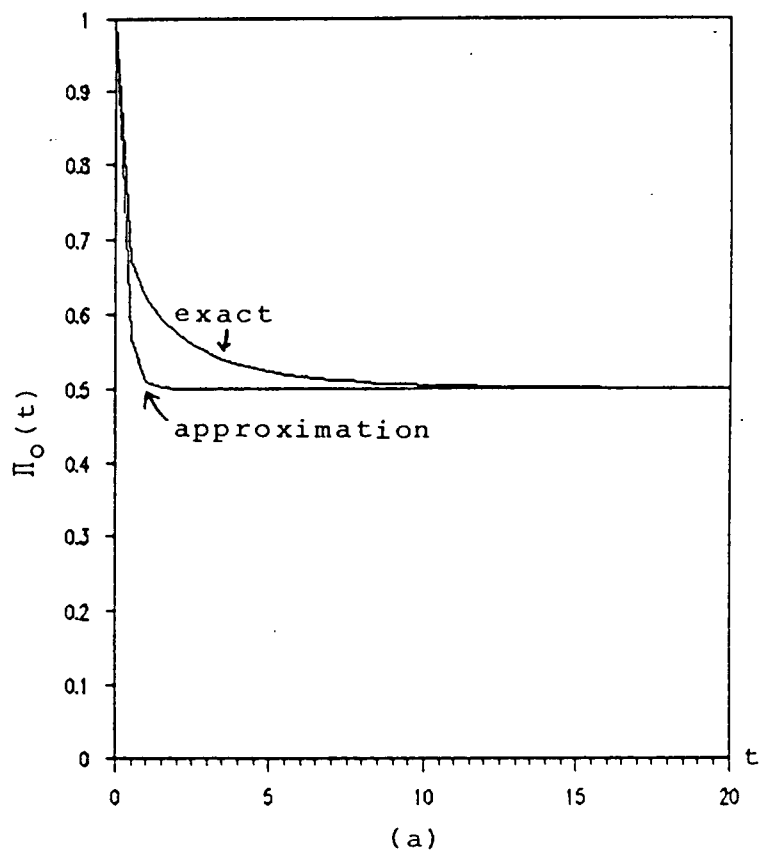
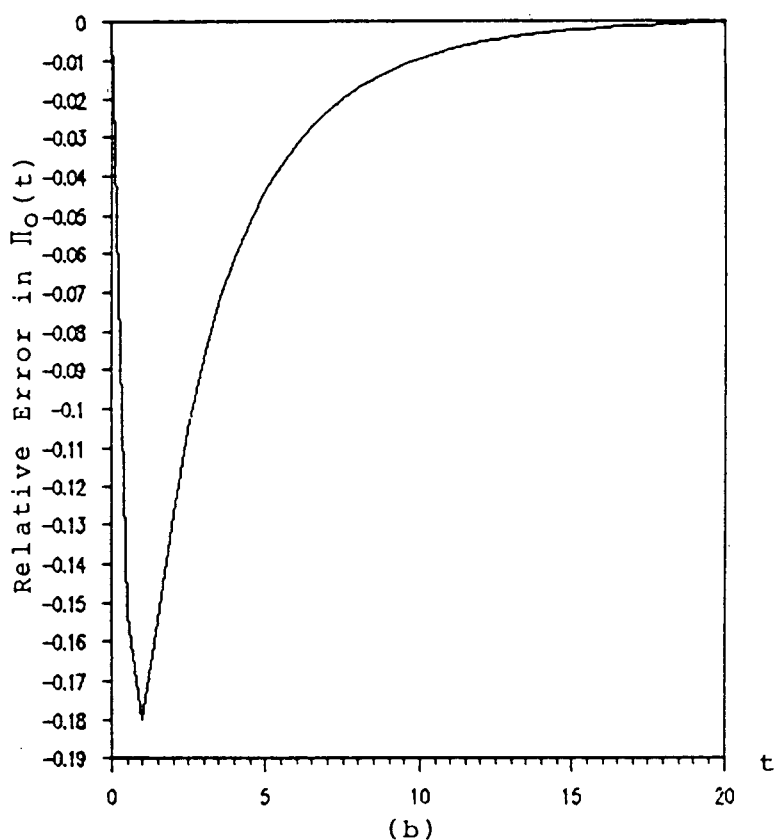


Figure 2.1a compares exact values of  $\Pi_O(t)$  for a constant parameter calculation ( $\lambda = 2.0$ ,  $\mu = 4.0$ ) with  $\Pi_O(t)$  computed using the approximation Method 1A, equation (2.2.5). Both curves converge quickly to the steady state value  $\Pi_O = 0.5$ . Figure 2.1.b is a plot of the relative error in the approximate  $\Pi_O(t)$  shown in Figure 2.1a. For these calculations, the system was assumed initially empty, i.e.,  $\Pi_O(0) = 1.0$ .



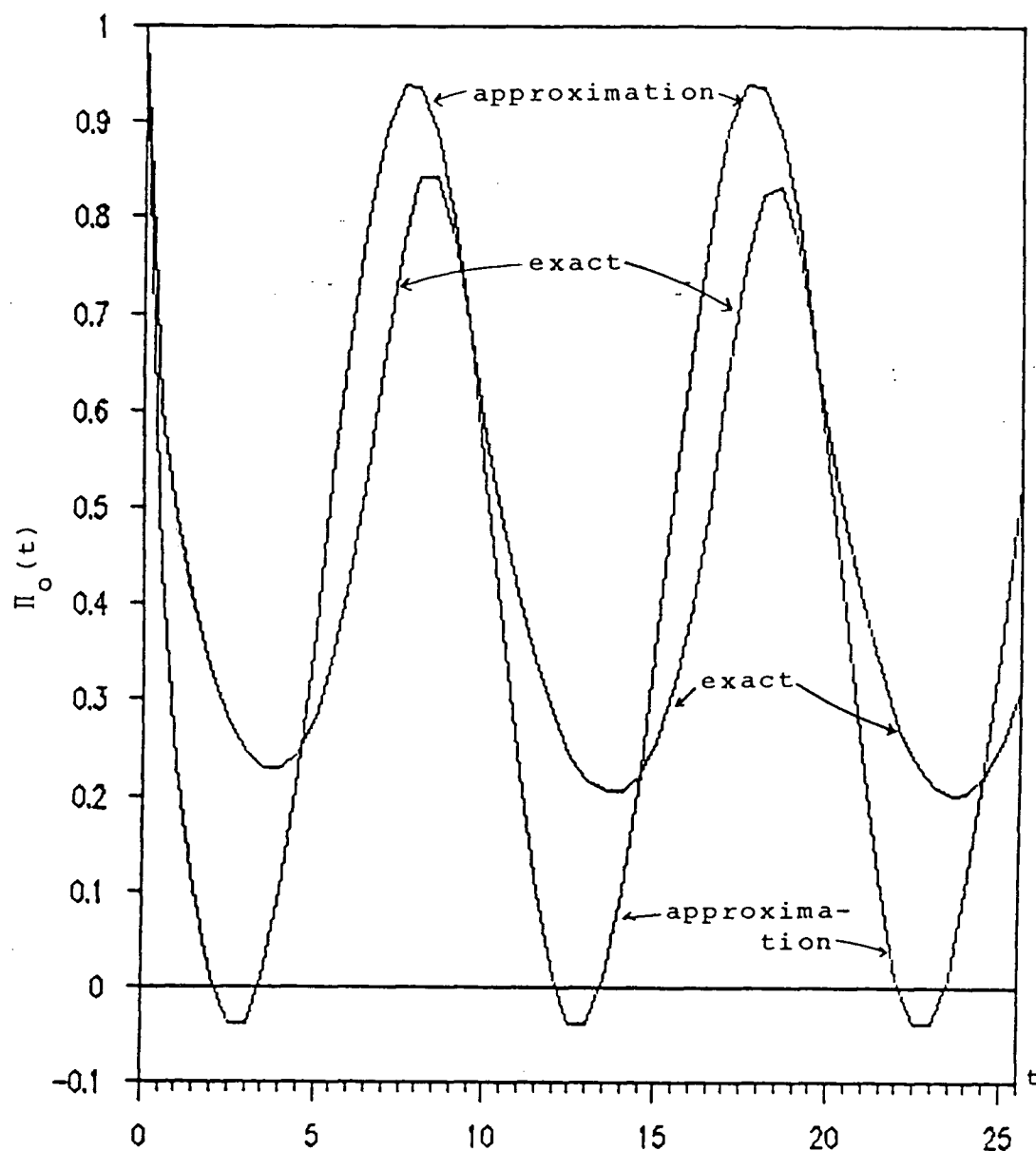
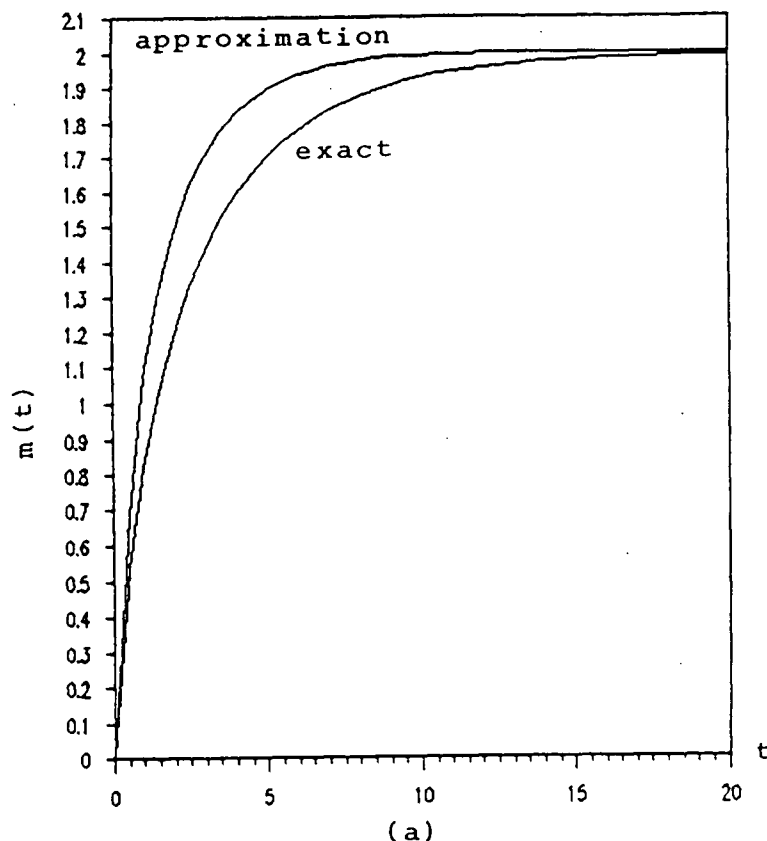


FIGURE 2.2 Comparison of  $\Pi_0(t)$  Computed Approximately Using Method 1A With Its Exact Values in the Oscillatory Parameter Case.

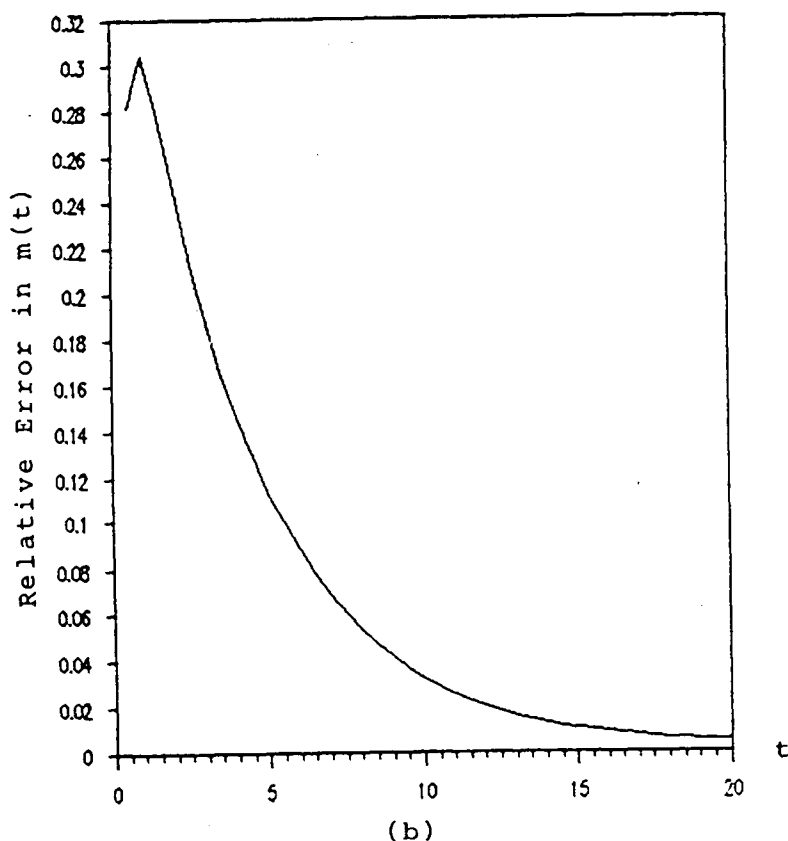
The graph of the approximation is the one which swings wider at all local maxima and minima. The calculation was done with  $\lambda_0 = 2.0$ ,  $\nu_0 = 2.0$ ,  $\mu = 4.0$ , and a period of oscillation of 10 time units. Note the small but disconcerting excursions of the approximation curve below the horizontal axis.

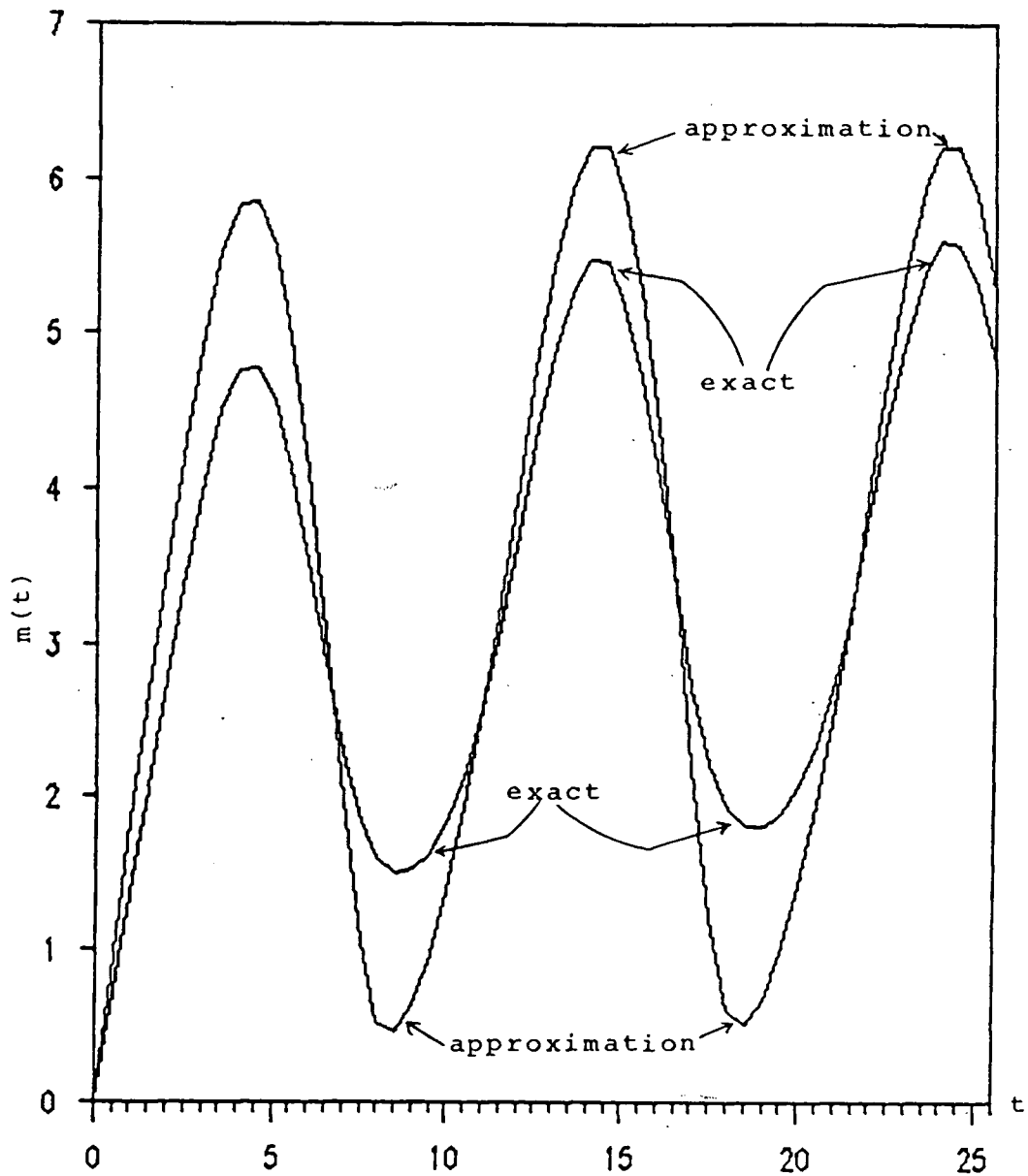


**FIGURE 2.3** Mean Number of Customers,  $m(t)$ , in the Retrial Queue Computed Using Method 1B in the Constant Parameter Case.

The lower curve in Figure 2.3a is the exact  $m(t)$  when  $\lambda = 2.0$ ,  $\mu = 4.0$ , and  $\nu = 2.0$ . The upper curve is  $m(t)$  computed using Method 1B, equation (2.2.10).

Figure 2.3b shows the corresponding relative error in this approximation  $m(t)$ . The closure approximation estimate of  $m(t)$  approaches the steady state value of 2.0 much too fast. In these calculations, it was assumed that the system was initially empty, i.e.  $m(0) = 0$ .





**FIGURE 2.4** Mean Number of Customers,  $m(t)$ , in the Retrial Queue Computed Using Method 1B in the Oscillatory Parameter Case.

For the curves plotted here,  $\lambda_0 = 2.0$ ,  $\mu(t) = 4.0$ ,  $\nu_0 = 2.0$ , and the period of oscillation is 10 time units. Initially, the system is empty, i.e.,  $m(0) = 0.0$ .

### 2.3 Two-Equation Closure Methods: $\Pi_0$ and $m$

*'I should like to buy an egg please,' she said timidly. 'How do you sell them?' 'Fivepence farthing for one -- two pence for two,' the sheep replied. 'Then two are cheaper than one?' Alice said in a surprised tone, taking out her purse. 'Only you must eat them both, if you buy two,' said the Sheep.*  
[Through the Looking Glass, Lewis Carroll]

#### Methods 2A and 2B ( $\Pi_0(t)$ and $m(t)$ )

The first and fourth equations in Table 2 can be used as the basis of a much more accurate closure approximation than those examined for retrial queues in the previous section. We have

$$\begin{aligned}\Pi'_0(t) &= -\lambda\Pi_0(t) + \mu\Pi_1(t) - \nu m(t) + \nu m_1(t) , \\ m'_1(t) &= \lambda - \mu\Pi_1(t) .\end{aligned}\tag{1}$$

In order to turn these equations into a system involving just  $\Pi_0(t)$  and  $m(t)$ , it is necessary to eliminate  $\Pi_1(t)$  and  $m_1(t)$ .

As before, we have the exact relation  $\Pi_1(t) = 1 - \Pi_0(t)$ . However, for  $m_1(t)$  it is necessary to resort to some sort of approximation. At steady state, we have

$$m_1 = \frac{\rho(c\rho + 1)}{1 - \rho} ,\tag{2}$$

whereas

$$m = \frac{\rho(c + 1)}{1 - \rho} , \quad \Pi_0 = 1 - \rho , \quad (\Pi_1 = \rho) .\tag{3}$$

We would like to express  $m_1$  at steady state in terms of both  $m$  and  $\Pi_0$ , since then account of the values of both quantities will be taken in computing  $\Pi'_0(t)$ . Just by inspection, we note that

$$m_1 = \frac{\rho(c+1)}{1-\rho} \cdot \frac{c\rho+1}{c+1} = m \cdot \frac{c\Pi_1+1}{c+1}. \quad (4)$$

This formula reflects the type of relationship expected between changes in  $m_1$  and corresponding changes in  $m$  and in  $\Pi_1$ . As  $m$  increases, we expect both of its components,  $m_0$  and  $m_1$ , to increase since in this situation there would be larger numbers of customers present when the server is busy as well as when the server is idle. Also,  $m_1$  is the component of  $m$  corresponding to the server being busy. As  $\Pi_1$  increases, the proportion of the time the server is busy is larger and so the component  $m_1$  is expected to reflect a larger fraction of  $m$ .

A second obvious approach is simply to replace  $\rho$  by  $\Pi_1$  everywhere in (2) (or  $1-\rho$  by  $\Pi_0$ ) to obtain the closure formula

$$m_1 = \frac{\Pi_1(c\Pi_1+1)}{\Pi_0}, \quad (5)$$

which effectively expresses  $m_1$  in terms of  $\Pi_0$  only. Again, this formula has qualitative plausibility since it expresses the fact that, for reasons cited just above,  $m$  will tend to increase directly (though not necessarily in strict direct proportion) with increases in  $\Pi_1$  and inversely with increases in  $\Pi_0$ .

We will have a problem here if  $\Pi_0(t)$  attains the value zero. However, in that

case all  $p_{0n}(t)$ ,  $n = 0, 1, 2, \dots$ , must be zero, and so  $m_0(t)$  is zero and we have that  $m_1(t) = m(t)$ . It is unlikely that  $\Pi_0(t)$  will vanish except when the system is started with the server assumed busy (then  $\Pi_0(0) = 0$ ). However, it is likely that this proposed closure formula will be at its worst when  $\Pi_0(t)$  gets closer to zero.

Adopting (4) as the relationship between  $m_1(t)$ ,  $\Pi_1(t)$  and  $m(t)$  at all times in equations (1) leads to method '2A', say, and (5) leads to method '2B'.

### Numerical Performance of Methods 2A and 2B

A set of test calculations were done using (1) with method 2A and with method 2B in the constant parameter case with  $\lambda = 2$ ,  $\mu = 4$ , and  $\nu = 2$ . Relative errors are plotted against time in Figures 2.5a and 2.5b for each of these two options. The maximum relative error in  $\Pi_0(t)$  in this example is around 4% for both methods. The maximum relative error in  $m(t)$  is around 10% for both methods. It appears that the 2B method is slightly better in this constant parameter case than is the 2A method, though the difference is very slight.

A corresponding set of calculations was done using these two closure methods in the oscillatory parameter case with  $\lambda_0 = 2$  and  $\nu_0 = 2$ , leading to the relative error plots in Figures 2.6a and 2.6b. The oscillation period here, as usual, is ten time units with maximum arrival rates occurring at  $t = 2.5, 12.5, \dots$ , and minimum arrival rates occurring at  $t = 7.5, 17.5, \dots$ . The 2A-method gives  $\Pi_0(t)$  with relative errors oscillating between roughly +0.1 and -0.1, which is comparable

to the relative errors in  $p_0(t)$  computed using Rothkopf and Oren's approach (see Figure 1.7a). The relative errors in  $\Pi_0(t)$  computed using the 2B-method here oscillated between + 0.2 and -0.2 roughly.

On the other hand, relative errors in  $m(t)$  as computed using the 2A-method vary between 0 and 0.6 (except for an initial region where the errors in  $m(t)$  are negative<sup>1</sup>), whereas the 2B-method results in relative errors in  $m(t)$  lying in the much narrower range of -0.2 to +0.1 . Keep in mind that magnitudes of relative errors can be somewhat misleading, particularly where the quantity under study has rather small numerical magnitudes.

#### Method 2C ( $\Pi_0(t)$ and $m(t)$ )

What is more interesting about the two sets of Figures, 2.5 and 2.6, is that where the 2A-method results in a positive and maximal error, the 2B-method roughly gives a negative and minimal error. This strongly suggests that an intermediate method '2C' should prove even more accurate than either of these two.

For any one of the exact solutions we have for particular test cases, we could construct a combination of (4) and (5) to minimize in some sense the errors in the solution produced by the hybrid method '2C'. However, here we will simply test the potential value of such a hybrid closure method by quite arbitrarily

---

<sup>1</sup>Throughout, errors have always been computed as 'approximate - exact', and relative errors are this quantity divided by the exact value.

adopting an equal mix of the two approximations above, and for method 2C simply writing <sup>2</sup>

$$m_1(t) = \frac{1}{2} \left\{ m(t) \frac{c\Pi_1(t) + 1}{c + 1} + \frac{\Pi_1(t) \{c\Pi_1(t) + 1\}}{\Pi_0(t)} \right\}. \quad (6)$$

### Constant Parameter Calculations Using Method 2C

Straightforward application of (6) in the constant parameter calculations done to produce Figures 2.5a,b indicates the maximum relative error in  $\Pi_0(t)$  has been reduced to something like 0.013 and in  $m(t)$  to something like 0.04. This represents a decrease in the maximum relative error by factors of three roughly.

A more extended series of calculations was carried out using this method 2C. With  $\nu$  fixed at 2.0 and  $\mu$  at 4.0, calculations were run for  $\lambda = 0.4, 1.2, 2.0, 2.8, 3.6,$  and  $3.8$  (corresponding to  $\rho = 0.1, 0.3, 0.5, 0.7, 0.9,$  and  $0.95$ ) in the constant parameter case. Then, with  $\mu$  fixed at 4.0 and  $\lambda$  fixed at 2.0, constant parameter calculations were run for  $\nu = 10, 4, 2, 1, 0.4,$  and  $0.2$  (corresponding to  $c = 0.2, 0.5, 1, 2, 5,$  and  $10$  respectively). Plots of the relative errors in  $\Pi_0(t)$  for each of these values of  $\lambda$  appear in Figure 2.7a and for each of these values of  $\nu$  in Figure 2.7b. Similar plots of the relative error in  $m(t)$  are given in Figures 2.8a and 2.8b.

---

<sup>2</sup>The value of varying the way simple closure formulas are 'mixed' is considered at greater length in section 2.4 .

There is nothing very complicated about these plots. Generally, the maximum relative errors increase in magnitude for increasing  $\lambda$  and for decreasing  $\nu$  for both  $\Pi_0(t)$  and  $m(t)$ . Also, the usually observed decrease in relative error as  $t$  gets large takes longer to reach the vicinity of zero for those calculations resulting in larger peaks in the relative error, mainly because this decrease has farther to go and not because it is remarkably slower.

The other main thing to note is that over this fairly wide range of parameter values, the relative errors do not become unreasonably large. In the worst cases shown, they peak in the vicinity of 0.10, whereas in the application of Rothkopf and Oren's method to the simple M/M/1 queue, these maximal relative errors were more in the vicinity of 0.04, as seen in Figures 1.4 and 1.5. A simple direct comparison of Figures 2.7 and 2.8 with Figures 1.4 and 1.5 is not easy, since the relative errors found here for the retrial queue calculations are quite sensitive to the retrial rate -- much more so for  $m(t)$  than for  $\Pi_0(t)$ . The current method appears viable over a wide range of operational conditions for the retrial queue.

It becomes quite difficult computationally to obtain exact solutions when  $\lambda$  or  $\nu$  are very large. For example, for the calculation with  $\nu = 10$ ,  $\lambda = 2$ , and  $\mu = 4$ , the exact calculation of the  $p_{\xi_n}(t)$  from which  $\Pi_0(t)$  and  $m(t)$  were computed in half-integer steps over the time range  $t = 0$  to  $t = 25.5$  required just over 10,000 evaluations of the derivative vector and about 25 seconds of computer processor time. On the other hand, the closure method required under 2000 derivative evaluations and less than 0.4 seconds of computer processor time to obtain the approximate values of  $\Pi_0(t)$  and  $m(t)$  at these points. As mentioned

before, no rigorous attempt was made to optimize any of the programs developed for these calculations and so the actual amounts of processor time required are not significant. But, what is significant is that at a cost of incurring relatively small errors, this closure approximation in the case cited reduced the computing time required by a factor of 60. For most queue parameter value combinations, such a dramatic reduction in computing time would not be observed, yet the reductions are still significant.<sup>3</sup>

### Oscillatory Parameter Calculations Using Method 2C

A corresponding series of calculations was carried out using this method in the oscillatory parameter case. Calculations were done with  $\mu(t) = 4$ ,  $\nu_0 = 2$ , and  $\lambda_0 = 0.4, 1.2, 2.0, 2.8$  and  $3.6$  (corresponding to  $\rho_0 = 0.1, 0.3, 0.5, 0.7$ , and  $0.9$ , respectively). Another set was done where  $\mu(t) = 4$ ,  $\lambda_0 = 2$ , and  $\nu_0$  varied over the six values  $10, 4, 2, 1, 0.4$ , and  $0.2$ .

No calculations were done with  $\lambda_0 = 3.8$  in the oscillatory parameter case, since the term '1.1' in equation (1.4.4) results in an average arrival rate of approximately 4.2 over a period of oscillation. This exceeds the average service rate of 4.0 over each period, and so the system is divergent. In fact, with  $\lambda_0 = 3.6$ , the time-average arrival rate is 3.96, corresponding to a time-averaged utilization of  $\rho_0 = 0.99$ , which is a system very near saturation. Since our

---

<sup>3</sup>In general, exact calculations become very costly for both large values of  $\lambda$  (near  $\mu$ ) and large values of  $\nu$ . Large values of  $\lambda$  require system truncation at large dimensions for acceptable accuracy. Large values of  $\nu$  lead the RKF45 algorithm to use very short internal step lengths, for reasons which are not clear, and so calculations require large numbers of derivative evaluations per time increment.

calculations were limited to the first 2-1/2 periods, and we start with an empty system, we do not have to deal with the large mean queue lengths that would occur by the time this system reaches the equilibrium periodic state.

In Figure 2.9a are plotted exact values of  $\Pi_0(t)$  against time, giving an indication of the nature of the variation of this quantity with time in the oscillatory parameter case. It is seen that the influence of the initial state here is more pronounced as  $\rho_0$  increases. Also, the regions of maximal value of  $\Pi_0(t)$  become increasingly sharper peaks as  $\rho_0$  increases, whereas the troughs become wider. Remember that large values of  $\Pi_0(t)$  correspond to less customers in the system. The wide smooth troughs (compared to the sharp peaks) are likely due to the more uniform utilization of the server because of the increased effective retrial rate,  $n\lambda(t)$ , when the number of customers present is larger. It would be interesting to repeat these calculations using an oscillatory service rate (either in-step with the arrival rate or with a slight delay relative to the oscillating arrival rate).

In Figure 2.9b the relative error in  $\Pi_0(t)$  as computed using this hybrid closure formula is plotted against  $t$  for  $\rho_0 = 0.1, 0.3, 0.5, 0.7$ , and  $0.9$ . The higher values of  $\rho_0$  lead to wider swings in this plot of relative error. On the scale used, relative errors in  $\Pi_0(t)$  when  $\rho_0 = 0.1$  show up as a periodic thickening of the  $t$ -axis only. A plot of the exact and approximate  $\Pi_0(t)$  against time is given in Figure 2.9c. The closure approach does not produce as wide swings from minimum to maximum values of  $\Pi_0(t)$  around  $t = 4, 14, 24, \dots$  units.

The plots in Figure 2.9c give the impression that the closure approach  $\Pi_0(t)$  is decaying to the periodic steady state more quickly than the exactly computed  $\Pi_0(t)$ . This pattern of error was also noted for the one-variable closure methods examined in the preceding section.

Relative mean absolute errors are listed in Table 3a for these calculations. They too increase sharply for increasing  $\rho_0$  as well as for increasing time at each value of  $\rho_0$ . However, this increase in the relative mean absolute errors is again due at least partly to the decreasing size with time of the quantity to which the mean absolute errors are being compared, since the mean absolute errors appear to decrease with time.

Figures 2.10a-d deal with the computation of  $m(t)$  in the oscillatory parameter case. Figure 2.10a shows  $m(t)$  against time for  $\rho_0 = 0.1, 0.3, 0.5, 0.7$ , and  $0.9$ . For  $\rho_0 = 0.1, 0.3$ , and  $0.5$ , the periodic steady state seems to be established very quickly, and there is little significant difference between successive periods of variation from the beginning. A more quantitative indication of this is given by the mean values of  $m(t)$  over the first two ten-unit time intervals, listed in Table 3a.

For  $\rho_0 = 0.7$  and  $0.9$ , the local maxima and minima of  $m(t)$  are still increasing very markedly after two such periods, and by comparison with steady state values of  $m$  computed using  $\rho = 1.1\rho_0$  and  $c = 1.1$ , we would expect the periodic steady state will not be established until perhaps twenty or more additional periods of ten time units have elapsed. By that time,  $m(t)$  will be oscillating

around a mean value of something in excess of 200 when  $\rho_0 = 0.9$ .<sup>4</sup>

In Figure 2.10b is compared  $m(t)$  computed exactly when using this closure method with  $\lambda_0 = 3.6$ . The approximate  $m(t)$  tracks the exact  $m(t)$  very well, giving an estimate of the exact  $m(t)$  which is slightly too large in this case over the time interval considered. At least over the first 2-1/2 periods, there is no sign of the accuracy of the approximation deteriorating. The largest gaps occur in the vicinity of the local maxima of  $m(t)$ , which occur at  $t = 5, 15, 25, \dots$ . When  $t=5$ , the approximation error is 1.442 (with  $m(5) = 15.167$ , this gives a relative error of 0.095), when  $t = 15$ , the approximation error is 1.515 (corresponding to a relative error of 0.065 since  $m(15) = 23.168$ ), and when  $t = 25$ , the approximation error is 1.308 (corresponding to a relative error of 0.045, since  $m(25) = 29.145$ ).

It is seen from Table 3a that  $\rho_0 = 0.9$  leads to the greatest mean absolute errors in  $m(t)$  over all of the test calculations represented in Table 3a, though not the greatest relative mean absolute errors because the mean values of  $m(t)$  over one period are much greater when  $\rho_0 = 0.9$  than for any of the other values of  $\rho_0$  for which calculations were done here.

---

<sup>4</sup>We can compare the period-averaged values of  $\Pi_0$  and  $m$  given in Table 3a with the corresponding steady-state values of  $\Pi_0$  and  $m$  computed using  $\rho = 1.1\rho_0$  and  $c = 1.1$ . For  $\Pi_0$  these latter are 0.89, 0.67, 0.45, 0.23 and 0.01, respectively, for  $\rho_0 = 0.1, 0.3, 0.5, 0.7$  and  $0.9$ . For  $m$ , the corresponding steady-state values are 0.2596, 1.034, 2.567, 7.030, and 207.9. The period-averaged values of  $\Pi_0(t)$  are similar to the steady state values obtained when  $\rho$  and  $c$  are period-averaged in those cases when  $\rho_0$  is small enough for our calculations to give an apparent periodic steady state. For  $m(t)$ , the similarity between the early time period-averages and steady-state values computed using mean values of  $\rho$  and  $c$  is less obvious from the data in Table 3a.

Figures 2.10c and 2.10d illustrate variations of relative errors in values of  $m(t)$  computed using this closure method. Plots of these relative errors for the cases  $\rho_0 = 0.7$  and  $0.9$  were done separately in Figure 2.10d to allow adoption of a larger scale and to make it easier to trace the rather complicated curves than would have been possible if all five relative error curves in these two figures had been plotted together. It is noteworthy that the local maxima in these relative error plots quite clearly decrease in value as  $t$  increases (though the magnitude of the local minima do increase in some cases shown -- the  $\lambda_0 = 1.2$  and  $\lambda_0 = 2.0$  cases being examples).

It would be interesting, though potentially very expensive computationally, to push these comparisons between exact and approximate for, say,  $\lambda_0 = 2.8$ , to much greater values of  $t$  to determine if these oscillatory relative errors tend to a mean of zero over a complete period of oscillation, with the extremes of the relative error over such a period tending to zero as well. In Figure 2.10d, there are two swings of the relative error in  $m(t)$  from relative maximum to relative minimum. The difference between these two successive pairs of extrema are 0.982 and 0.969 respectively, perhaps indicating a trend.

One other feature of Figures 2.10c and 2.10d is the complexity of these relative error curves. They are by no means primarily simple sine curves. This complexity seems to repeat itself each period though some smoothing out of irregularities is noticeable on comparing the portions of the curves in the second full period to their parts in the first full period. This complexity seems to be a

feature of relative error plots.

As mentioned earlier, a second series of oscillatory parameter calculations was carried out, keeping  $\lambda_0$  at 2.0,  $\mu(t)$  constant at 4.0, and considering the values  $v_0 = 10, 4, 2, 1, 0.4$ , and 0.2. These calculations were done to assess the accuracy of this closure method when the mean retrial rate changed relative to new customer arrival rates and service rates.

While the steady state value of  $\Pi_0$  does not depend upon the retrial rate for reasons already given,  $\Pi_0(t)$  does show some variation with  $v(t)$  in the oscillatory parameter case, as illustrated in Figure 2.11a. The server is more likely to be idle if orbiting customers attempt only rarely to obtain service than if they do so relatively frequently. The degree to which  $\Pi_0(t)$  is affected by varying  $v_0$  is not as great as when  $\lambda_0$  is varied.  $\Pi_0(t)$  is much more sensitive to the rate at which new customers enter the system than to the rate at which unsuccessful customers reapply for service. Figure 2.11b shows the variation of the relative error in the approximate  $\Pi_0(t)$  computed using the current closure method for the sequence of values of  $v_0$  considered. Figure 2.11c compares the approximate  $\Pi_0(t)$  with its exact counterpart when  $v_0 = 10$ . The worst correspondence between the two occurs in the vicinity of the local minima of  $\Pi_0(t)$  which tends to magnify local maxima of relative errors. The closure method gives local minima in  $\Pi_0(t)$  which are somewhat shallower, broader, and shifted compared to those of the exact  $\Pi_0(t)$ .

Corresponding results for  $m(t)$  computed for this sequence of values of  $\nu_0$  are depicted in Figures 2.12a-c. Other than the observation that relative errors in the approximate  $m(t)$  for the smaller values of  $\nu_0$  show very weak oscillatory behaviour as a function of time, these figures require little discussion.

Relative mean absolute errors are listed in Table 3b for this series of oscillatory parameter calculations corresponding to various retrial rates. The interpretation of these numbers corresponds to that applying to Table 3a, and the discussion will not be repeated.<sup>5</sup>

The relative mean absolute errors given in Tables 3a and 3b for oscillatory parameter calculations done using this closure method can be compared with their counterparts for  $p_0(t)$  and  $m(t)$  in Table 1 where the Rothkopf and Oren method has been applied to the simple M/M/1 queue. Generally the values in Table 1 are smaller than the corresponding ones in Tables 3a and 3b, in some cases by large factors. The relative mean absolute errors for  $m(t)$  when  $\rho_0 = 0.1$  are smaller in Table 1 by a factor of near 50 than those in Table 3a. However, in this instance, the relative mean absolute errors in  $m(t)$  from Table 3a are still under 2% when  $\rho_0 = 0.1$ , and so this large difference in accuracy is really the difference between an exceedingly accurate approximation and an acceptably accurate approximation. Where relative mean absolute errors become larger, this retrial queue closure method gives results which do not exceed the numbers in Table 1 by more than a

---

<sup>5</sup>For the calculations represented in Table 3b, the period-averaged value of  $\rho$  is 0.55. It would be very interesting to determine if for sufficiently large values of  $t$ , the period-averaged value of  $\Pi_0(t)$  approached  $\Pi_0 = 1 - \rho = 0.45$ . There is some support for expecting such a trend from the data in Table 3b, though the data given apply only to the first two full periods of oscillation.

factor of two, and in most cases, the correspondence between the relative mean absolute errors in the two sets of calculations is much better.

#### Computation of Period-Averages of $\Pi_0(t)$ and $m(t)$

Before leaving discussion of this two-variable closure method, we note the comparison of exact and approximate period-averaged values of  $\Pi_0(t)$  and  $m(t)$  given in Tables 4a and 4b. Exact and approximate values of  $\Pi_0(t)$  and  $m(t)$  were computed for integer and half-integer values of  $t$ . The period-averages were obtained from these equally time-spaced values using Simpson's rule. In a footnote in section 1.4 above, we have explained why we believe that this approach leads to no significant error in the numerical integration step, so that discrepancies between period-averages of exact quantities and period-averages of their approximations computed using this closure method are effectively totally attributable to the closure approximation.

It is seen from Tables 4a and 4b that these period averages can be computed very accurately from the closure method results. The relative error in the mean of  $\Pi_0(t)$  is far below 1% for most of the test values of  $\lambda_0$  and  $\nu_0$  and rises to 2.5% in the first ten-unit time period when  $\nu_0 = 0.2$  and 4.3% in the first such period when  $\rho_0 = 0.9$ . These cases are near the computationally most difficult end of the range of values  $\lambda$  and  $\nu$  may have. The relative errors in the period average of  $m(t)$  are quite uniform over the set of computations for various values of  $\nu_0$ , being in the 8% - 10% range for the first ten-unit time period, and in the 6% - 10% range for the second such period. With  $\nu_0$  fixed at

2.0, the relative errors in the period averaged  $m(t)$  vary with  $\lambda_0$  from a low of just under 1% when  $\rho_0 = 0.1$ , to a 6% - 9% range when  $\rho_0 = 0.5, 0.7$ , and  $0.9$ .

In both Tables 4a and 4b, the relative errors in the period averages of  $m(t)$  generally are smaller in the second ten-unit time period than in the first. This occurs in part because these averages are increasing in successive periods as the system evolves from its initial empty state to its eventual periodic steady state. The absolute errors are also given in these tables and are seen to be relatively unchanged in the two periods for which numbers were obtained and in most cases slightly larger in the second period compared to the first. The same trend in relative errors is not observed for the period averages of  $\Pi_0(t)$ , which generally decrease slowly through successive periods.

**TABLE 3a**      **Exact Period-Mean Values of  $\Pi_0(t)$  and  $m(t)$  and Mean Absolute and Relative Mean Absolute Errors When These Quantities are Computed Using Closure Method 2C for Various Values of  $\rho_0$ .**

The closure approximation values of  $\Pi_0(t)$  and  $m(t)$  on which these numbers are based were calculated using Method 2C, equations (2.3.1) and (2.3.6).

The first column gives the value of  $\rho_0$ . The second gives the time interval to which the computed mean values apply. The next two columns give the exact mean values of  $\Pi_0(t)$  and  $m(t)$  respectively over these time intervals, the next two give the mean absolute errors in the closure approximation estimate of these quantities over the indicated time interval (computed as integrated absolute errors divided by the interval length), and the last two columns give the corresponding relative mean absolute errors (computed as the mean absolute error divided by the mean value of the quantity in question).

		Exact Mean Value		Mean Absolute Error		Relative Mean Absolute Error	
$\rho_0$		$(\Pi_0)$	$(m)$	$(\Pi_0)$	$(m)$	$(\Pi_0)$	$(m)$
0.1	[0,10]	.8930	.1509	.0005177	.002962	.000579	.01962
	[10,20]	.8900	.1571	.0004485	.002699	.000504	.01718
0.3	[0,10]	.6840	.8979	.007470	.03718	.01092	.04141
	[10,20]	.6701	.9789	.008760	.03943	.01307	.04028
0.5	[0,10]	.4969	2.7446	.02868	.2226	.05771	.08112
	[10,20]	.45595	3.4260	.02946	.2257	.06461	.06588
0.7	[0,10]	.3560	5.9320	.05283	.5610	.1484	.09457
	[10,20]	.2799	9.0347	.05231	.5591	.1869	.06189
0.9	[0,10]	.2638	10.1175	.06480	.8653	.2457	.08552
	[10,20]	.1717	18.0434	.05733	1.0238	.3340	.05674

**TABLE 3b**      **Exact Period-Mean Values of  $\Pi_0(t)$  and  $m(t)$  and Mean Absolute and Relative Mean Absolute Errors When These Quantities are Computed Using Closure Method 2C for Various Values of  $c_0$  or  $v_0$ .**

The closure approximation values of  $\Pi_0(t)$  and  $m(t)$  on which these numbers are based were calculated using Method 2C, equations (2.3.1) and (2.3.6).

The first column gives the value of  $v_0$ , the mean retrial rate, and the second column gives the value of  $c_0 = \lambda_0 / v_0$ . The third column gives the time interval to which the computed mean values apply. The next two columns give the exact mean values of  $\Pi_0(t)$  and  $m(t)$  respectively over these time intervals, the next two give the mean absolute errors in the closure approximation estimate of these quantities over the indicated time interval (computed as integrated absolute errors divided by the interval length), and the last two columns give the corresponding relative mean absolute errors (computed as the mean absolute error divided by the mean value of the quantity in question).

$v_0$	$c_0$		Exact Mean Value		Mean Absolute Error		Relative Mean Absolute Error	
			$(\Pi_0)$	$(m)$	$(\Pi_0)$	$(m)$	$(\Pi_0)$	$(m)$
10.0	0.2	[0,10]	.4734	1.9893	-.002197	.1648	.08629	.1427
		[10,20]	.4505	2.1655	-.0004056	.1346	.09539	.1285
4.0	0.5	[0,10]	.4821	2.3078	-.002023	.1852	.06878	.1075
		[10,20]	.4519	2.6483	-.0008732	.1459	.07753	.09209
2.0	1.0	[0,10]	.4969	2.7446	.00006752	.2224	.05771	.08112
		[10,20]	.4559	3.4260	-.001340	.1992	.06461	.06588
1.0	2.0	[0,10]	.5230	3.3794	.005831	.2816	.05080	.08332
		[10,20]	.4676	4.7900	-.001536	.3062	.05807	.06392
0.4	5.0	[0,10]	.5714	4.3977	.008987	.4167	.04722	.09476
		[10,20]	.5032	7.5169	.0002891	.5898	.05478	.07846
0.2	10.0	[0,10]	.6088	5.1257	.01527	.5568	.04565	.1086
		[10,20]	.5434	9.8624	.005137	.9539	.05165	.09672

**TABLE 4a**      **Period Averages of  $\Pi_0(t)$  and  $m(t)$  for Various New Customer Arrival Rates, Based on Method 2C.**

This table gives the exact period-averages of  $\Pi_0(t)$  and  $m(t)$  (*third column*), the corresponding averages of the same quantities calculated using the two-equation closure approximation 'C' (*fourth column*), and the actual errors (*fifth column*) and errors relative to the exact values of these approximate values (*sixth column*) for various values of  $\rho_0$ . For these calculations,  $v_0$  was fixed at 2.0, and  $\mu(t)$  was constant in time with value 4.0. The second column indicates the time interval over which the averages were computed, and represent complete periods of oscillation in the new customer arrival rate and retrial rate.

$\Pi_0(t) :$		Exact Mean	Approximate Mean	Absolute Error in Mean	Relative Error in Mean
$\rho_0$					
0.1	[0, 10]	.8930	.8930	.00001828	.00002047
	[10, 20]	.8900	.8900	-.00000025	-.000000281
0.3	[0, 10]	.6840	.6842	.0001832	.0002678
	[10, 20]	.6701	.6713	.001154	.001722
0.5	[0, 10]	.4969	.4970	.00006752	.0001359
	[10, 20]	.45595	.45461	-.001340	-.002938
0.7	[0, 10]	.3560	.3596	.003644	.01024
	[10, 20]	.2799	.2757	-.004167	-.01489
0.9	[0, 10]	.2638	.2752	.01144	.043355
	[10, 20]	.1717	.1708	-.0008418	-.004904
$m(t) :$					
$\rho_0$					
0.1	[0, 10]	.1509	.1521	.001120	.007419
	[10, 20]	.1571	.1585	.001488	.009475
0.3	[0, 10]	.8979	.9284	.03057	.03405
	[10, 20]	.9789	1.0101	.03118	.03185
0.5	[0, 10]	2.7446	2.9671	.2224	.08104
	[10, 20]	3.4260	3.6252	.1992	.05813
0.7	[0, 10]	5.9320	6.4930	.5572	.09457
	[10, 20]	9.0347	9.59196	.5610	.06168
0.9	[0, 10]	10.1175	10.9828	.8653	.08552
	[10, 20]	18.0434	19.0672	1.0238	.05674

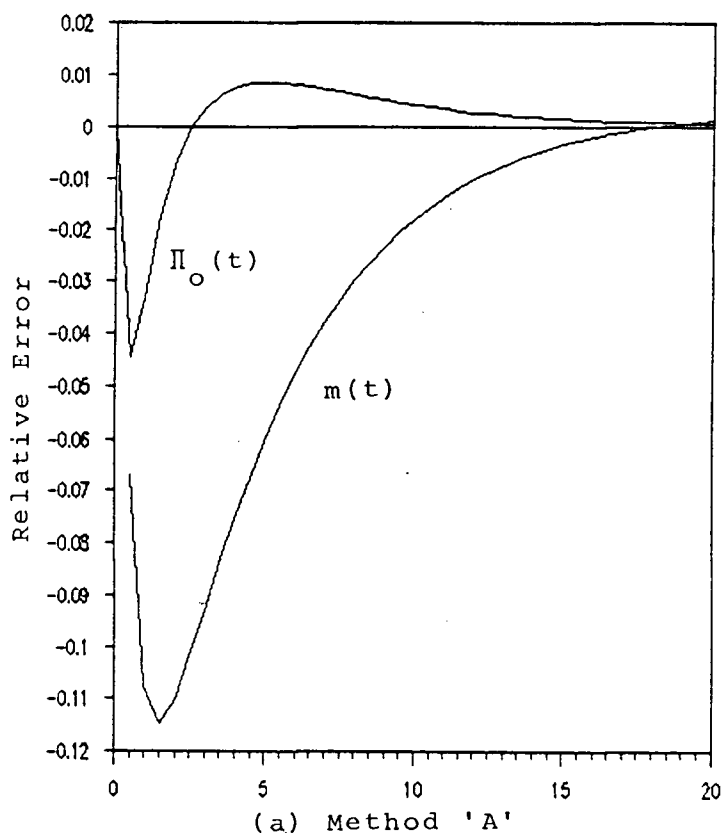
**TABLE 4b**      **Period Averages of  $\Pi_0(t)$  and  $m(t)$  for Various Retrial Rates for the Single-Server Retrial Queue, Based on Method 2C.**

This table gives the exact period-averages of  $\Pi_0(t)$  and  $m(t)$  (*fourth column*), the corresponding averages of the same quantities calculated using the two-equation closure approximation 'C' (*fifth column*), and the actual errors (*sixth column*) and errors relative to the exact values of these approximate values (*seventh column*) for various values of  $\nu_0$ , the mean retrial rate for orbiting customers. For these calculations,  $\lambda_0$  was fixed at 2.0, and  $\mu(t)$  was constant in time with value 4.0. The third column indicates the time interval over which the averages were computed, which comprise complete periods of oscillation in the new customer arrival rate and retrial rate. The second column gives  $c_0 = \lambda_0 / \nu_0$ .

$\Pi_0(t) :$			Exact	Approximate	Absolute Error	Relative Error
$\nu_0$	$c_0$		Mean	Mean	in Mean	in Mean
10.0	0.2	[0, 10]	.4734	.4711	-.002197	-.004640
		[10, 20]	.4505	.4501	-.0004056	-.0009004
4.0	0.5	[0, 10]	.4821	.4801	-.002023	-.004195
		[10, 20]	.4519	.4510	-.0008732	-.001932
2.0	1.0	[0, 10]	.4969	.4970	.0006753	.0001359
		[10, 20]	.4559	.4546	-.001340	-.002915
1.0	2.0	[0, 10]	.5230	.5262	.005831	.01121
		[10, 20]	.4676	.4661	-.001536	-.003284
0.4	5.0	[0, 10]	.5714	.5804	.008987	.01573
		[10, 20]	.5032	.5035	.0002891	.0005746
0.2	10.0	[0, 10]	.6088	.6241	.01527	.02508
		[10, 20]	.5434	.5485	.005137	.009453

$m(t) :$

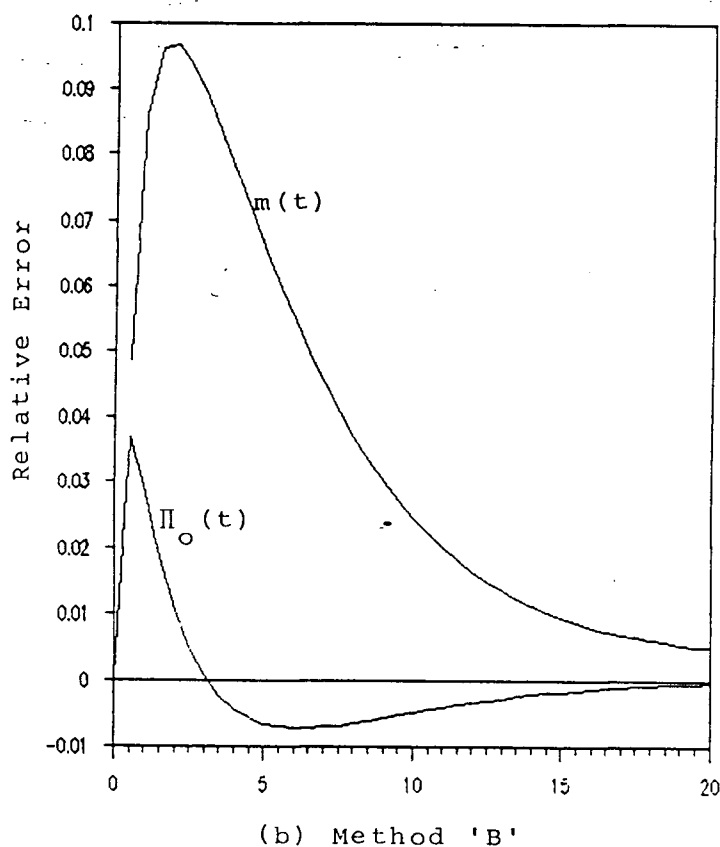
$\nu_0$	$c_0$					
10.0	0.2	[0, 10]	1.9893	2.1541	.1648	.08285
		[10, 20]	2.1655	2.3001	.1346	.06218
4.0	0.5	[0, 10]	2.3078	2.4930	.1852	.08025
		[10, 20]	2.6483	2.7942	.1459	.05508
2.0	1.0	[0, 10]	2.7446	2.9671	.2224	.08104
		[10, 20]	3.4260	3.6252	.1992	.05813
1.0	2.0	[0, 10]	3.3794	3.6609	.2816	.08332
		[10, 20]	4.7900	5.0962	.3062	.06392
0.4	5.0	[0, 10]	4.3977	4.8145	.4167	.09476
		[10, 20]	7.5169	8.1067	.5898	.07846
0.2	10.0	[0, 10]	5.1257	5.6826	.5568	.1086
		[10, 20]	9.8624	10.8164	.9539	.09672

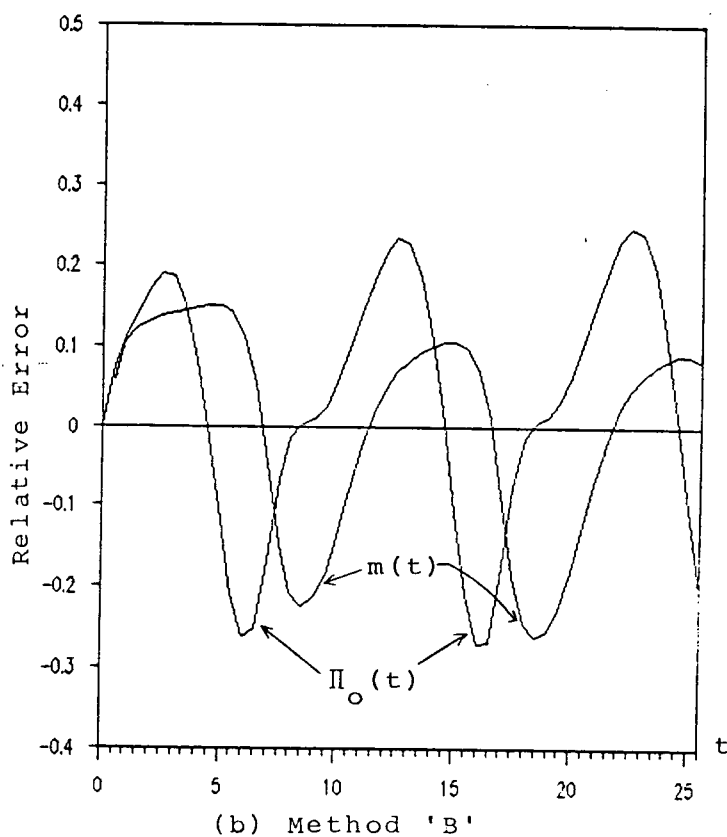
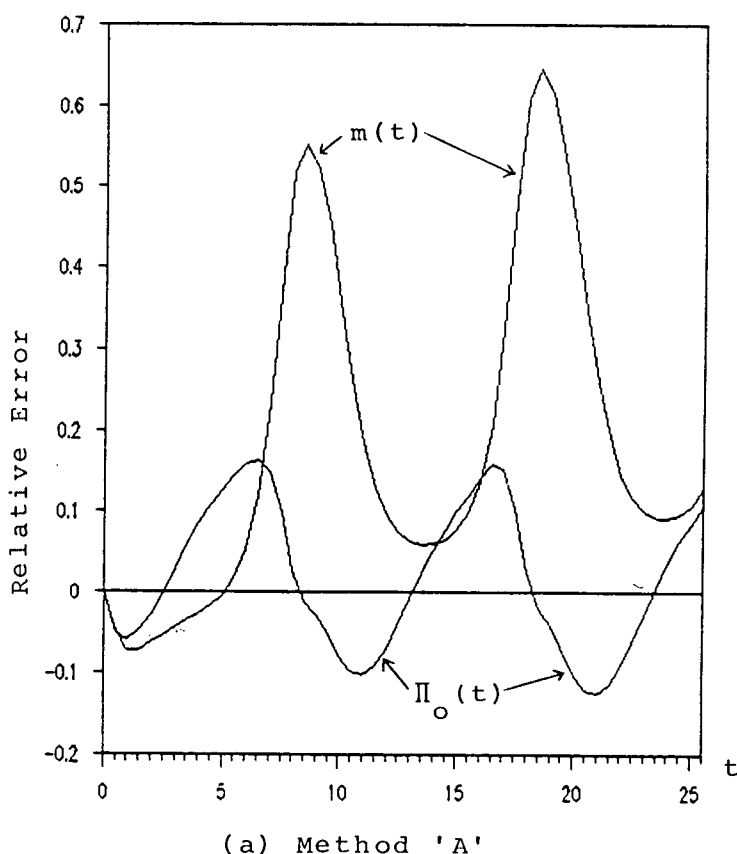


**FIGURE 2.5** Relative Errors in  $\Pi_0(t)$  and  $m(t)$  When Computed Using a Two-Equation Closure Approximation With Basic Methods 'A' and 'B' in the Constant Parameter Case.

Figure 2.5a results from a calculation using equation (2.3.4) to eliminate  $m_1(t)$  (the so-called method 'A'), whereas Figure 2.5b results from a calculation using equation (2.3.5) (the so-called method 'B'). What is noteworthy here is not so much the magnitudes of the relative errors as the fact that the two figures are very nearly reflections of each other in the  $t$ -axis. This observation suggests the hybrid method 'C' described in section 2.3.

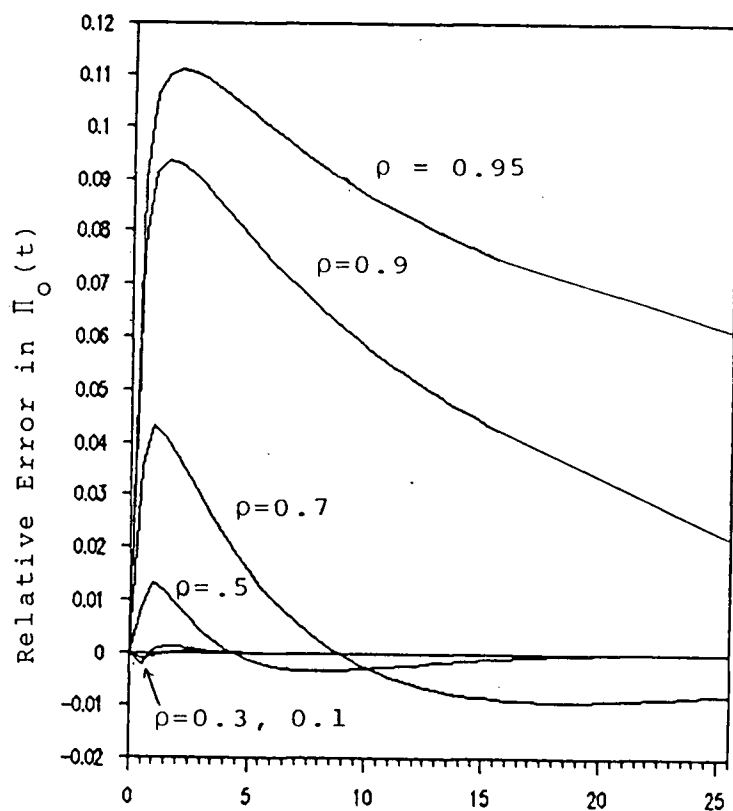
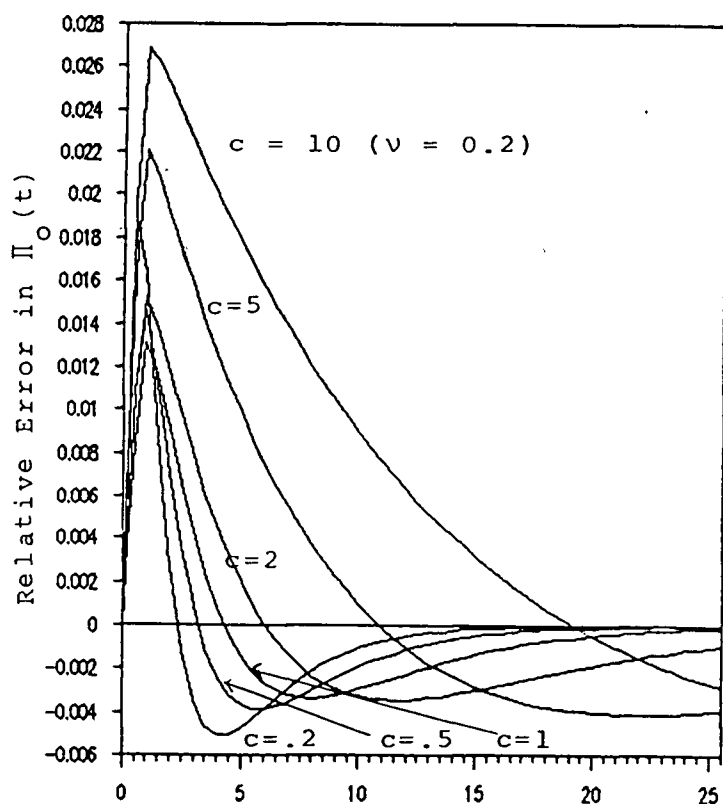
Calculations here correspond to  $\lambda = 2.0$ ,  $\mu = 4.0$ , and  $\nu = 2.0$ .





**FIGURE 2.6** Relative Errors in  $\Pi_O(t)$  and  $m(t)$  When Computed Using a Two-Equation Closure Approximation With Basic Methods 'A' and 'B' in the Oscillatory Parameter Case.

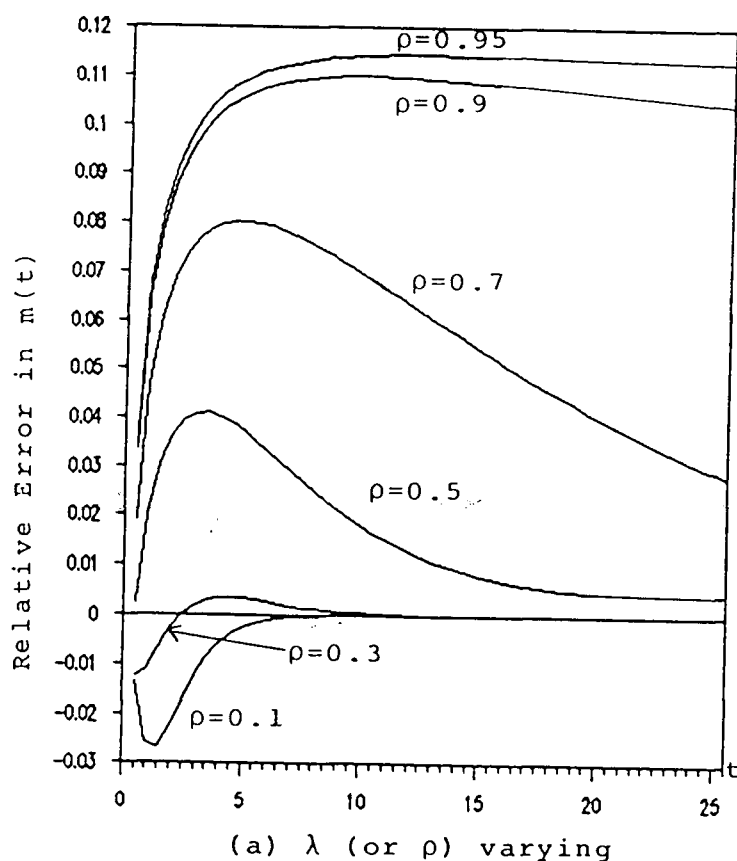
These two figures are the oscillatory parameter counterparts of Figures 2.5a and 2.5b, corresponding to  $\lambda_O = 2.0$ ,  $\nu_O = 2.0$  and  $\mu(t) = 4.0$ . Because these relative error curves oscillate, it is not so easy to see at a glance that those in Figure 2.6a go through their maximal and minimal values in the regions where those in Figure 2.6b go through their minimal and maximal values respectively. However, comparison of the  $t$ -coordinates of these local extrema indicates that this correspondence in their locations is virtually exact. The shapes of the corresponding curves here are not as nearly mirror images as in Figures 2.5a and 2.5b. However, the overall range in values of relative errors observed for each quantity is not much different between Figure 2.6a and Figure 2.6b if one ignores the first five or so units of time. It is also noteworthy that the maximum magnitudes of relative errors in this oscillatory parameter case are two to six times those observed for the corresponding constant parameter cases.

(a)  $\lambda$  (or  $\rho$ ) varying(b)  $v$  (or  $c$ ) varying

**FIGURE 2.7** Relative Errors in  $\Pi_0(t)$  Using the Two-Equation Closure Method 'C' in the Constant Parameter Case With Various Arrival and Retrial Rates.

Calculations represented in Figure 2.7a were done with  $\mu(t) = 4.0$  and  $v(t) = 2.0$  in all cases, with  $\lambda(t)$  taking on the values 0.4, 1.2, 2.0, 2.8, 3.6, and 3.8 in sequence. Calculations represented in Figure 2.7b were done with  $\mu(t) = 4.0$  and  $\lambda(t) = 2.0$  in all cases, with  $v(t)$  taking on the values 10.0, 4.0, 2.0, 1.0, 0.4, and 0.2 in sequence. The  $\rho = 0.5$  calculation in Figure 2.7a is the same as the  $c = 1$  calculation in Figure 2.7b, corresponding to the calculations shown in Figure 2.5 for  $\Pi_0(t)$ . Note that the maximum relative error here is smaller by a factor of three to four for  $\Pi_0(t)$  in that particular case.

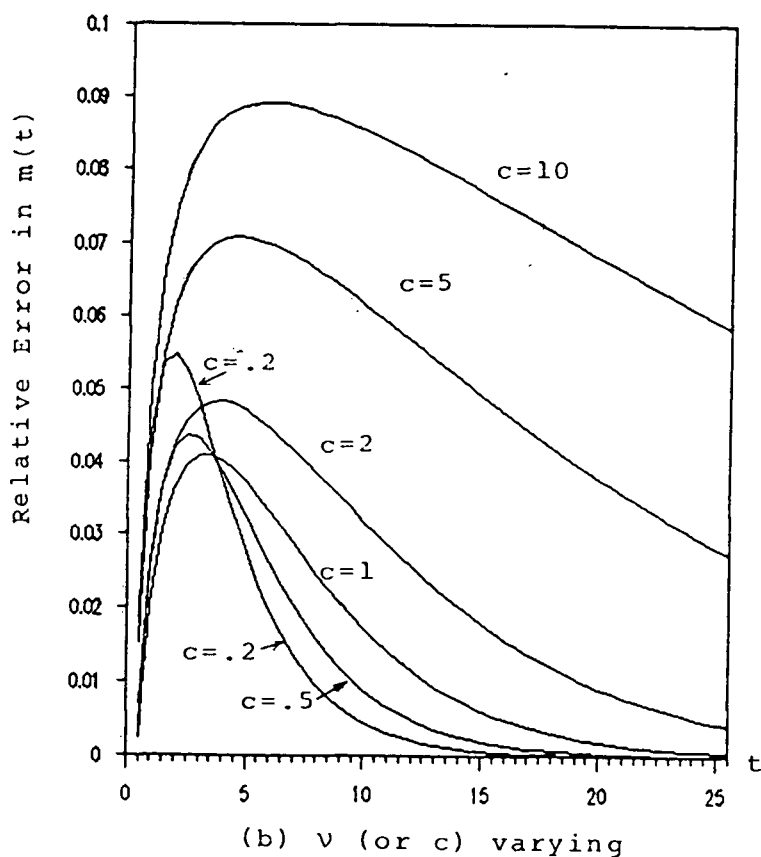
Appendix A gives details of calculations done to obtain exact values required in determining these relative errors.

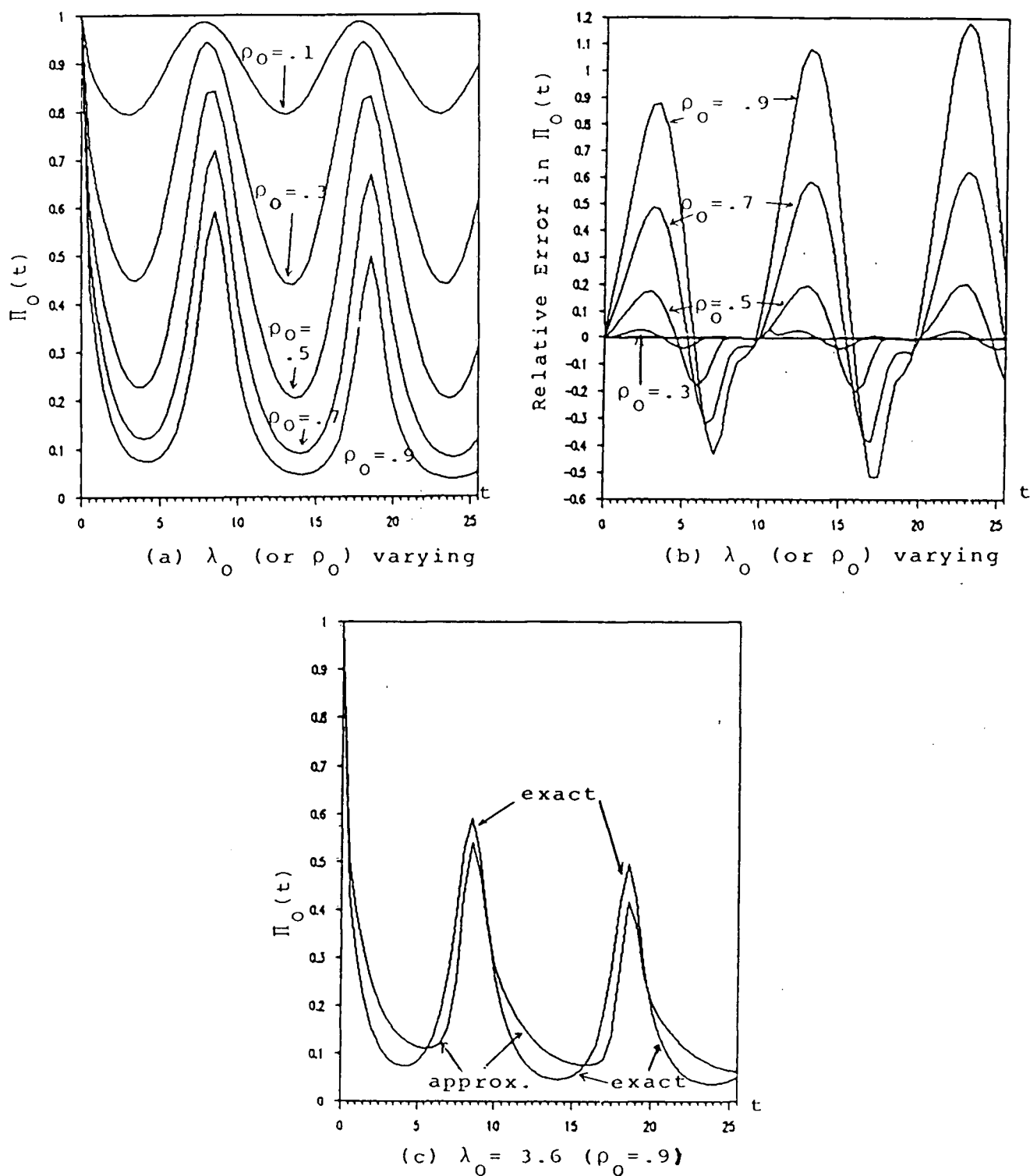


**FIGURE 2.8** Relative Errors in  $m(t)$  Using the Two-Equation Closure Method 'C' in the Constant Parameter Case With Various Arrival and Retrieval Rates.

Calculations represented in Figure 2.8a were done with  $\mu(t) = 4.0$  and  $\nu(t) = 2.0$  in all cases, with  $\lambda(t)$  taking on the values 0.4, 1.2, 2.0, 2.8, 3.6, and 3.8 in sequence. Calculations represented in Figure 2.8b were done with  $\mu(t) = 4.0$  and  $\lambda(t) = 2.0$  in all cases, with  $\nu(t)$  taking on the values 10.0, 4.0, 2.0, 1.0, 0.4, and 0.2 in sequence. The  $\rho = 0.5$  calculation in Figure 2.8a is the same as the  $c = 1$  calculation in Figure 2.8b, corresponding to the calculations shown in Figure 2.5 for  $m(t)$ . Note that the maximum relative error here is smaller by a factor of 2.5 to 3 for  $m(t)$  in that particular case.

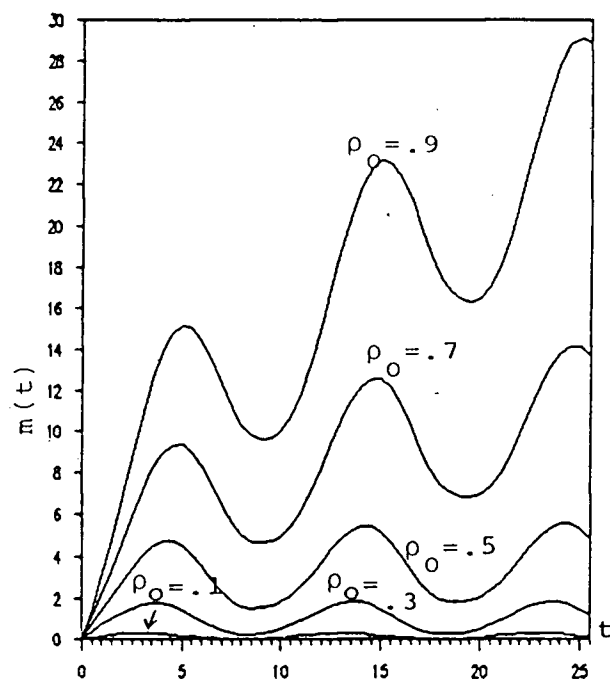
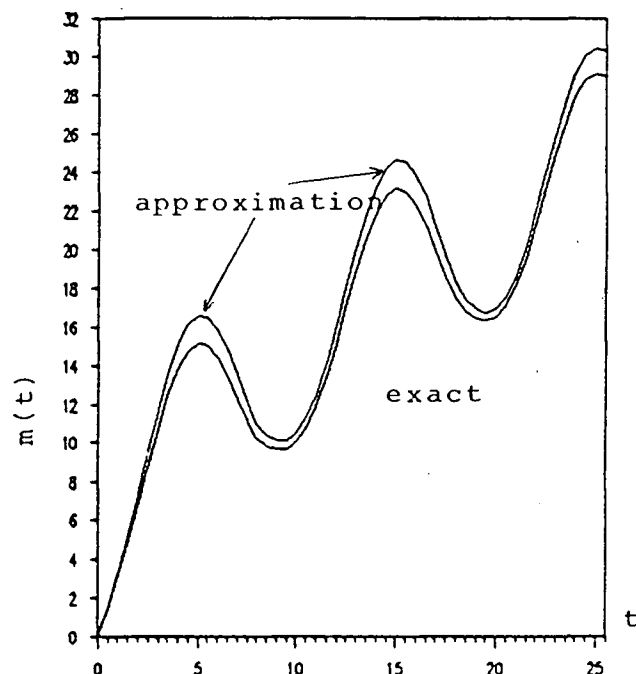
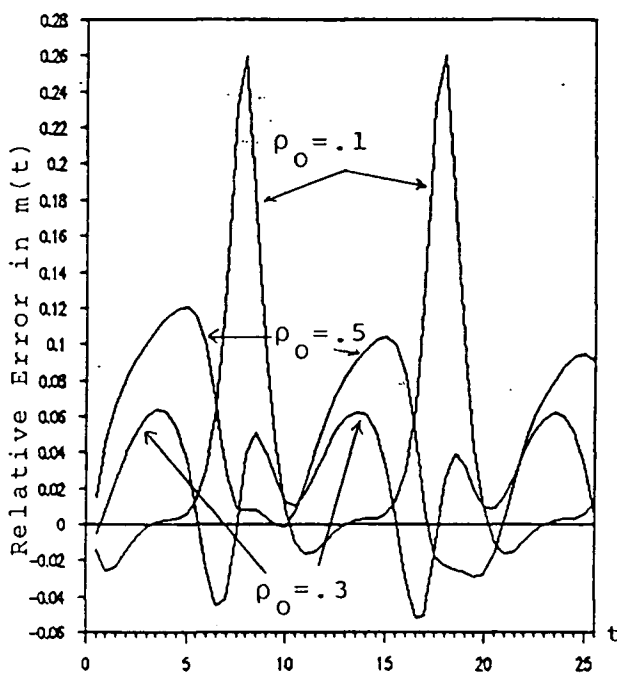
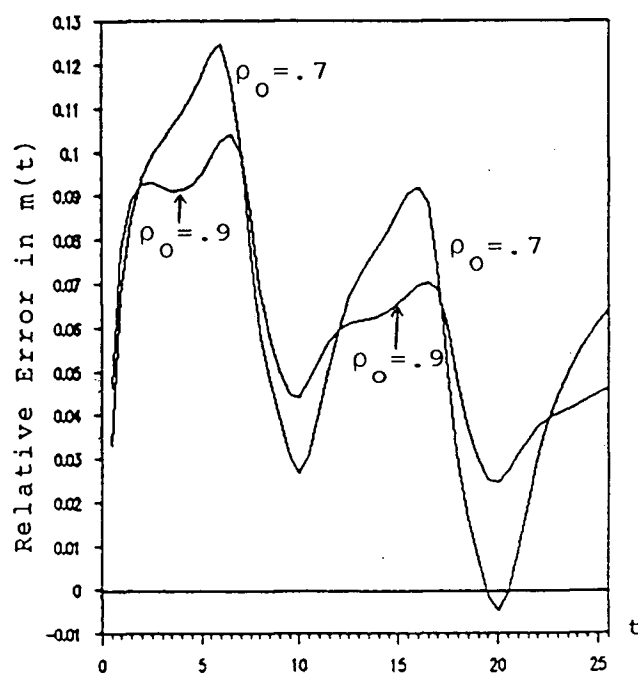
Appendix A gives details of calculations done to obtain exact values required in determining these relative errors.





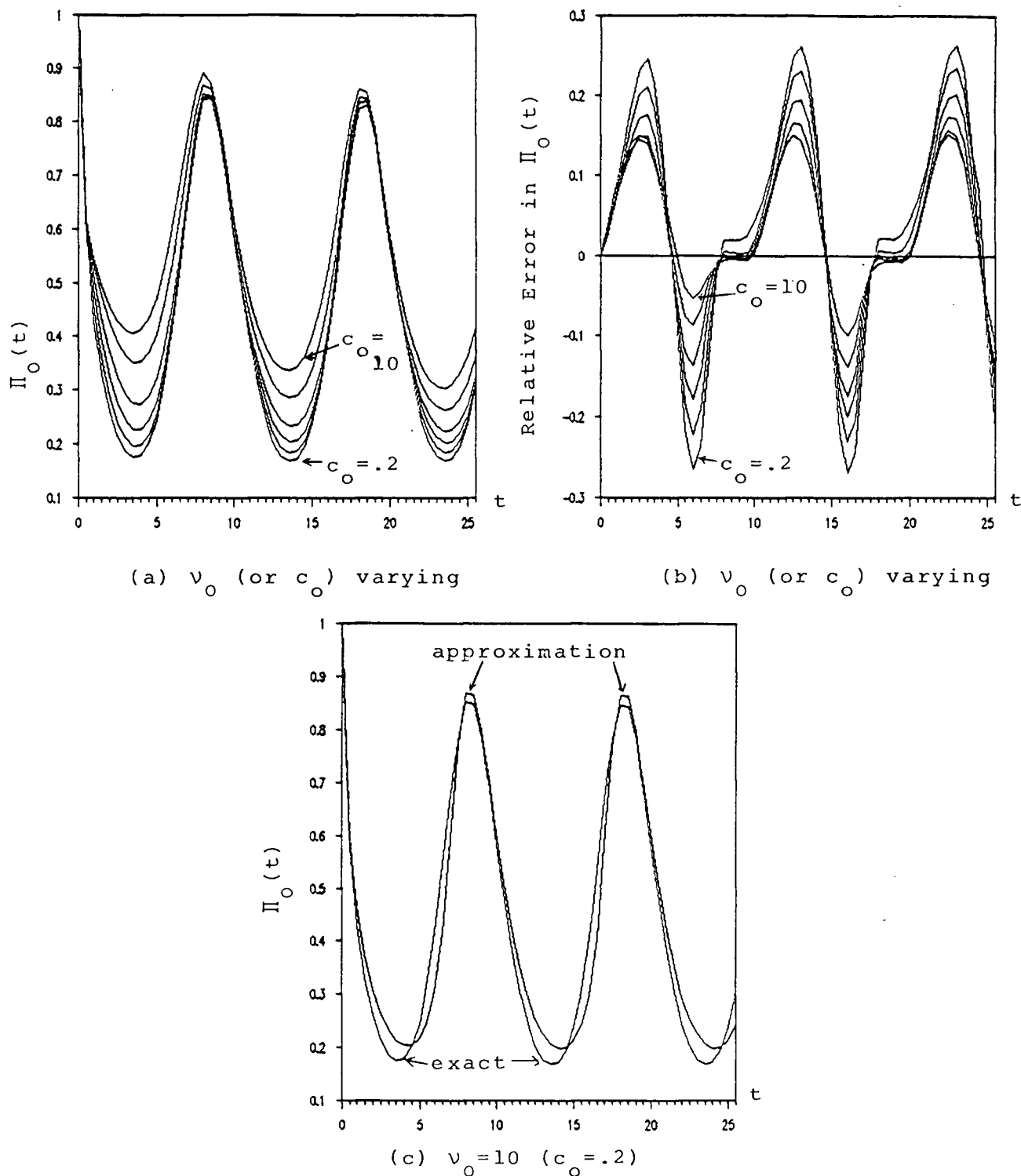
**FIGURE 2.9**  $\Pi_0(t)$  Computed Using the Two-Equation Closure Method 'C' in the Oscillatory Parameter Case for Various Values of  $\rho_0$ .

Figure 2.9a simply displays the exact  $\Pi_0(t)$  for various values of  $\rho_0$ . Figure 2.9b displays the corresponding relative error in  $\Pi_0(t)$  calculated using closure method 2C. Figure 2.9c compares the exact and approximate  $\Pi_0(t)$  when  $\rho_0 = 0.9$ . Note that on the scale employed in Figure 2.9b, the relative error curve for  $\rho_0 = 0.1$  is indistinguishable from the  $t$ -axis. Appendix A gives details of calculations done to obtain exact values used in preparing these plots.

(a)  $\lambda_0$  (or  $\rho_0$ ) varying(b)  $\lambda_0 = 3.6$  ( $\rho_0 = 0.9$ )(c)  $\rho_0 = 0.1, 0.3, 0.5$ (d)  $\rho_0 = 0.7, 0.9$ 

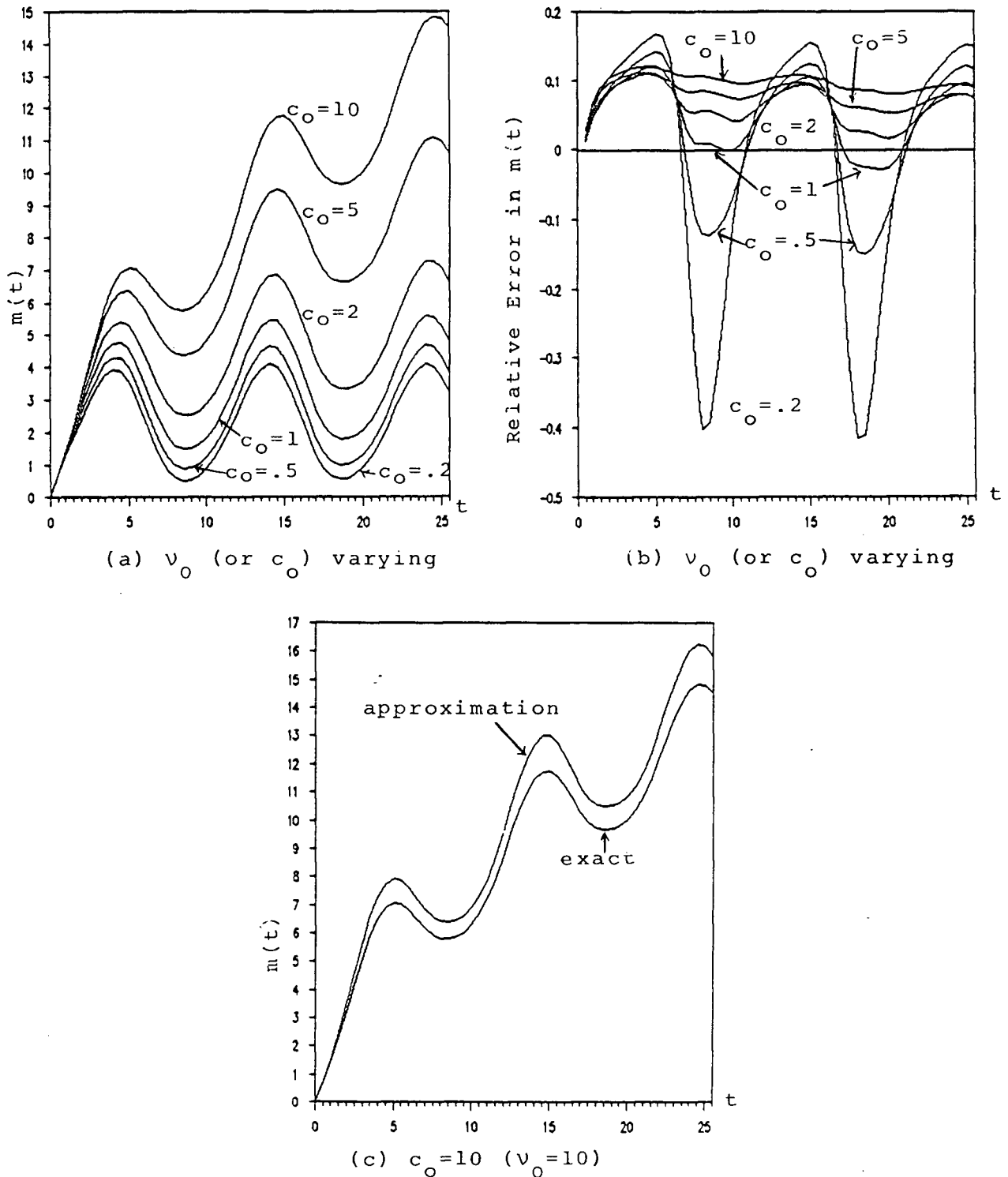
**FIGURE 2.10**  $m(t)$  Computed Using the Two-Equation Closure Method 'C' in the Oscillatory Parameter Case for Various Values of  $\rho_0$ .

Figure 2.10a simply displays the exact  $m(t)$  for various values of  $\rho_0$ . Figure 2.10b compares the closure approximation  $m(t)$  with its exact counterpart when  $\rho_0 = 0.9$ . The plots of relative errors in the closure approximation  $m(t)$  are displayed in Figures 2.10c and 2.10d. Appendix A gives details of calculations done to obtain exact values required in determining these relative errors.



**FIGURE 2.11**  $\Pi_0(t)$  Computed Using the Two-Equation Closure Method 'C' in the Oscillatory Parameter Case for Various Values of  $c_0$ .

Figure 2.11a shows the exact  $\Pi_0(t)$  for various values of  $c_0$ . At the local minima,  $c_0 = 0.2$  corresponds to the lowest curve and  $c_0 = 10$  to the highest. Figure 2.11b shows corresponding relative errors in the closure approximation  $\Pi_0(t)$ , with the widest extremes corresponding to  $c_0 = 0.2$ , and the narrowest to  $c_0 = 10$ . In Figure 2.11c is compared the exact and approximate  $\Pi_0(t)$  in the worst case  $c_0 = 0.2$ .



**FIGURE 2.12**  $m(t)$  Computed Using the Two-Equation Closure Method 'C' in the Oscillatory Parameter Case for Various Values of  $c_0$ .

Figure 2.12a shows the exact  $m(t)$  for various values of  $c_0$ . Throughout,  $c_0 = 0.2$  corresponds to the lowest curve and  $c_0 = 10$  to the highest. Figure 2.12b shows corresponding relative errors in the closure approximation  $m(t)$ , with the curve having the lowest local minima corresponding to  $c_0 = 0.2$ , and the one with the highest local minima corresponding to  $c_0 = 10$ . In Figure 2.12c is compared the exact and approximate  $m(t)$  when  $c_0 = 10.0$ .

## 2.4 Two-Equation Closure Methods: $\Pi_0$ , $m$ , and $v$

*Then he got up on top with a tip of his hat.  
'I call this game FUN-IN-A-BOX', said the cat.  
'In this box are two things I will show to you  
now.*

*You will like these two things', said the cat  
with a bow.*

[The Cat in the Hat, Dr. Seuss]

### Method 2D ( $\Pi_0$ , $m$ , $v$ )

Because the steady-state probabilities for the single-server retrial queue form the negative binomial distribution, it is possible to write down an analogue of the Rothkopf and Oren method using the following equations from Table 2:

$$\begin{aligned} m'(t) &= \lambda - \mu \{1 - \Pi_0(t)\}, \\ v'(t) &= \lambda + \mu - \mu \Pi_0(t) \{2m(t) + 1\} + 2\mu m_0(t). \end{aligned} \quad (1)$$

As before,  $m(t)$  is the mean number of customers in the system at time  $t$ ,  $m_0(t)$  is the contribution to  $m(t)$  by server idle terms,  $v(t)$  is the variance of the number of customers in the system, and  $\Pi_0(t)$  is the probability that there are no customers in the system at time  $t$ . At steady state,

$$\Pi_0 = 1 - \rho, \quad m_0 = c\rho, \quad m = \frac{\rho(c + 1)}{1 - \rho}$$

and (2)

$$v = \frac{\rho(c + 1)}{(1 - \rho)^2}$$

The object here is to eliminate  $\Pi_0(t)$  and  $m_0(t)$  in equations (1) to obtain a system of two equations which can be used to compute values of  $m(t)$  and  $v(t)$ , at least approximately.

Following Rothkopf and Oren, we note that

$$p_{00} = (1 - \rho)^{c+1} = \left( \frac{m}{v} \right)^{m^2 / (v-m)} \quad (3)$$

Thus,

$$\Pi_0 = (1 - \rho) = p_{00}^{1/(c+1)} = \left( \frac{m}{v} \right)^{m^2 / \{ (v-m) (c+1) \}} \quad (4)$$

When  $v - m = 0$ , we follow Rothkopf and Oren in writing

$$\Pi_0 = e^{-m/(c+1)} \quad (5)$$

In the simple M/M/1 case, the last term in the equation above for  $v'(t)$  does not appear.

Examination of the formulas in (2) easily yield a number of relationships for  $m_0$  at steady state. Among these are

$$m_0 = c\rho = c(1 - \Pi_0), \quad (6)$$

obtained by direct replacement of  $\rho$  by  $1 - \Pi_0$ ;

$$m_0 = m\Pi_0 - \Pi_1, \quad (7)$$

obtained by replacing  $c\rho$  by  $m_0$  in the numerator of the formula for  $m$ , solving for  $m_0$ , and replacing  $\rho$  by  $\Pi_1$  or  $1 - \rho$  by  $\Pi_0$ ; and finally,

$$m_0 = v\Pi_0^2 - \Pi_1, \quad (8)$$

obtained as is (7) but starting with the formula for  $v$ .

Formula (7) has the most intuitive appeal, since it relates  $m_0(t)$  in a direct way to  $m(t)$  and  $\Pi_0(t)$ . We would expect  $m_0(t)$  to increase with both  $m(t)$  and with  $\Pi_0(t)$  in general. Unfortunately, substitution of (7) into the right-hand side of the equation for  $v'(t)$  causes it to become identical to the right-hand side of the equation for  $m'(t)$  in (1), which is unrealistic. The subtracted terms in (7) and (8) are also cause for concern, and computational tests quickly indicate that using (8),  $v'(t)$  is prone to become negative when  $v$  is near zero (as is the usual case at the beginning of a calculation when the system is in a known definite state). This leads to  $v(t)$  becoming negative, and so  $(m/v)$  is negative, and the right hand side of (3) is not real.

As a result, we will initially consider using only formula (6) in conjunction with (5) and (1) to comprise a closure method for computing approximations to  $m(t)$  and  $v(t)$ . We will call this approach Method 2D.

### Numerical Performance of Method 2D

A series of calculations was carried out using Method 2D. The results were not very encouraging. In constant parameter calculations, relative error maxima were several times larger than they were for corresponding M/M/1 queue calculations using the Rothkopf and Oren method. This is best seen by comparing Figures 2.13a, 2.13b, and 2.13c directly with Figures 1.4, 1.5, and 1.6. Figure 2.13 shows the relative errors in  $\Pi_0$ ,  $m$ , and  $v$ , as functions of time for a sequence of new customer arrival rates.

For the calculations represented in Figure 2.14, the new customer arrival rate is held constant and the retrial rate is varied. When the retrial rate is comparable or smaller than the new customer arrival rate, relative accuracy over the time range shown in the figures is poor. Of particular interest are Figures 2.13c and 2.14c, dealing with the variance, since no other closure technique has been discovered for computing this quantity for the retrial queue so far.

Figure 2.15a shows the exact  $v(t)$  and  $v(t)$  obtained using this proposed closure method when  $\rho = 0.5$  and  $c = 10$  (actually  $\lambda = 2.0$ ,  $\mu = 4.0$ , and  $v = 0.2$ ). This figure gives an alternative impression of the approximation errors arising.

The situation is much the same for oscillatory parameter calculations using this approach. Figures 2.15b, 2.15c, and 2.15d contain plots of the exact  $v(t)$  and the approximate  $v(t)$  for the most difficult cases studied ( $\rho_o = 0.7$ ,  $c_o = 1.0$ , in Figure 2.15b;  $\rho_o = 0.9$ ,  $c_o = 1.0$ , in Figure 2.15c; and  $\rho_o = 0.5$ ,  $c_o = 10.0$  in Figure 2.15d). The approximation errors in  $\Pi_o(t)$  and  $m(t)$  computed using this method are not as relatively large as for  $v(t)$ . However, they are significantly greater than when the best of the methods developed in the preceding section are used.

Figures 2.16a and 2.16b show the approximate  $\Pi_o(t)$  along with its exact values in the oscillatory parameter case when  $\lambda_o = 3.8$ ,  $\mu = 4.0$ ,  $v_o = 2.0$ , and when  $\lambda_o = 2.0$ ,  $\mu = 4.0$ , and  $v_o = 0.2$ , respectively. The corresponding comparisons of approximate  $m(t)$  with exact  $m(t)$  in these two worst-of-the-cases-

studied situations are displayed in Figures 2.16c and 2.16d. Figures 2.16a and 2.16c go with Figure 2.15c, and Figures 2.16b and 2.16d go with Figure 2.15d.

In Table 5 are given period averages of  $\Pi_0(t)$ ,  $m(t)$ , and  $v(t)$  as well as relative mean absolute errors. These numbers indicate further the breakdown of the accuracy of this approach as the problematic large  $\rho_0$  or large  $c_0$  situation is approached. Error measures in Table 5 should be compared with corresponding quantities in Tables 4 and 1.

#### Method 2E ( $\Pi_0$ , $m$ , $v$ )

There are two approximations in the method just described. The first is to use (4) to express  $\Pi_0(t)$  in terms of  $m(t)$  and  $v(t)$ . This rather complicated formula is unlikely to be a significant source of error since its analogue in the Rothkopf and Oren method for simple M/M/1 queues worked extremely well. Also, relative errors in  $\Pi_0(t)$  computed using this method are much smaller generally than for  $m(t)$  and  $v(t)$ .

The second approximation in the computations just described was in using equation (6) to eliminate  $m_0(t)$  from the formula for  $v'(t)$ . Equation (6) was obtained simply by substituting  $1 - \Pi_0(t)$  for  $\rho$  in the steady-state formula for  $m_0$ , and so is analogous to formula (2.3.5) in the previous section. For the closure method developed in section 2.3, there was a second expression, equation (2.3.4), which expressed the component  $m_1(t)$  in terms of  $m(t)$ . The analogue of equation (2.3.4) in the present context is equation (7), which, as already

explained, cannot be used by itself to eliminate  $m_0(t)$  in equations (1). However, in section 2.3, it was found that a superior approximation for  $m_1(t)$  was obtained by combining formulas (2.3.4) and (2.3.5). And, in fact, the difficulty in using equation (7) here does not arise if a combination of (6) and (7) is used:

$$m_0(t) = \eta c(1 - \Pi_0) + (1 - \eta)\{m\Pi_0 - \Pi_1\} \quad , \quad (9)$$

$$0 < \eta < 1 \quad .$$

This formula, (9), is the analogue of equation (2.3.6) in the preceding section. There, the value  $\eta = 0.5$  was adopted somewhat arbitrarily. Here we will look a little more closely at the effect of varying the mixing factor  $\eta$  because of the large errors which can arise in the approximate  $v(t)$  when  $\eta = 1$ .

Of course, one could stretch the reader's indulgence further by starting with a mixture of formulas (6), (7), and (8), but the author's advancing age precludes an adequate study of such a scheme. In addition, formula (8) would not appear to have a strong intuitive appeal, and as well, contains  $v(t)$ , which is a quantity we are not able to estimate very accurately at present.

#### Constant Parameter Calculations Using Method 2E

The calculations already described in this section correspond to  $\eta = 1$ . The highest  $\rho$  or  $\rho_0$ , and the highest  $c$  or  $c_0$  (corresponding to smallest  $v$  or  $v_0$ ) calculations and the intermediate  $\rho$  (or  $\rho_0$ ) = 0.5,  $c$  (or  $c_0$ ) = 1.0 calculations were repeated with  $\eta = 0.75, 0.5, 0.25, 0.1$ , and 0.05. Since the

formula (7) is exact when  $\Pi_0(t) = 1.0$ , a scheme was also tested in which the mixing factor  $\eta$  varied with time by computing  $\eta(t) = 1 - \Pi_0(t)$ .

Results by Method 2E in constant parameter calculations were very encouraging. Figure 2.17 illustrates the relative errors in  $v(t)$  for the various mixing factor values. In the three situations illustrated,  $\eta = 0.25$  seems to give the lowest overall errors of all of the values of  $\eta$  tested.

When  $\lambda = 2$  and  $\nu = 2$ , the smallest of the local maxima of absolute relative errors observed is 0.06, occurring in the calculation with  $\eta = 0.25$ . This is less than one-quarter of the maximum relative error occurring when  $\eta = 1$  is used. Actually, for this particular calculation, the relative error starts off with a value of 0.102 when  $t = 0.5$ , but its value drops rapidly toward zero. For low values of  $t$ ,  $v(t)$  also has a smaller value and so larger relative errors are not as significant. When  $\lambda = 3.8$  and  $\nu = 2$ , the  $\eta = 1$  calculation gives a maximum absolute relative error in  $v(t)$  of approximately 0.48, whereas with  $\eta = 0.25$ , this maximum is likely to be around 0.12<sup>1</sup>, again smaller by a factor of near four.

When  $\lambda = 2.0$  and  $\nu = 0.2$ , the maximum relative error in  $v(t)$  decreases monotonically with the values of  $\eta$  used, and has a value of 0.52 when  $\eta = 0.05$ , the smallest value of  $\eta$  for which calculations were done. This compares with a

---

<sup>1</sup>The absolute relative error in  $v(t)$  when  $\eta = 0.25$  occurs at a time greater than  $t = 25.5$ , the largest value of  $t$  for which numbers were obtained in these calculations. However, as seen in Figure 2.17b, the relative error curve is very nearly horizontal by  $t = 25.5$ , and at this point the relative error is -0.1138. It is reasonable to infer that the actual turning point of the curve will not be far below -0.12.

maximum of 0.81 when  $\eta = 1$ . Thus, as  $\nu$  becomes small relative to  $\lambda$ , use of the formula (9) with  $\eta < 1$  does lead to more accurate estimates of  $v(t)$ , though the improvement is not as great as for the other two examples illustrated in Figure 2.17.

Only in the intermediate case ( $\lambda = 2.0$ ,  $\nu = 2.0$ ) is the variable  $\eta$  approach competitive with fixing  $\eta$  at some constant value. In the other two cases studied, it is actually no better than a poor intermediate between the worst case  $\eta = 1$  calculation and the calculation using the absolute relative error minimizing constant value of  $\eta$ . Jumping ahead a bit, we note here that this conclusion regarding the accuracy of the variable  $\eta$  approach has also been born out by comparing relative mean absolute errors given in Table 6 for corresponding oscillatory parameter calculations.

Figure 2.18 gives a more graphic impression of the limitations of formula (9) when  $\nu$  is small relative to  $\lambda$  in the constant parameter case. For the calculations represented in Figures 2.17a and 2.17b,  $\lambda = \nu$  and  $\lambda \approx 2$  respectively, whereas in Figure 2.17c,  $\lambda = 10\nu$ . If, for the first two cases, curves of one approximate quantity are plotted on the same grid as the exact quantity for different values of  $\eta$ , the gap between the  $\eta = 1$  approximation and the exact curve is nearly evenly filled by curves corresponding to a sequence of values of  $\eta$  between 1 and the best value of  $\eta$  for the given queue parameter values. Figure 2.18d illustrates this by plotting  $v(t)$  when  $\lambda = 3.8$  and  $\nu = 2.0$ , and the approximations to  $v(t)$  computed with  $\eta = 1, 0.75, 0.5$ , and  $0.25$ .

Figures 2.18a, 2.18b, and 2.18c give corresponding plots for  $\Pi_0(t)$ ,  $m(t)$ , and  $v(t)$ , respectively, when  $\lambda = 2.0$  and  $\nu = 0.2$ . In each case, it is seen that the approximate curves for the various values of  $\eta$  form a tight band well separated from the exact curve. This means that the comparative lack of improvement observed in using formula (9) when  $\nu$  is quite small relative to  $\lambda$  is due to the intrinsic limitations of formula (9) and not the result of an inadequate search for the best value of  $\eta$ .

For constant parameter calculations with  $\lambda$  not large compared to  $\nu$ , the formula (9) gives an closure approximation of comparable accuracy to those of the previous section when a value of  $\eta$  somewhere between 0.5 and 0.25 is used. When the retrial rate,  $\nu$ , is low relative to the new customer arrival rate  $\lambda$ , we are still without a closure method for computing  $v(t)$  with truly acceptable accuracy, though formula (9) represents best (and significant) improvement over (6) if a value of  $\eta$  near 0.1 is used.

#### Oscillatory Parameter Calculations Using Method 2E

The oscillatory parameter case is, as usual, more complicated to analyze here. Figure 2.19a shows the exact  $v(t)$  plotted on the same grid as the approximate  $v(t)$  curves obtained using  $\eta = 1, 0.75, 0.5$  and  $0.25$ . When  $\lambda_0 = \nu_0 = 2.0$ , this figure indicates that a highly accurate calculation of  $v(t)$  should result if a value of  $\eta$  approximately midway between 0.5 and 0.25 is used. Figure 2.19b shows a similar comparative plot when  $\lambda_0 = 3.6$  and  $\nu_0 = 2.0$ . Again, it appears that the best value of  $\eta$  is somewhere between 0.5 and 0.25.

Figure 2.19c gives a similar comparative plot for  $v(t)$  in the oscillatory parameter case when  $\lambda_0 = 2.0$  and  $\nu_0 = 0.2$ . Here, the  $\eta = 0.75$  curve has been left out and  $\eta = 0.1$  and  $0.05$  curves included. Unlike the corresponding constant parameter calculations when the new customer arrival rate is much larger than the retrial rate, here the use of formula (9) with a value of  $\eta$  in the neighborhood of  $0.1$  results in remarkable improvement in accuracy in the computation of  $v(t)$ . The improvement on using (9) is not as dramatic for  $\Pi_0(t)$  and  $m(t)$  in this oscillatory parameter case, though still significant.

Without belabouring the discussion, we note that a closure method based on (9) appears to give acceptably accurate approximations for  $v(t)$  in this oscillatory parameter case using similar values of  $\eta$  to those found best for corresponding constant parameter calculations. This tends to allay suspicion that formula (9) is simply a numerical trick, and to strengthen the presumption that it actually approximates the relationship between  $m_0(t)$  and  $\Pi_0(t)$  and  $m(t)$  in some generality.

#### Computation of Period-Averages of $m(t)$ and $v(t)$

In Table 6 are found period-averages and mean absolute relative errors of  $\Pi_0(t)$ ,  $m(t)$  and  $v(t)$  for various values of  $\eta$  in the three cases of oscillatory parameter calculations done. These entries can be compared with the entries for  $\eta = 1$  in each case to get a numerical measure of the improvements in accuracy formula (9) represents over formula (6).

The entries for  $v(t)$  in Table 6 are of greatest interest, and the results are quite remarkable. When the parameters have intermediate values,  $\lambda_0 = \nu_0 = 2.0$  with  $\mu(t) = 4.0$  as usual, the smallest of the relative errors observed in the mean value of  $v(t)$  and in the mean absolute errors occurred for  $\eta = 0.35$ .<sup>2</sup> When  $\eta = 0.35$  was used, the relative error in the computed mean value of  $v(t)$  over the time intervals  $[0,10]$  and  $[10,20]$  were just under 5% and 6% respectively, compared to just over 33% and 18% respectively when  $\eta = 1$  was used. This compares well with the values both near 16% found for the corresponding calculation using Rothkopf and Oren's method on the M/M/1 queue (see the  $\rho_0 = 0.5$  entries in Table 1). Thus the accuracy of the computed mean values of  $v(t)$  over these two periods of oscillation improves by factors of more than six and more than three respectively.

Similarly, the relative mean absolute error in  $v(t)$  over these two time intervals drops from 40% and 28% to 8% and 8%, respectively. These compare with 15% and 16% as the relative mean absolute errors in  $v(t)$  observed using Rothkopf and Oren's method on the simple M/M/1 queue in the corresponding  $\rho_0 = 0.5$  case (see Table 1). These are the first instances observed in which the relative accuracy of a closure technique for the retrial queue is better than that of the Rothkopf and Oren method for the corresponding M/M/1 queue, and the increase in accuracy here is significant.

---

<sup>2</sup>The "mean values" of  $\Pi_0(t)$ ,  $m(t)$ , and  $v(t)$  or their absolute errors referred to here are always computed over 10-unit time intervals, the period of oscillation of  $\lambda(t)$  and  $\nu(t)$  used throughout in oscillatory parameter calculations. For discussion purposes throughout this thesis, such mean values are always computed over two "standard" time intervals,  $[0,10]$  and  $[10,20]$ .

The corresponding improvements in accuracy shown when  $\lambda_0 = 3.6$  and  $v_0 = 2.0$  are even more dramatic. When  $\eta = 1$ , the relative errors in the computed mean values of  $v(t)$  over the first two complete periods of oscillation are 78% and 55% respectively, whereas when  $\eta = 0.35$ , these relative errors drop to 8% and 0.5% respectively, an improvement by factors in the neighborhood of 10 or better. The relative mean absolute errors in  $v(t)$  in this case drop from 78% and 55% when  $\eta = 1$  (the fact that both types of relative errors discussed here are equal indicates that the approximate  $v(t)$  is always greater than the exact  $v(t)$ ) to 8% and 2% when  $\eta = 0.35$ . These numbers compare with 20% and 25% respectively for the Rothkopf and Oren method applied to the simple M/M/1 queue. Because the relative error in the computed period-mean of  $v(t)$  is positive when  $\eta = 1$  and negative when  $\eta = 0.1$ , and is apparently continuous,<sup>3</sup> one could in principle find values of  $\eta$  which would make this period-mean of  $v(t)$  exact in any pre-specified period.

Thus, it is noteworthy that errors in the mean value of  $v(t)$  in the two disjoint periods for which calculations were done become small for the same or nearly the same values of  $\eta$ . But more significant is the fact that the mean absolute error in  $v(t)$  decreases so dramatically here, since in computing this quantity, there is not the potential for large positive error components in one

---

<sup>3</sup>The claim of "apparent continuity" is based on the numbers in Table 6 for various values of  $\eta$  between 1 and 0.1. This looks like the sort of assertion that could be proven rigorously without a great difficulty by someone more familiar with the theory of ordinary differential equations.  $\eta$  is effectively a simple constant numerical parameter in equations (1), and the claim would be that the solution of the system (1) under these circumstances is continuous in  $\eta$ .

region to cancel large negative error components in another region as can occur in computing errors in the period-mean itself.

The numbers in Table 6 for the calculation of  $v(t)$  when  $\lambda_0 = 2.0$  and  $\nu_0 = 0.2$  are not as good as for the two cases just described, though use of equation (9) with an appropriately chosen value of  $\eta$  still does give significant improvements in the accuracy of computed approximate mean values compared to calculations when  $\eta = 1$ . When  $\eta = 1$ , the relative error in computed mean values of  $v(t)$  are 63% and 40% respectively for the two standard periods, when  $\eta = 0.05$ , these are -4% and -18%, respectively. One would expect that at some value of  $\eta$  between 0.1 and 0.05, both relative errors would be found roughly between 10% and -10%, which is an improvement by factors of 4 - 6 compared to the  $\eta = 1$  case. The same is true for the relative mean absolute errors in this case.

In all three of the cases covered in Table 6 for  $v(t)$ , the variable  $\eta$  method gives, if not the poorest relative accuracy, then close to the poorest as far as the computation of these period-mean values is concerned.

The dependence of the relative error in the computed period-mean values of  $\Pi_0(t)$  and  $m(t)$ , and their relative mean absolute errors on  $\eta$  is not as great as for  $v(t)$ . Because these relative errors are generally much smaller than for  $v(t)$ , we simply note here that inspection of Table 6 indicates that while minimal relative errors for  $\Pi_0(t)$  and  $m(t)$  seem to occur when  $\eta$  is closer to 1 rather than 0.35 when retrial rates are not much smaller than new customer arrival rates, their values are generally well under 10% over most intermediate values of

$\eta$  and compare well with the numbers observed for  $p_0(t)$  and  $m(t)$  of the M/M/1 queue with Rothkopf and Oren's method.

#### Estimation of $m_0(t)$ Using Method 2E

It appears that in these cases where the retrial rate is not much smaller than the new customer arrival rate, equation (9) is not only compensating for deficiencies in equation (6), but also to some extent for deficiencies in equation (4), since with optimal values of  $\eta$ , we are observing significantly smaller relative errors here than for the corresponding results for  $v(t)$  for the M/M/1 queue with Rothkopf and Oren's original method.

Evidence for this claim is found in the fourth part of Table 6 dealing with  $m_0(t)$  and in Figure 2.20. If it was simply a matter of equation (9) giving more accurate estimates of  $m_0(t)$ , then we would expect that the value of  $\eta$  leading to the most accurate  $v(t)$  should also give much improved estimates of  $m_0(t)$ . Not only is this not so according to Table 6, but the relative errors found for  $m_0(t)$  in virtually all cases are extremely large, indicating that equation (9) generally would be a completely unacceptable way to compute  $m_0(t)$  even though it is an integral part of a method which produces quite acceptable estimates for  $\Pi_0(t)$ ,  $m(t)$ , and  $v(t)$ .

The numbers in Table 6 for  $m_0(t)$ , and even more so, the plots in Figure 2.20, are disturbing because they cast at least some doubt on the entire rationale of closure approaches as described by Rothkopf and Oren in their paper, and in this

thesis. Of the several closure methods examined so far, this is the first which required two distinct closure formulas, and so it is the first instance where accuracy in some of the quantities computed may be enhanced by taking advantage of the interplay between closure formulas used.

What is also remarkable after seeing the plots in Figure 2.20 is that the  $\eta = 1$  calculations give as accurate values for  $\Pi_0(t)$ ,  $m(t)$ , and  $v(t)$  as they do, since the estimates of  $m_0(t)$  when  $\eta = 1$  are almost as bad as they can be without being totally absurd numerically. However, it is not valid to conclude that closure formulas are nothing but fancy or complicated "fudge-formulas", since here,  $\Pi_0(t)$  is also estimated using such an imposed formula, and the resulting estimates reflect exact values very well.

A more viable approach for computing  $m_0(t)$  is described in the next section. Also, the method described in the preceding section produces an estimate of  $m_1(t)$  as well as  $m(t)$ , and the difference of these will give an estimate of  $m_0(t)$ . No attempt has been made to determine the accuracy of such estimates of  $m_0(t)$  however. If one was going to do that, it would probably be worthwhile to first explore the improvements possible by varying the mixing factor in equation (2.3.6). Since such an approach to computing  $m_0(t)$  would involve taking the difference of two approximate values, it would be desirable to maximize the accuracy of both approximate values first.

**TABLE 5a**      **Period Averages and Mean Absolute Errors in  $\Pi_0(t)$ ,  $m(t)$  and  $v(t)$  for Various New Customer Arrival Rates Using Method 2E.**

The entries in this table correspond to oscillatory parameter calculations using Method 2E with  $v_0 = 2.0$  and  $\mu(t) = 4.0$ , a constant, in all cases. The first column below gives the value of  $\rho_0 = \lambda_0/\mu$ , the second gives the time interval over which the mean values were computed, the third column gives the exact mean, the fourth column gives the closure method mean value, the fifth column gives the error in these period-mean values, and the sixth column gives the relative error in the closure method mean values. The last two columns give, respectively, the mean absolute error over the indicated time intervals, and the corresponding relative mean absolute error.

$\Pi_0(t)$  :

0.1	[0,10]	.8930	.8935	.004987	.0005585	.002452	.002746
	[10,20]	.8900	.8900	.000001	.0000011	.001842	.002069
0.3	[0,10]	.6840	.6857	.001743	.002548	.02159	.03157
	[10,20]	.6701	.6700	-.0001209	-.0001804	.01834	.02736
0.5	[0,10]	.4969	.4915	-.005414	-.001089	.05286	.1064
	[10,20]	.4560	.4523	-.003608	-.007912	.04824	.1058
0.7	[0,10]	.3560	.3453	-.01071	-.03008	.08535	.2398
	[10,20]	.2799	.2631	-.01680	-.06000	.07880	.2815
0.9	[0,10]	.2638	.2530	-.01084	-.04108	.09822	.3723
	[10,20]	.1717	.1452	-.02649	-.1543	.07521	.4381

$m(t)$  :

0.1	[0,10]	.1509	.1545	.003536	.02342	.01131	.07491
	[10,20]	.1571	.1604	.003371	.02146	.01121	.07135
0.3	[0,10]	.8979	.9237	.02580	.02874	.1268	.1412
	[10,20]	.9789	.9934	.01442	.01473	.1217	.1243
0.5	[0,10]	2.7446	2.8200	.07538	.02746	.3838	.1398
	[10,20]	3.4260	3.2611	-.1650	-.04815	.3078	.08983
0.7	[0,10]	5.9320	6.1062	.1741	.02935	.6280	.1059
	[10,20]	9.0347	8.4946	-.5401	-.05978	.5806	.06427
0.9	[0,10]	10.1175	10.4763	.3588	.03546	.7707	.07616
	[10,20]	18.0434	17.4799	-.5635	-.03123	.6127	.03396

Table 5a, continued....

v(t) :

0.1	[0,10]	.1787	.1834	.004647	.02600	.01399	.07826
	[10,20]	.1862	.1908	.004538	.02437	.01412	.07583
0.3	[0,10]	1.5713	1.7450	.1737	.1106	.3396	.2161
	[10,20]	1.7420	1.9002	.1582	.09083	.3398	.1951
0.5	[0,10]	6.4256	8.5757	2.3980	.3732	2.5981	.4043
	[10,20]	8.8236	10.4124	1.5888	.1801	2.4490	.2776
0.7	[0,10]	15.3710	24.4532	9.0822	.5909	9.2383	.6010
	[10,20]	28.5567	38.4299	9.8732	.3457	10.6962	.3746
0.9	[0,10]	25.5950	45.5821	19.9872	.7809	19.9872	.7809
	[10,20]	57.8233	89.3676	31.5443	.5455	31.5443	.5455

**TABLE 5b**      **Period Averages and Mean Absolute Errors in  $\Pi_0(t)$ ,  $m(t)$  and  $v(t)$  for Various Retrial Rates Using Method 2E.**

The entries in this table correspond to oscillatory parameter calculations using Method 2E with  $\lambda_0 = 2.0$  and  $\mu(t) = 4.0$ , a constant, in all cases. The first column below gives the value of  $v_0$  and the second column gives the value of  $c_0 = \lambda_0/v_0$ . The remaining seven columns are arranged as are the last seven columns of table 5a.

$\Pi_0(t)$  :

10.0	0.2	[0,10]	.4734	.4745	.001115	.002354	.01078	.02276
		[10,20]	.4505	.4506	.0000775	.0001720	.01066	.02232
4.0	0.5	[0,10]	.4821	.4815	-.0005578	-.001157	.02742	.05687
		[10,20]	.4519	.4511	-.0007688	-.001701	.02485	.05498
2.0	1.0	[0,10]	.4969	.4915	-.005414	-.01089	.05286	.1064
		[10,20]	.4560	.4523	-.003608	-.007912	.04824	.1058
1.0	2.0	[0,10]	.5230	.5090	-.01402	-.02680	.09094	.1739
		[10,20]	.4676	.4560	-.01159	-.02478	.08200	.1754
0.4	5.0	[0,10]	.5714	.5552	-.01614	-.02825	.1482	.2594
		[10,20]	.5032	.4742	-.02902	-.05768	.1290	.2564
0.2	10.0	[0,10]	.6088	.6146	.005798	.009523	.1864	.3062
		[10,20]	.5434	.5096	-.03386	-.06231	.1608	.2959

$m(t)$  :

10.0	0.2	[0,10]	1.9893	1.9789	-.01040	-.005227	.06651	.03343
		[10,20]	2.1655	2.4170	-.01847	-.008531	.07212	.03331
4.0	0.5	[0,10]	2.3078	2.3280	.02017	.008738	.1663	.07205
		[10,20]	2.6483	2.5961	-.05223	-.01972	.1484	.05605
2.0	1.0	[0,10]	2.7746	2.8200	.07538	.02746	.3838	.1398
		[10,20]	3.4260	3.2611	-.1650	-.04815	.3078	.08983
1.0	2.0	[0,10]	3.3794	3.6067	.2273	.06726	.7285	.2156
		[10,20]	4.7900	4.4184	-.3717	-.07759	.5725	.1195
0.4	5.0	[0,10]	4.3977	5.2172	.8195	.1863	1.2978	.2951
		[10,20]	7.5169	7.2490	-.2679	-.03563	.9721	.1293
0.2	10.0	[0,10]	5.1257	6.8207	1.6949	.3307	1.7413	.3397
		[10,20]	9.8624	10.7406	.8782	.08905	1.5219	.1543

Table 5b, continued....

v(t) :

10.0	0.2	[0,10]	5.3642	6.4752	1.1110	.2071	1.1223	.2092
		[10,20]	6.0539	7.2249	1.1710	.1934	1.1834	.1955
4.0	0.5	[0,10]	5.8539	7.4099	1.5565	.2659	1.7028	.2909
		[10,20]	7.1297	8.5744	1.4447	.2026	1.6789	.2355
2.0	1.0	[0,10]	6.4256	8.5757	2.3980	.3732	2.5981	.4043
		[10,20]	8.8236	10.4124	1.5888	.1801	2.4490	.2776
1.0	2.0	[0,10]	7.0719	10.1098	3.0379	.4296	3.8331	.5420
		[10,20]	11.3991	13.2489	1.8499	.1623	3.7199	.3263
0.4	5.0	[0,10]	7.8444	12.2899	4.4455	.5667	5.3012	.6758
		[10,20]	15.2618	19.0427	3.7809	.2477	6.1736	.4045
0.2	10.0	[0,10]	8.3430	13.5835	5.2405	.6281	5.7930	.6944
		[10,20]	17.6819	24.7199	7.0380	.3980	8.6585	.4897

**TABLE 6** Period Averages and Relative Mean Absolute Errors in  $\Pi_0(t)$ ,  $m(t)$  and  $v(t)$  Using Method 2E With Various Amounts of Mixing,  $\eta$ , in the Closure Formula for  $m_0(t)$ .

The three triads of numbers in the rightmost nine columns below are the approximate mean value of the quantity indicated (as calculated from closure method values), the relative error in this approximate mean value, and the relative mean absolute error over the intervals indicated, respectively, for three different oscillatory parameter retrieval queues with the parameter values listed at the bottom of each section of this table. The number,  $\eta$ , tabulated in the first column, is the mixing factor,  $\eta$ , in equation (2.4.9).

(a) $\Pi_0(t)$ :										
$\eta$										
1.0	[0,10]	.4915	-.01089	.1064	.2529	-.04108	.3723	.6146	.009523	.3062
	[10,20]	.4523	-.007912	.1058	.1452	-.1543	.4381	.5096	-.06230	.2959
0.75	[0,10]	.4941	-.005766	.1087	.2468	-.06455	.3901	.6142	.008771	.3074
	[10,20]	.4530	-.006486	.1082	.1425	-.1698	.4704	.5090	-.06339	.2966
0.5	[0,10]	.4966	-.0006466	.1159	.2528	-.1026	.4119	.6128	.006545	.3107
	[10,20]	.4536	-.005041	.1224	.1318	-.1956	.5015	.5303	-.06581	.3032
0.4	[0,10]	.4972	.0004573	.1198	.2309	-.1246	.4199			
	[10,20]	.4539	-.004393	.1315	.1351	-.2126	.5122			
0.35	[0,10]	.4973	.0006131	.1219	.2274	-.1378	.4243			
	[10,20]	.4541	-.004035	.1366	.1332	-.2239	.5172			
0.25	[0,10]	.4968	-.0002608	.1299	.2349	-.1705	.4351	.6074	-.002449	.3182
	[10,20]	.4545	-.003270	.1474	.1210	-.2552	.5248	.5301	-.07169	.3219
0.1	[0,10]	.4934	-.007194	.1571	.1981	-.2489	.4631	.5954	-.02212	.3310
	[10,20]	.4547	-.002783	.1694	.1113	-.3518	.5212	.4998	-.08037	.3461
0.05	[0,10]	.4901	-.01372	.1732	.1862	-.2942	.4739	.5860	-.03747	.3409
	[10,20]	.4542	-.003845	.1823	.09924	-.4219	.5283	.4947	-.08961	.3563
vble	[0,10]	.5045	.01510	.1185	.2576	-.02330	.3951	.6176	.01445	.3127
	[10,20]	.4565	.001275	.1515	.1564	-.08899	.5064	.5166	-.04939	.3181
Exact	[0,10]	.4969			.2638			.6088		
	[10,20]	.4560			.1717			.5434		
	$\rho_0$	1.0			0.9			1.0		
	$c_0$	1.0			1.0			10.0		
	$\lambda_0$	2.0			3.6			2.0		
	$v_0$	2.0			2.0			0.2		
	$\mu$	4.0			4.0			4.0		

Table 6, continued....

(b)  $m(t)$ :

$\eta$	1.0	[0,10]	2.8200	.02746	.1398	10.4763	.03546	.07617	6.8207	.3307	.3397
		[10,20]	3.2611	-.04815	.08983	17.4799	-.03123	.03396	10.7406	.08905	.1543
0.75		[0,10]	2.7710	.009625	.1284	10.3148	.01950	.07767	6.7776	.3223	.3343
		[10,20]	3.2515	-.05095	.09186	17.2285	-.04516	.04516	10.6638	.08126	.1523
0.5		[0,10]	2.6836	-.02224	.1280	10.0988	-.001849	.08200	6.6984	.3068	.3272
		[10,20]	3.2367	-.05526	.1065	16.8022	-.06879	.06879	10.5263	.06732	.1499
0.4		[0,10]	2.6294	-.04199	.1336	9.9897	-.01263	.08533			
		[10,20]	3.2220	-.05955	.1164	16.5391	-.08337	.08337			
0.35		[0,10]	2.5959	-.05420	.1400	9.9286	-.01867	.08741			
		[10,20]	3.2092	-.06327	.1221	16.3753	-.09245	.09245			
0.25		[0,10]	2.5106	-.08527	.1582	9.7895	-.03242	.09287	6.5200	.2720	.3198
		[10,20]	3.1613	-.07728	.1357	15.9511	-.1160	.1160	10.2291	.03718	.1485
0.1		[0,10]	2.3004	-.1618	.2166	9.5172	-.05933	.1068	6.2550	.2203	.3257
		[10,20]	2.9319	-.1442	.1745	14.8405	-.1775	.1775	9.7145	-.01500	.1466
0.05		[0,10]	2.1819	-.2050	.2545	9.3953	-.07138	.1144	6.0817	.1865	.3335
		[10,20]	2.7225	-.2053	.2120	14.1551	-.2155	.2155	9.2632	-.06076	.1508
vble		[0,10]	2.7856	.01492	.1123	10.3706	.02501	.07113	6.6968	.3065	.3231
		[10,20]	3.6729	.07206	.1377	17.9813	-.003442	.03371	10.8931	.1045	.1633
Exact		[0,10]	2.7446			10.1175			5.1257		
		[10,20]	3.4260			18.0434			9.8624		
	$\rho_0$	1.0				0.9			1.0		
	$c_0$	1.0				1.0			10.0		
	$\lambda_0$	2.0				3.6			2.0		
	$v_0$	2.0				2.0			0.2		
	$\mu$	4.0				4.0			4.0		

Table 6, continued....

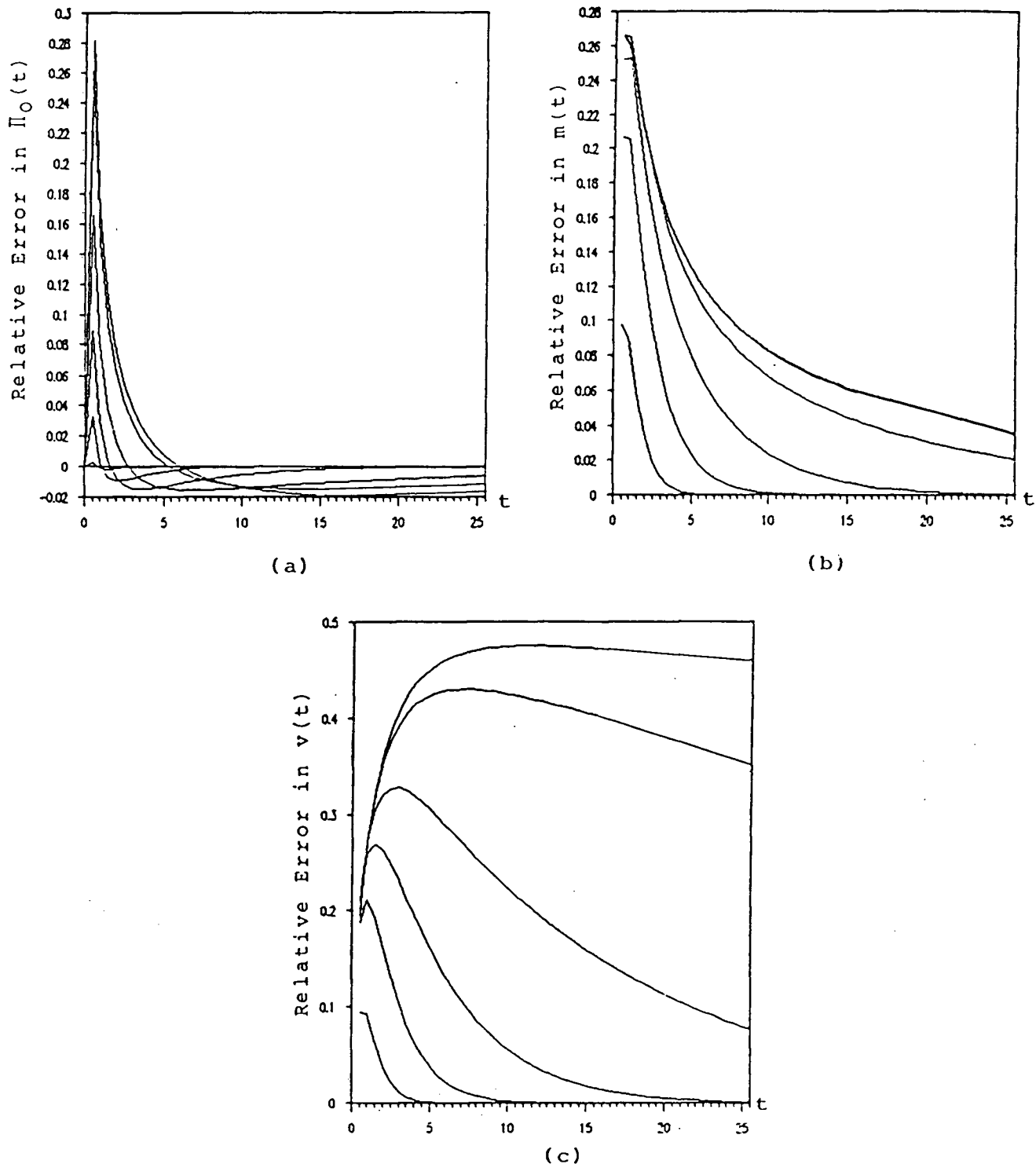
(c)  $v(t)$ :

$\eta$										
1.0	[0,10]	8.5757	.3346	.4043	45.5821	.7809	.7809	13.5835	.6281	.6944
	[10,20]	10.4124	.1801	.2776	89.3676	.5455	.5455	24.7199	.3980	.4897
0.75	[0,10]	8.1823	.2734	.3032	40.2226	.5715	.5715	13.3437	.5994	.6511
	[10,20]	10.1664	.1522	.2016	79.9766	.3831	.3831	24.1721	.3671	.4423
0.5	[0,10]	7.4808	.1642	.1645	33.1423	.2949	.2949	12.8761	.5433	.5657
	[10,20]	9.7751	.1078	.1098	67.8332	.1731	.1731	23.1664	.3102	.3538
0.4	[0,10]	7.0277	.09371	.09371	29.6465	.1583	.1583			
	[10,20]	9.5223	.07918	.07918	61.6341	.06590	.06590			
0.35	[0,10]	6.7415	.04918	.07762	27.7192	.08299	.08299			
	[10,20]	9.3387	.05838	.08096	58.0919	.004645	.02388			
0.25	[0,10]	6.0015	-.06600	.1333	23.4414	-.08414	.08515	11.6088	.3914	.3914
	[10,20]	8.7741	-.005612	.1186	49.7807	-.1390	.1390	20.9121	.1827	.1899
0.1	[0,10]	4.2022	-.3460	.3489	15.6918	-.3869	.3874	9.4407	.1316	.1365
	[10,20]	6.6994	-.2407	.2566	32.2777	-.4418	.4418	17.3672	-.01779	.03091
0.05	[0,10]	3.2710	-.4909	.4929	12.6375	-.5062	.5066	8.0350	-.03691	.1083
	[10,20]	5.1419	-.4173	.4173	23.9212	-.5863	.5863	14.5742	-.1758	.1758
vble	[0,10]	9.1767	.4282	.4282	45.9507	.7953	.7953	13.7861	.6524	.6524
	[10,20]	14.1814	.6072	.6072	98.9332	.7110	.7110	27.1265	.5341	.5341
Exact	[0,10]	6.4256			25.5950			8.3430		
	[10,20]	8.8236			57.8233			17.6819		
	$\rho_0$	1.0			0.9			1.0		
	$c_0$	1.0			1.0			10.0		
	$\lambda_0$	2.0			3.6			2.0		
	$v_0$	2.0			2.0			0.2		
	$\mu$	4.0			4.0			4.0		

Table 6, continued....

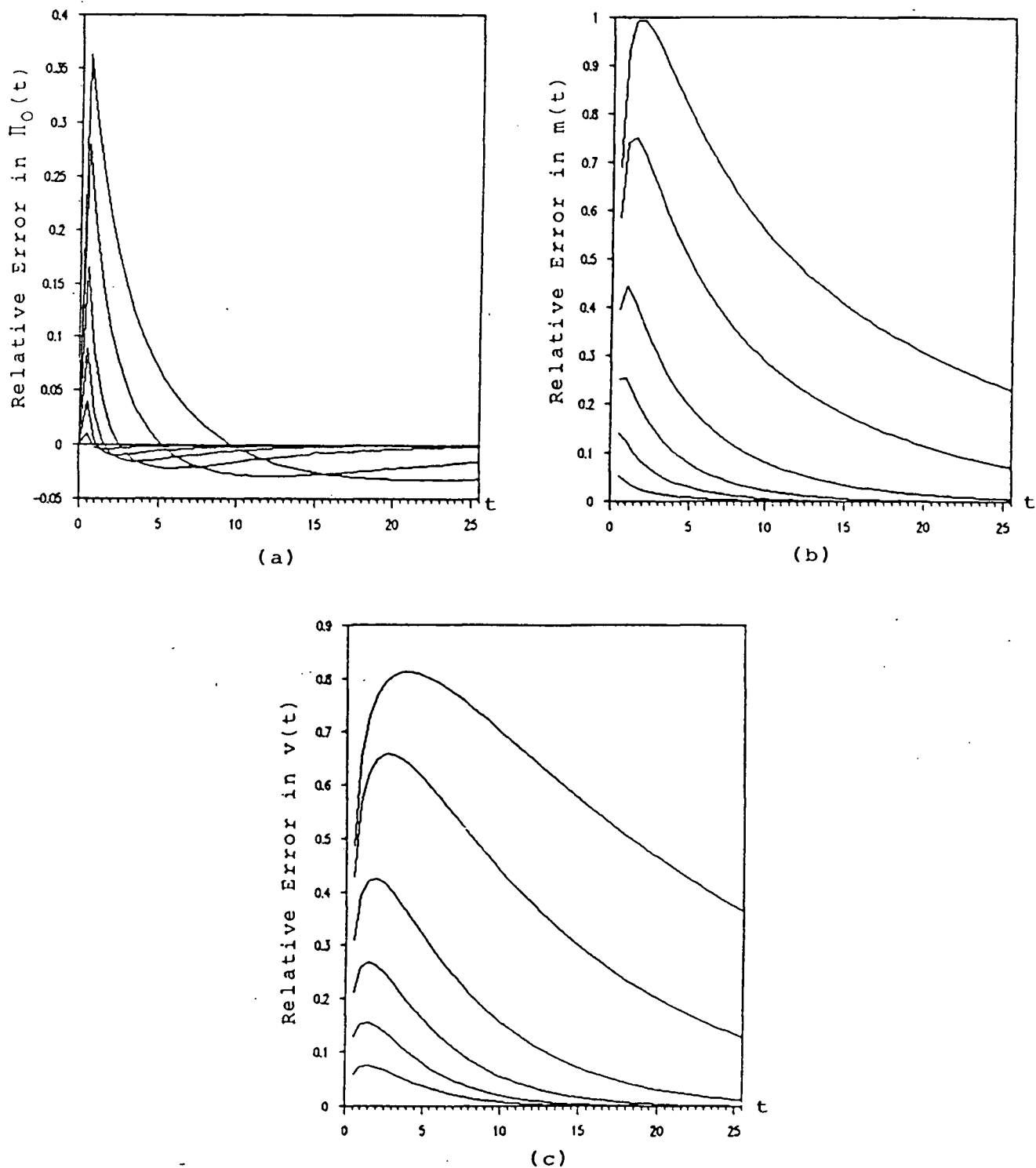
(d)  $\underline{m}_0(t):$ 

$\eta$										
1.0	[0, 10]	.5085	-.1948	.5402	.7470	-.5761	.6444	.3854	-.8619	.8619
	[10, 20]	.5477	-.2874	.4738	.8548	-.6442	.6675	.4904	-.9002	.9002
0.75	[0, 10]	.4995	-.2090	.4247	.7951	-.5488	.5860	1.1089	-.6026	.6241
	[10, 20]	.5477	-.2873	.3644	.9213	-.6165	.6165	1.5051	-.4606	.6938
0.5	[0, 10]	.4738	-.2498	.3636	.8309	-.5712	.6183	1.8010	-.3546	.5056
	[10, 20]	.5377	-.3003	.3261	.9113	-.6208	.6208	2.5505	-.4976	.4976
0.4	[0, 10]	.4905	-.2233	.3417	.4480	-.7458	.7867	2.1212	-.2398	.4745
	[10, 20]	.5488	-.2860	.3069	.5420	-.7744	.7744	2.9248	-.4050	.4289
0.35	[0, 10]	.4649	-.2638	.3727	.3355	-.8096	.8519	2.2252	-.2026	.4676
	[10, 20]	.5351	-.3037	.3412	.3738	-.8444	.8444	3.1747	-.3542	.4201
0.25	[0, 10]	.3893	-.3835	.4898	.6690	-.6858	.7409	2.3869	-.1446	.4721
	[10, 20]	.4880	-.3650	.4408	.7144	-.6985	.7152	3.4673	-.3302	.4243
0.1	[0, 10]	.2493	-.6053	.7050	.2710	-.8462	.9016	2.5424	-.08890	.5079
	[10, 20]	.3666	-.5230	.6119	.3687	-.8466	.8472	3.5388	-.2801	.4514
0.05	[0, 10]	.1588	-.7485	.8311	.1148	-.9349	.9903	2.4843	-.1097	.5433
	[10, 20]	.2608	-.6607	.7123	.1307	-.9456	.9456	3.4292	-.3024	.4753
vble	[0, 10]	.6483	.02662	.3014	1.3016	-.2613	.3581	2.2512	-.1932	.4792
	[10, 20]	.7903	.02836	.2723	1.4561	-.3939	.4106	3.0418	-.3812	.4379
Exact	[0, 10]	.6315			1.7622			2.7905		
	[10, 20]	.7685			2.4026			4.9159		
	$\rho_0$	1.0			0.9			1.0		
	$c_0$	1.0			1.0			10.0		
	$\lambda_0$	2.0			3.6			2.0		
	$v_0$	2.0			2.0			0.2		
	$\mu$	4.0			4.0			4.0		



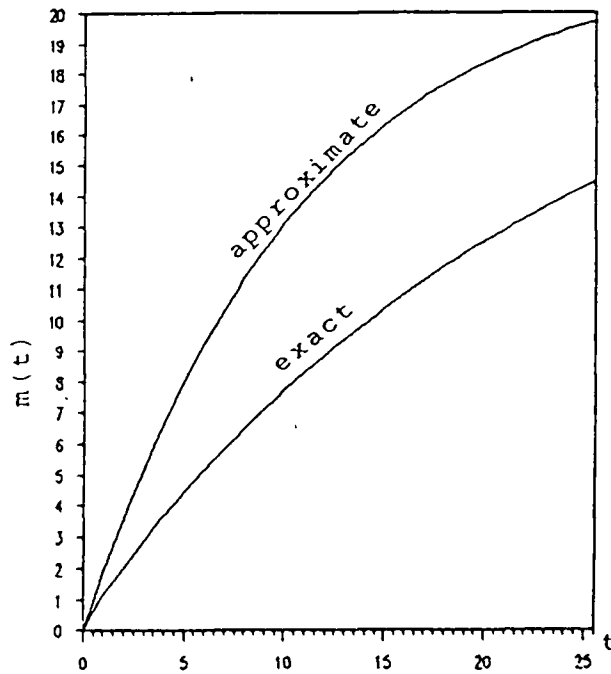
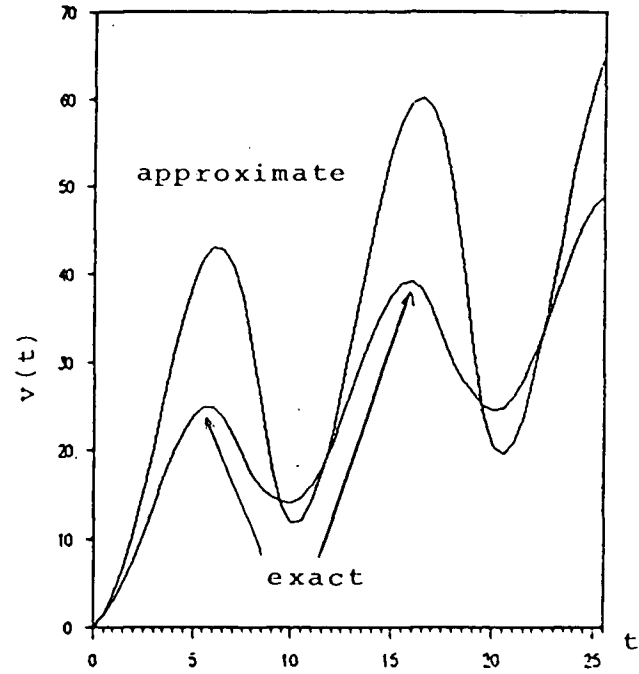
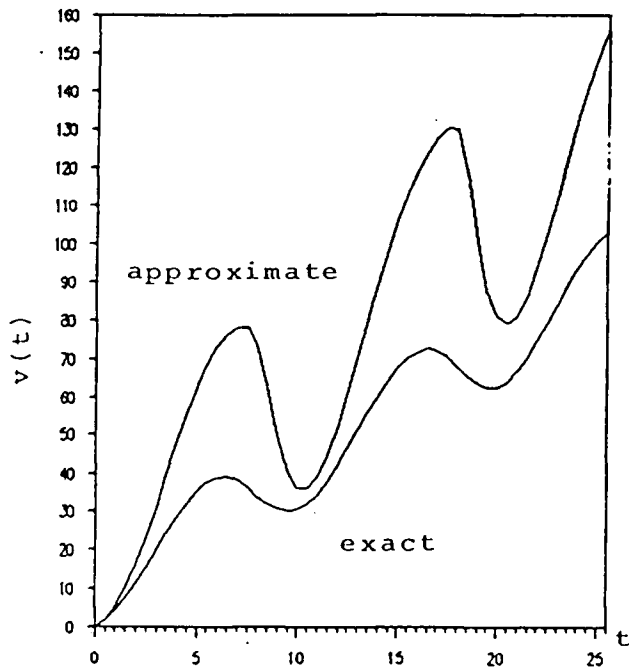
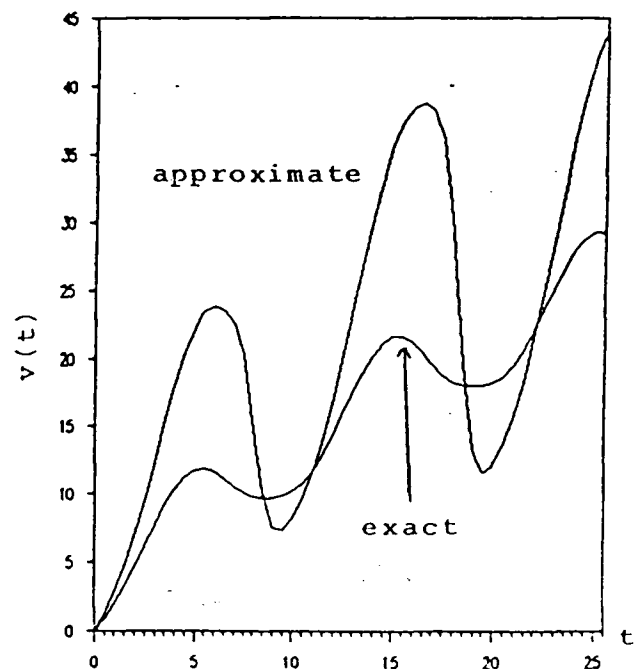
**FIGURE 2.13** Relative Errors in  $\Pi_0(t)$ ,  $m(t)$  and  $v(t)$  Using Closure Method 2D for Various Values of  $\rho$  in the Constant Parameter Retrial Queue.

The graphs were prepared from calculations based on equations (2.4.1) with (2.4.6). The six curves in each Figure correspond from lowest to highest to  $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$ , and  $0.95$ , respectively. In Figures 2.13a and 2.13b, the relative error curves corresponding to  $\rho = 0.9$  and  $\rho = 0.95$  are effectively indistinguishable at the resolution employed. Appendix A gives details of calculations done to obtain exact values required in determining these relative errors.



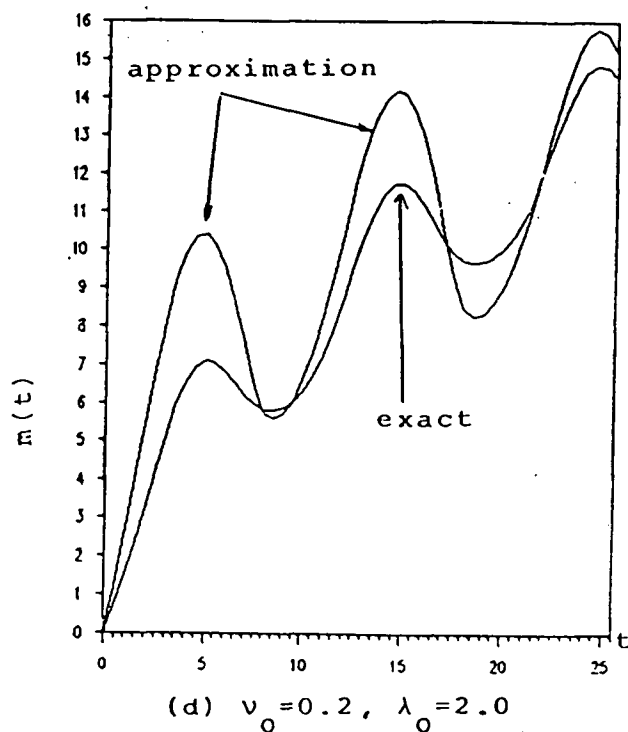
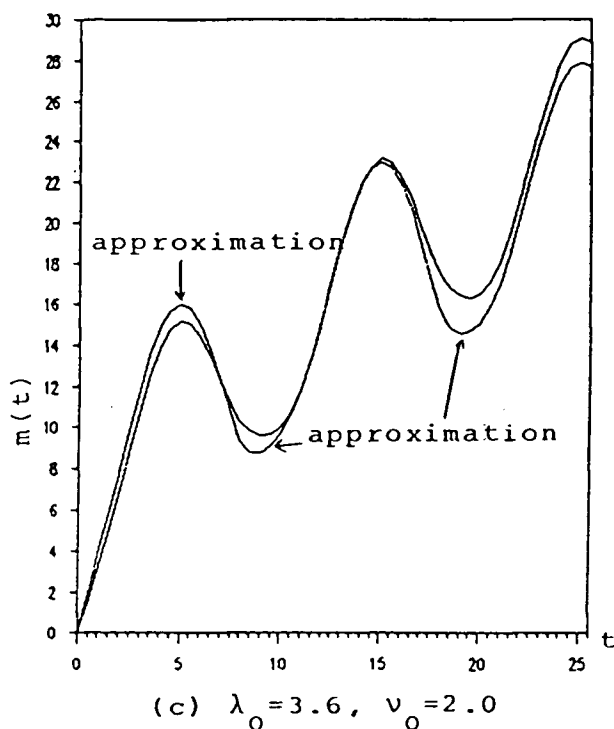
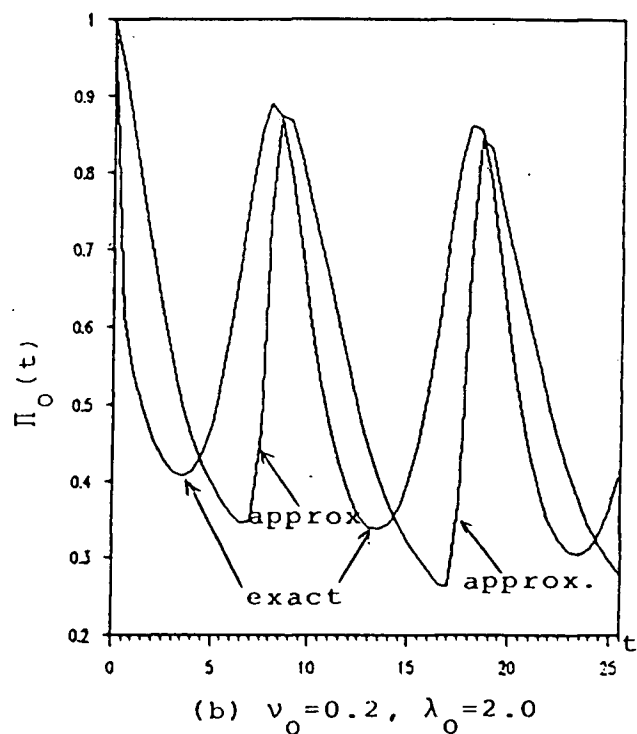
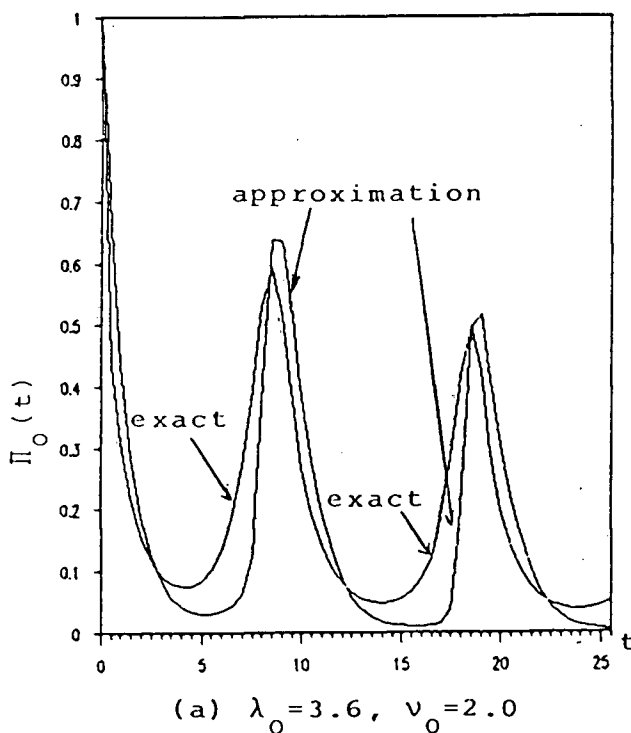
**FIGURE 2.14** Relative Errors in  $\Pi_0(t)$ ,  $m(t)$  and  $v(t)$  Using Closure Method 2D for Various Values of  $c$  in the Constant Parameter Retrial Queue.

The graphs were prepared from calculations based on equations (2.4.1) with (2.4.6). The six curves in each Figure correspond from lowest to highest to  $c = 0.2, 0.5, 1.0, 2.0, 5.0$ , and  $10.0$ , respectively. As always in this thesis, the system is assumed to be empty at  $t = 0$  (i.e.,  $\Pi_0(0) = 1.0$ ,  $m(0) = 0.0$  and  $v(0) = 0.0$ ). Appendix A gives details of calculations done to obtain exact values required in determining these relative errors.

(a)  $\lambda=2.0$ ,  $\mu=4.0$ ,  $\nu=2.0$ (b)  $\rho_0=0.7$ ,  $c_0=1.0$ (c)  $\rho_0=0.9$ ,  $c_0=1.0$ (d)  $\rho_0=0.5$ ,  $c_0=10.0$ 

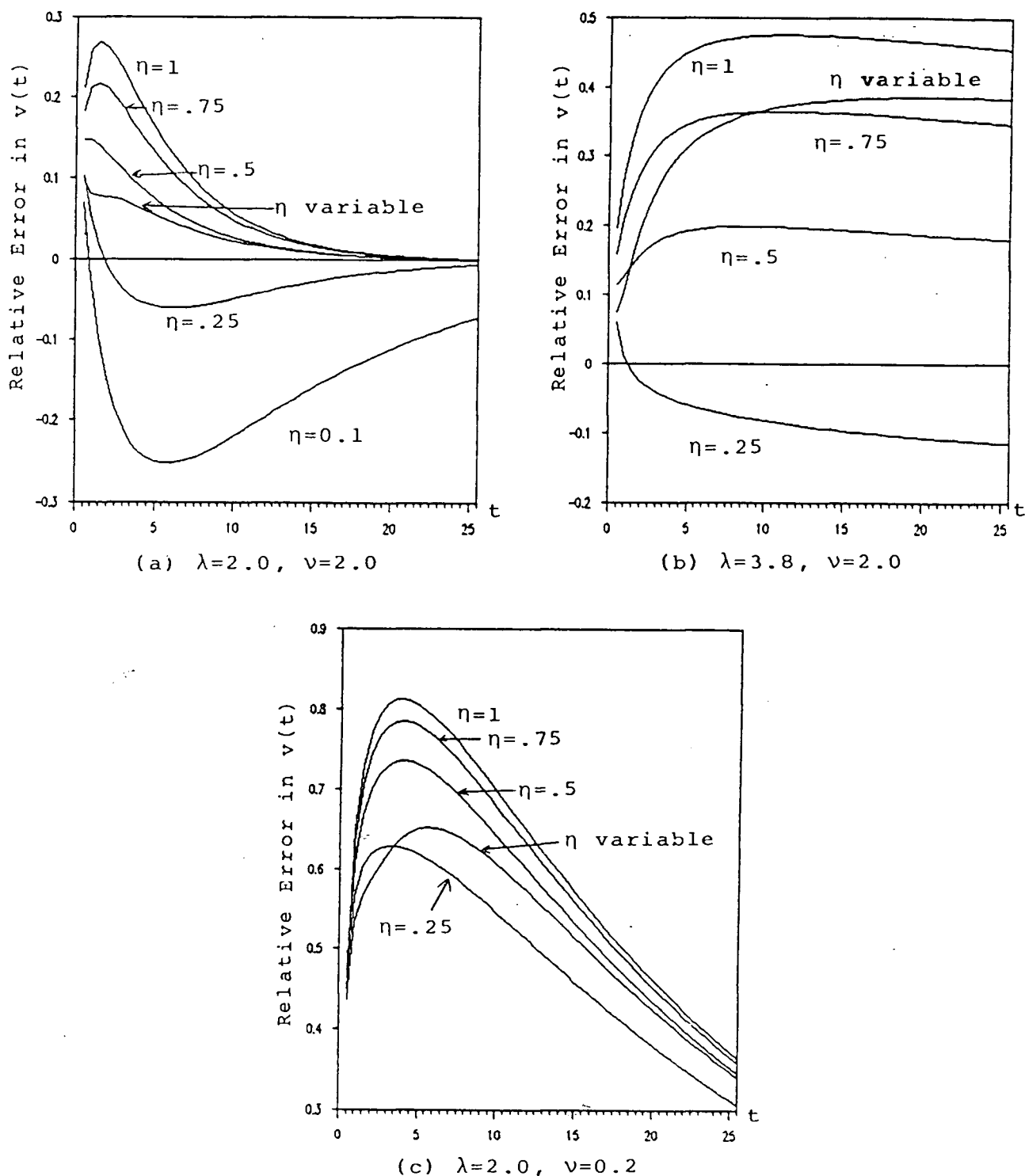
**FIGURE 2.15 Comparison of Exact and Approximate Quantities for the Retrial Queue Using Method 2D.**

The calculations corresponding to these Figures were based on equations (2.4.1) and (2.4.6). Figure 2.15a applies to the constant parameter case, whereas the other three Figures on this page apply to oscillatory parameter cases. The emphasis is on  $v(t)$  here, and the accuracy of the closure approximation is clearly not great in any of the cases displayed. Appendix A gives details of calculations done to obtain these exact quantities.



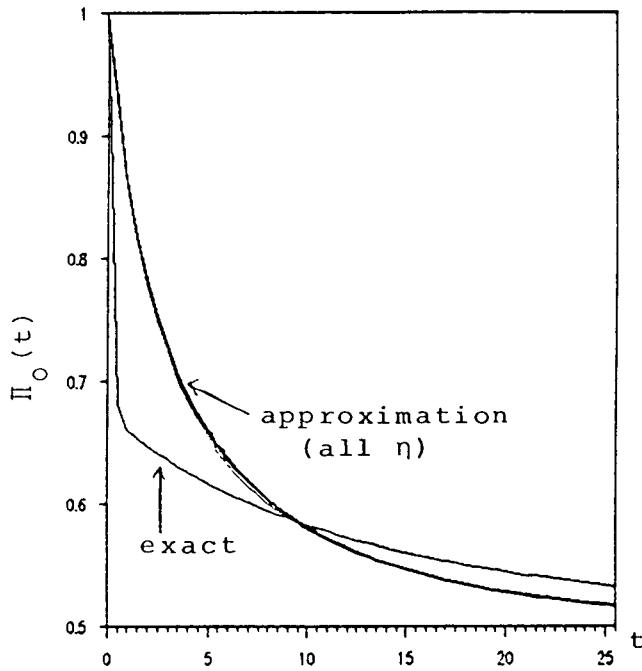
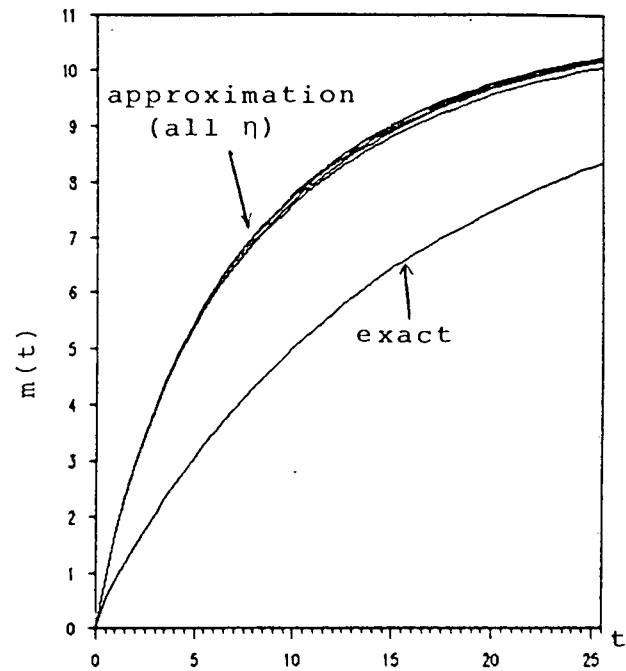
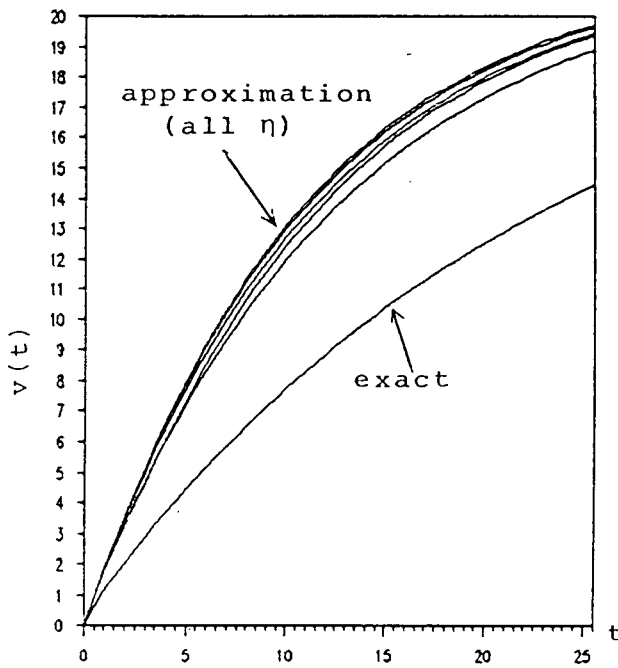
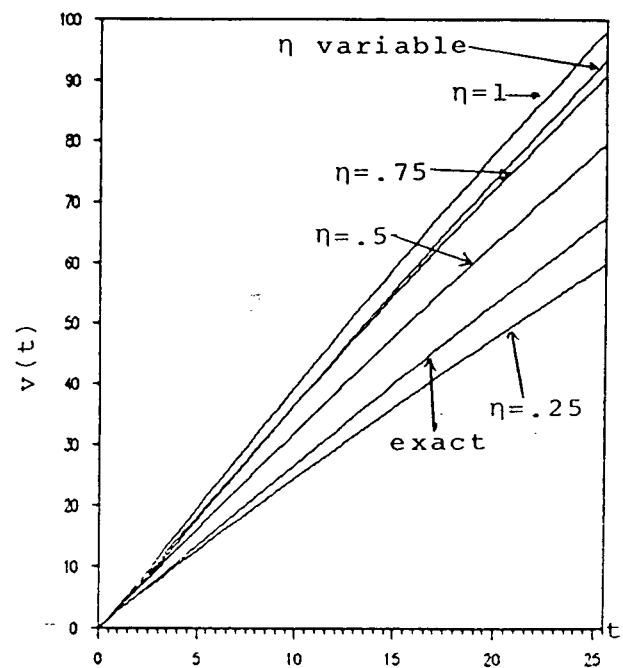
**FIGURE 2.16 Comparison of Exact and Closure Method 2D Approximation Values of  $\Pi_0(t)$  and  $m(t)$  for Oscillatory Parameter Retrial Queues.**

For all of the calculations represented here,  $\mu(t) = 4.0$ . The discrepancies between the exact and approximate  $\Pi_0(t)$  and  $m(t)$  are not as great as those between the exact and approximate  $\nu(t)$  as displayed in Figures 2.15c and 2.15d. However, particularly when  $\nu = 0.2$  and  $\lambda = 2.0$ , agreement between exact and approximation is still less than acceptable. Appendix A gives details of calculations done to obtain these exact quantities.



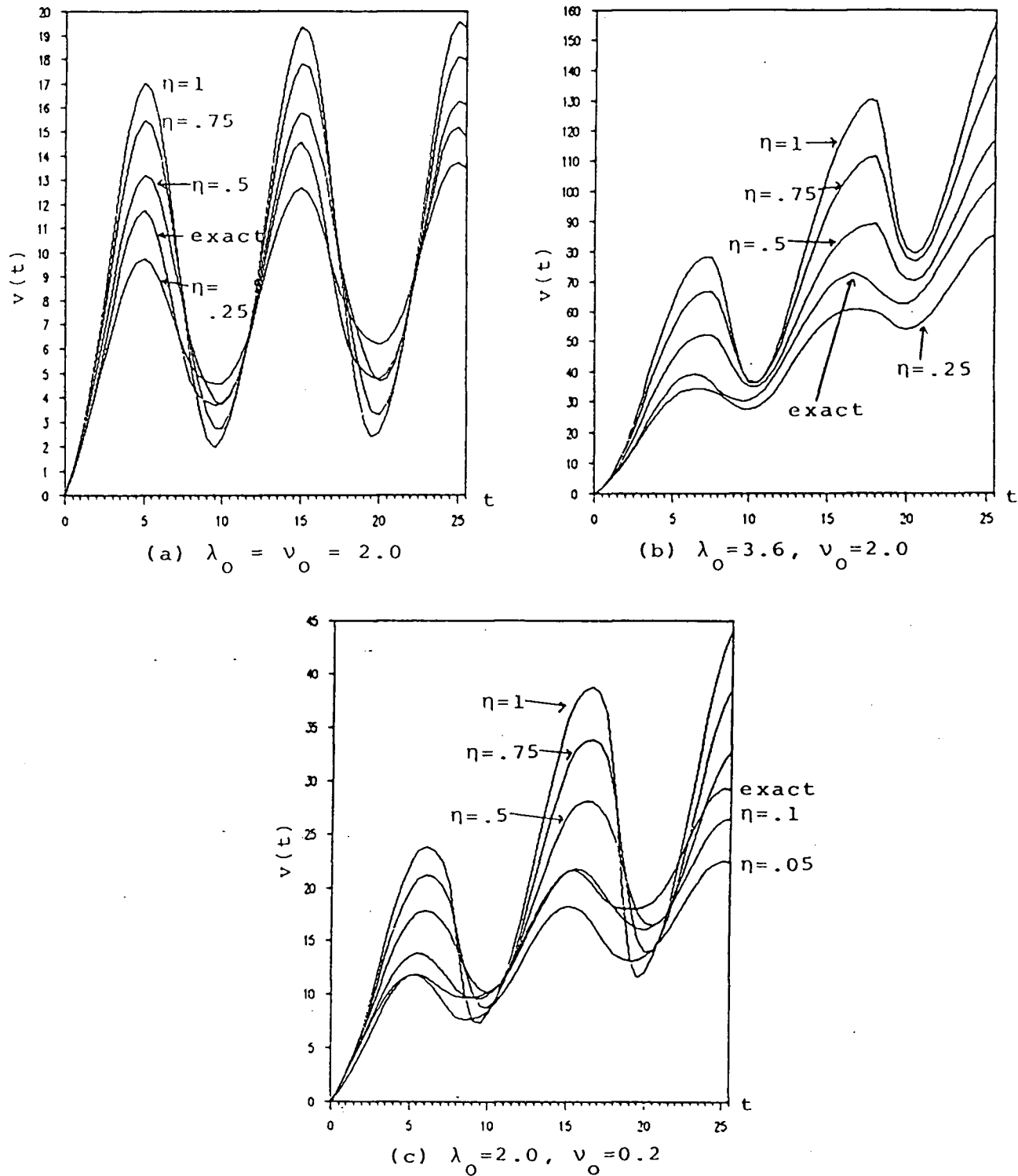
**FIGURE 2.17** Relative Errors in  $v(t)$  Computed Using the Closure Method 2E in the Constant Parameter Case.

The calculations represented here were done using equation (2.4.9) along with (2.4.1). Particularly when  $v$  is not small compared to  $\lambda$ , the relative errors are very sensitive to the value of  $\eta$  used in (2.4.9). In Figure 2.17c, the relative error curves for  $\eta = 0.1$  and  $\eta = 0.05$  peak lower yet than those shown, at values of approximately 0.54 and 0.52, respectively. Note that the vertical scale in Figure 2.17c is clipped. In all calculations represented here,  $\mu = 4.0$ .

(a)  $v=0.2, \lambda=2.0$ (b)  $v=0.2, \lambda=2.0$ (c)  $v=0.2, \lambda=2.0$ (d)  $v=2.0, \lambda=3.8$ 

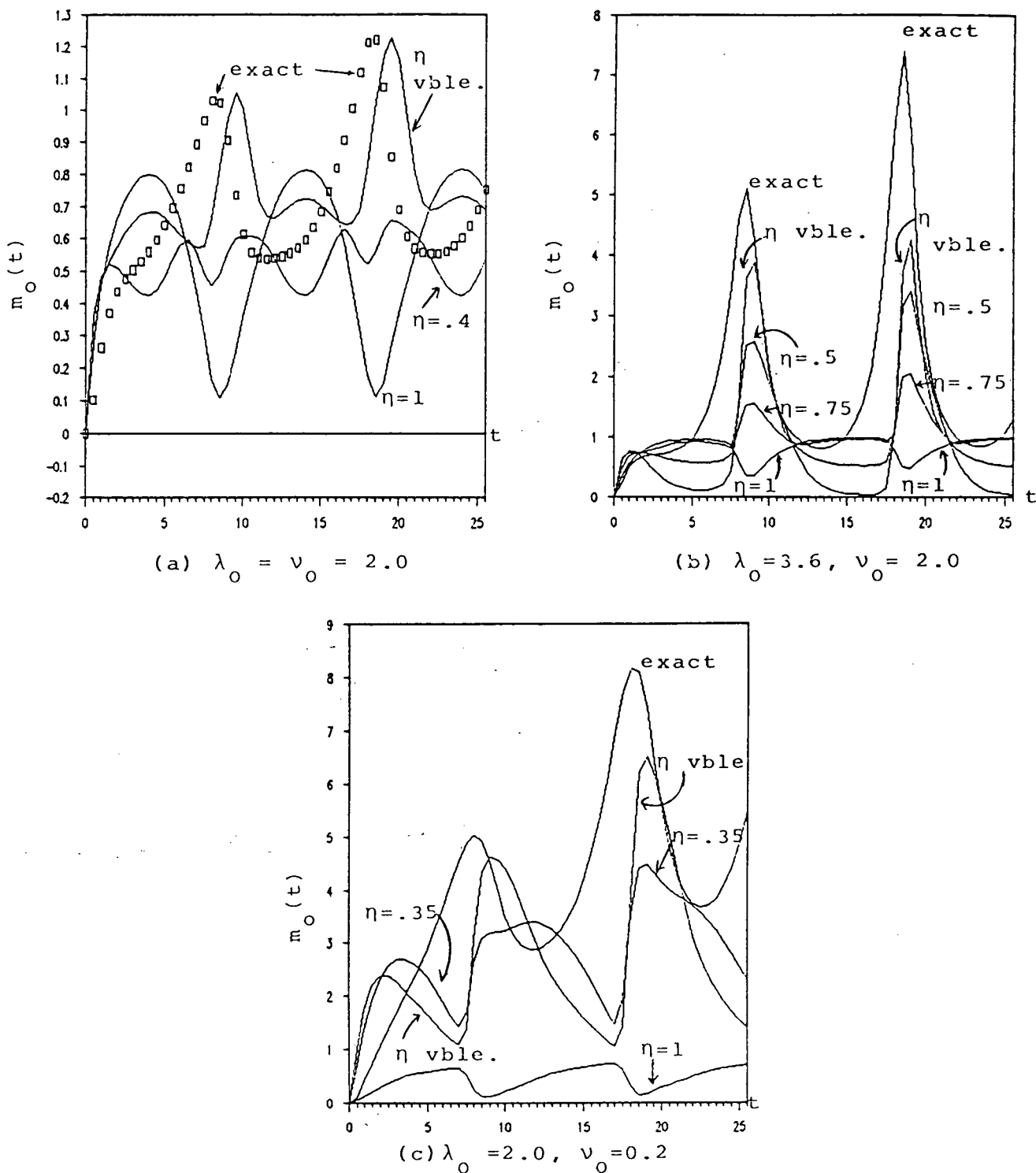
**FIGURE 2.18 Performance of the Closure Method 2E When the Retrial Rate is Much Smaller Than the New Customer Arrival Rate.**

Figures 2.18a, b, c, illustrated the limited improvement represented by equation (2.4.9) when the retrial rate is significantly lower than the new customer arrival rate. The notation 'all  $\eta$ ' means the sequence of values  $\eta = 1, 0.75, 0.5, 0.25, 0.1$ , and  $0.05$ , as well as the time varying  $\eta(t) = 1 - \Pi_0(t)$ . Figure 2.18d provides a contrast to the first three Figures here with a similar plot when  $\lambda$  and  $v$  have more similar values.



**FIGURE 2.19 Comparison of Exact and Method 2E Approximation  $v(t)$  for Various Values of  $\eta$  in Equation (2.4.9) in the Oscillatory Parameter Case.**

For all of the calculations represented here,  $\mu(t) = 4.0$ . Appendix A gives details of calculations done to obtain these exact quantities.



**FIGURE 2.20** Comparison of Exact  $m_0(t)$  with  $m_0(t)$  Given by Equation (2.4.9) for Selected Values of  $\eta$  in an Oscillatory Parameter Retrial Queue.

The reason equations (2.4.1) with (2.4.6) give the poor results observed earlier seems clear from comparing the  $\eta = 1$  curves here with the exact  $m_0(t)$ . What is surprising is the accuracy of the results achieved when equation (2.4.9) is used with (2.4.1) and intermediate values of  $\eta$ . It is clear that equation (2.4.9) does not generally constitute an acceptably accurate way to compute  $m_0(t)$  for its own sake.

## 2.5 Closure Methods With Three or More Equations

*Grandma Sonderquist's Conclusion: A chicken doesn't stop scratching just because the worms are scarce.*

[1001 Logical Laws..., Peers and Bennett]

### Three Equation Methods

Between the two two-equation closure methods described and analyzed in the preceding two sections, we are able to compute approximate values for all four the quantities listed at the beginning of section 2.1 . Little has been said about the computation of  $\gamma_o(t)$  so far, but since both of these methods produce estimates of  $\Pi_o(t)$  and  $m_o(t)$ , it is easy to compute  $\gamma_o(t)$ .

We have used only some of the equations listed in Table 2, and that Table can be extended indefinitely. A reasonably concerted search has been made for additional viable closure methods, perhaps based on systems of more than two differential equations. One such method is described below, but in general the results of this search have not been particularly encouraging.

The main problem encountered in widening the scope of equations from Table 2 which were considered is in finding viable closure formulas. The quantities which must be eliminated from many of the equations not considered until now usually include moments or components of moments of order two or greater. These quantities have much more complicated steady state formulas, and they tend to vary over much wider ranges of values as a function of time.

We end up trying to approximate higher order moments by complicated expressions involving lower order moments which are not only difficult to test comprehensively, but also difficult to justify intuitively. It is not surprising that such approximations are usually not simply poor, but lead to equations with solutions that violate some very basic requirements (such as the occurrence of negative values for moments of queue length, or probabilities with values exceeding 1).

However, a three-equation closure method was found for computing  $\Pi_0(t)$ ,  $m(t)$ , and  $\gamma_0(t)$ . We start with the first, second, and fourth equations in Table 2, modifying them trivially to get the system

$$\begin{aligned}\Pi'_0(t) &= -\lambda \Pi_0 - \mu \Pi_1 - \nu \Pi_0 \gamma_0, \\ m'(t) &= \lambda - \mu \Pi_1, \\ \gamma'_0(t) &= \frac{m'_0 \Pi_0 - m_0 \Pi'_0}{\Pi_0^2},\end{aligned}\tag{1}$$

where

$$m'_0(t) = -\lambda \Pi_0 \gamma_0 - \nu \zeta_0 + \mu m_1 - \mu \Pi_1.\tag{2}$$

Since we have the exact relations

$$\Pi_1 = 1 - \Pi_0, \quad m_1 = m - m_0, \quad \text{and} \quad \gamma_0 = \frac{m_0}{\Pi_0},$$

the only closure formula we need is to eliminate  $\zeta_0(t)$  from equation (2).

Now, at steady state,

$$\zeta_0 = \frac{\rho c (\rho c + 1)}{1 - \rho} ,$$

and so with a little effort, we can come up with a list of potential closure formulas, including the two simplest,

$$\zeta_0 = c m_1 , \quad (3)$$

since  $m_1 = \rho(c\rho + 1)/(1 - \rho)$ , giving Method 3A, and

$$\zeta_0 = \frac{m_0 (m_0 + 1)}{\Pi_0} , \quad (4)$$

since  $m_0 = c\rho$ , and  $\Pi_0 = 1 - \rho$ , giving Method 3B.

A number of more complicated flights of fancy were included briefly in test calculations, but all resulted in catastrophic numerical breakdown of the solution of the resulting system of differential equations within one or two steps of the starting point .

What formula (3) has going for it is simplicity. In constant parameter calculations with  $\lambda = 2.0$ ,  $\mu = 4.0$ , and  $\nu = 2.0$ , results were almost misleadingly good. Figure 2.21 includes a plot of relative errors as well as plots comparing the exact and approximate quantities for this set of constant parameters. Note from Figure 2.21d that the approximate  $m_0(t)$  takes a bit of a negative dive for a short interval initially, and that the relative error in  $m_0(t)$  is

rather large initially.<sup>1</sup>

Figure 2.22 shows what happens when the corresponding oscillatory parameter calculation is done using equation (3). There is no real need to discuss magnitudes of relative errors here, though, for what it's worth, the first clipped peak in Figure 2.22d actually rises to a value of approximately 150, which is around 60 times the true value of  $\gamma_0(t)$  at that point. Reservations regarding use of formula (3) above in a closure method for any sort of calculations are difficult to avoid in view of Figure 2.22. This is perhaps a good example of the risks in using a closure formula based only on simple algebraic coincidence between steady state formulas, since it is difficult to think of reasons for  $\zeta_0$  to vary in simple direct proportion to  $m_1$  in general.

On the other hand, the formula (4) leads to a very accurate closure method here. This could well be due to the fact that not only is (4) true at steady state in the constant parameter case, but one could perhaps argue that  $\zeta_0$  will vary in a manner similar to the square of  $m_0^2$ . At any rate, for the constant parameter retrieval queue with  $\lambda = 2.0$ ,  $\mu = 4.0$ , and  $\nu = 2.0$ , relative errors in  $\Pi_0$ ,  $m$ , and  $m_0$ , never exceed a magnitude of 0.03, and when plotted on the same grid at the level of resolution employed in figures in this thesis, graphs of the approximate and exact quantities are virtually indistinguishable. Figure 2.23 compares plots of exact and approximate values of  $\Pi_0$ ,  $m$ , and  $m_0$ , using (4) for the oscillatory

---

<sup>1</sup>Figure 2.21a has been clipped somewhat. At  $t = 0.5$ , the first positive value of  $t$  for which an approximate value of  $m_0(t)$  was computed, the relative error in  $m_0(t)$  is -1.7. While this is a rather large value, the magnitude of the relative error does decrease rapidly to a range of values comparable with the relative errors in the other two quantities being computed.

parameter case with  $\lambda_0 = 2.0$ ,  $\mu = 4.0$ , and  $v_0 = 2.0$ . Clearly, when formula (4) is combined with equations (1) and (2), a closure method results which is of comparable accuracy to those methods discussed in previous sections.<sup>2</sup>

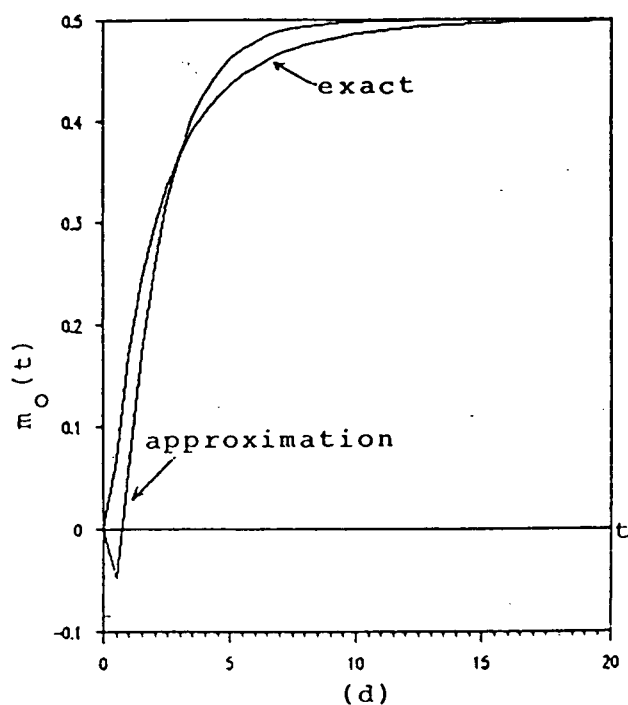
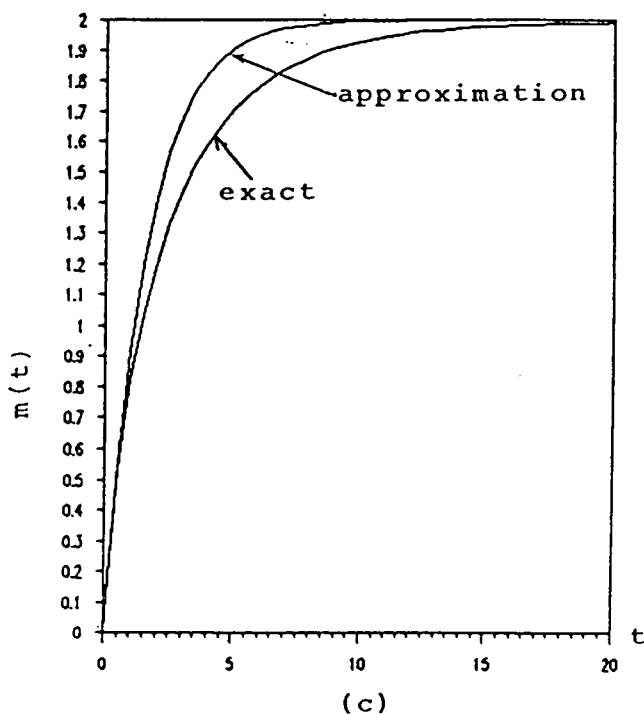
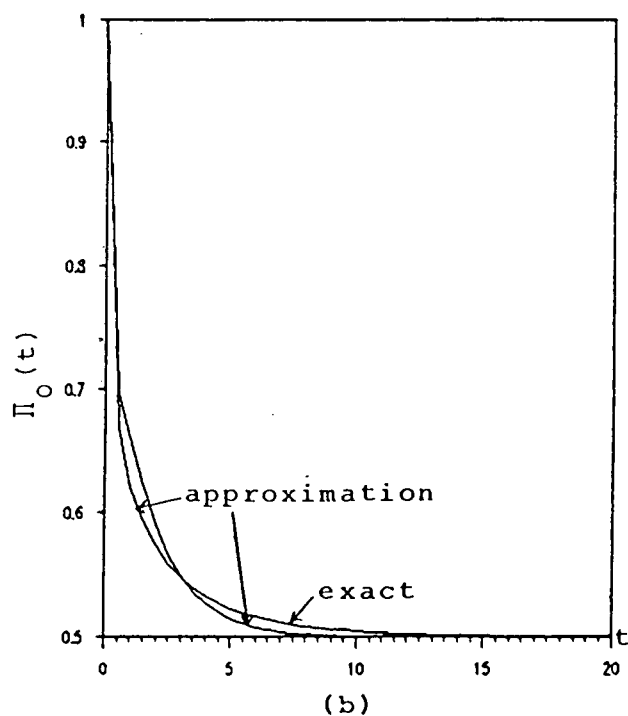
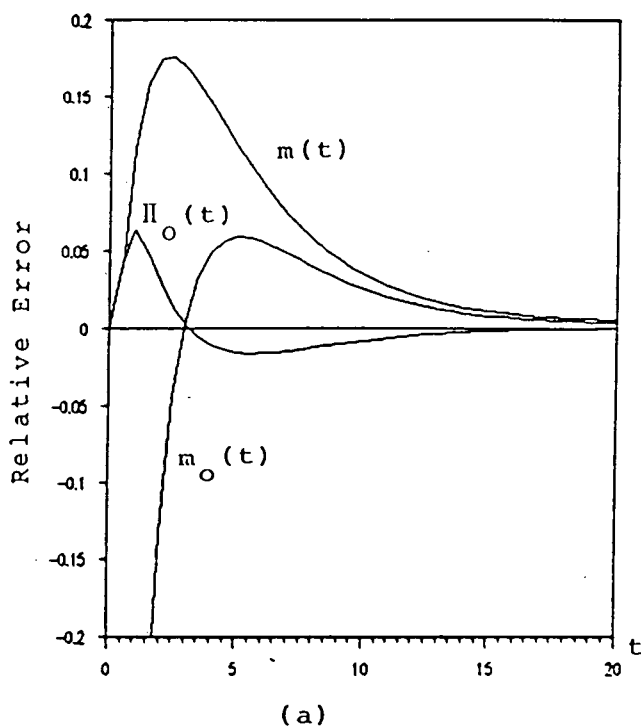
#### Closure Methods Involving Four or More Equations

Two attempts were made to find viable four-equation closure methods. One attempt involved the equations for  $\Pi'_0(t)$ ,  $m'(t)$ ,  $\zeta'_0(t)$ , and  $\zeta'_0(t)$  from Table 2 (a round-about way to get at  $\Pi_0$ ,  $m$ , and  $v$ ). This required a closure formula for  $m_0$  and  $\eta_0$ , the server-idle component of the third order moment of the system. A second line of work involved the equations for  $\Pi'_0(t)$ ,  $m'_0(t)$ ,  $m'(t)$ , and  $v'(t)$ , which required closure formulas for  $m_1$  and  $\zeta_0$ . All attempts here produced ludicrous results.

A five equation system using the equations for  $\Pi'_0(t)$ ,  $m'_0(t)$ ,  $m'(t)$ ,  $\zeta'_0(t)$ , and  $\zeta'_1(t)$  was tested briefly, but abandoned on the basis of early results. This system required only a single closure formula for  $\eta_0$ , but none of the four such potential expressions tested showed any promise at all.

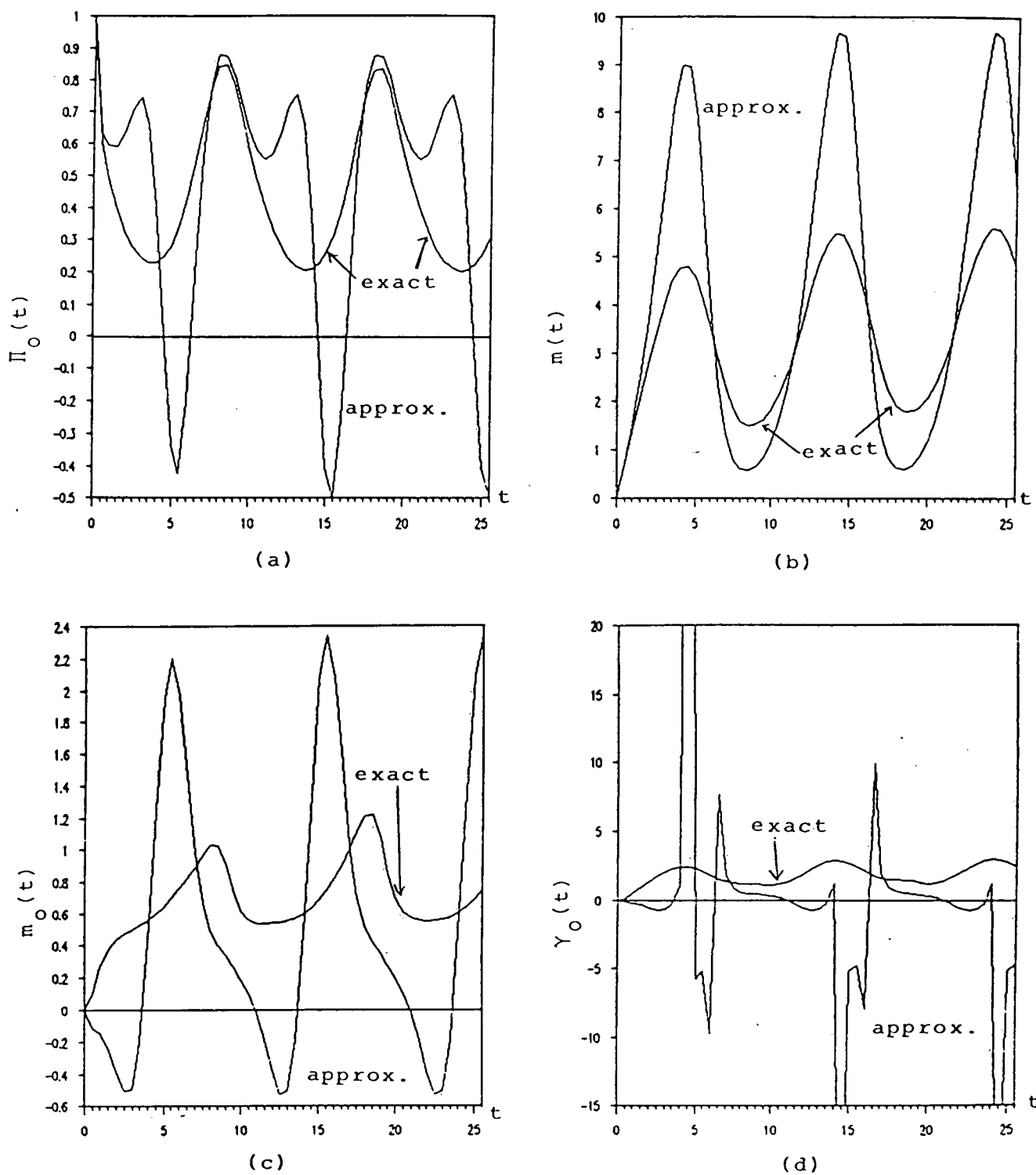
---

<sup>2</sup>Actually, one can use equations (1) and (2) as written to solve for  $\Pi_0(t)$ ,  $m(t)$  and  $\gamma_0(t)$ , or simply use the original equations for  $\Pi'_0(t)$ ,  $m'(t)$  and  $m'_0(t)$  in Table 2 with formula (4) to solve for  $\Pi_0(t)$ ,  $m(t)$  and  $m_0(t)$ , from which  $\gamma_0(t)$  can be computed easily.



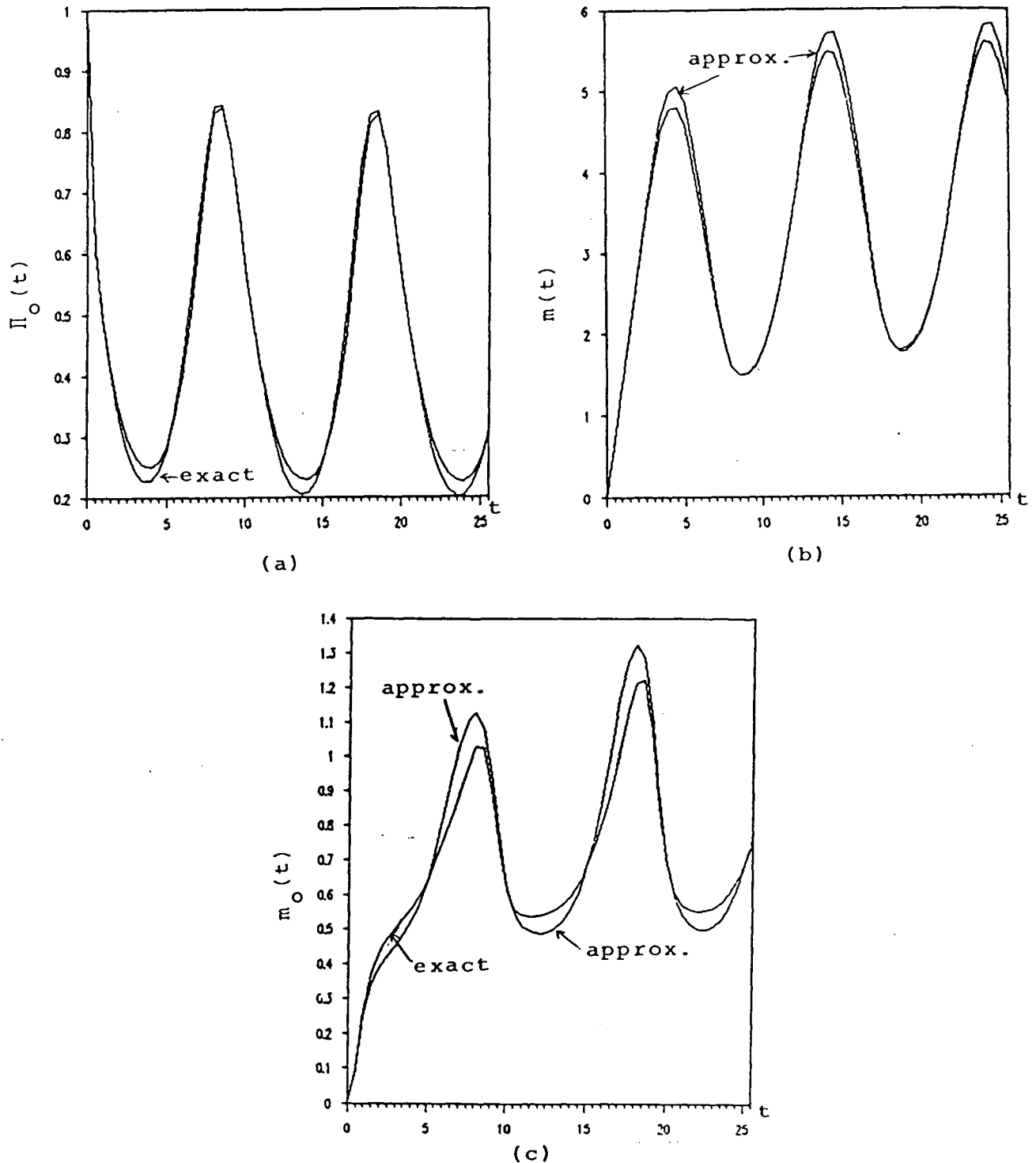
**FIGURE 2.21** Relative Errors and Comparison of Exact Quantities With the Results of the Closure Method 3A for a Constant Parameter Retrial Queue.

The calculations represented here correspond to  $\lambda = 2.0$ ,  $\mu = 4.0$  and  $\nu = 2.0$ . Appendix A gives details of calculations done to obtain the exact quantities used in the preparation of these plots.



**FIGURE 2.22 Comparison of Exact Quantities With the Results of the Closure Method 3A for an Oscillatory Parameter Retrial Queue.**

The calculations represented here correspond to  $\lambda_0 = 2.0$ ,  $\mu(t) = 4.0$ ,  $\nu_0 = 2.0$  and a period of oscillation of 10.0 time units. Appendix A gives details of calculations done to obtain these exact quantities.



**FIGURE 2.23** Comparison of Exact Quantities With the Results of the Closure Method 3B for an Oscillatory Parameter Retrial Queue.

The calculations represented here correspond to  $\lambda_0 = 2.0$ ,  $\mu(t) = 4.0$ ,  $\nu_0 = 2.0$  and a period of oscillation of 10.0 time units. Appendix A gives details of calculations done to obtain these exact quantities.

## 2.6 Summary and Conclusions

In each of sections 2.2 through 2.5, we have described and analyzed one or more distinct closure methods for the single server retrial queue along with computations illustrating the accuracy of the approximate quantities produced and discussed why we might expect a certain degree of accuracy in each case.

Of the methods studied, three were found to give results of the same degree of accuracy as the closure method due to Rothkopf and Oren [1979]. These are summarized in Table 7. Only one of these three serviceable methods produces estimates of the variance of the queue length,  $v(t)$ , whereas all three give estimates of the probability that the system is empty,  $\Pi_0(t)$ , and the mean number of customers in the system,  $m(t)$ . It is also possible to compute estimates of  $m_0(t)$ , the server-idle contribution to the mean number of customers in the system or  $\gamma_0(t)$ , the conditional expected number of customers in the system given the server is idle, using any of the three best methods described, but the accuracy of these estimates has been examined only for Method 3B.

The quantities  $\Pi_0(t)$  and  $m(t)$  can be obtained using any of the three closure methods found viable. Experience so far does not indicate that any one of these three methods is markedly superior to the others if  $\Pi_0(t)$  and  $m(t)$  are the only quantities of interest. Use of Method 2E in this case may be attractive, since that method has been studied most comprehensively.

The primary value of these closure techniques is in the reduction in computational labour they allow compared to determining the same quantities 'exactly' by solving an appropriately truncated subsystem of equations (1.1.3). Instead of solving a system of dozens or perhaps hundreds of simultaneous ordinary linear differential equations, we can get good approximations for these quantities by solving a system of two or three such equations only. The reduction in dimension of the systems to be solved when closure methods are used brings their solution within the capability of currently available microcomputer systems.

This work also demonstrates that Rothkopf and Oren's closure approach can be usefully modified to apply to more complicated queuing regimes, and no doubt to other probability models involving systems of ordinary linear differential equations.

We do note however, that the work in this thesis applies only to the single server retrial queue, with all inter-arrival, inter-retrial, and service times assumed exponentially distributed. The work of Rothkopf and Oren would seem to imply that extension of these techniques to multi-server retrial queues with arrival and service times exponentially distributed would be easier to do than to allow alternative and inevitably more complicated probability distributions. One would not be surprised if the development of closure methods turned out to be very difficult if not impossible for systems which could not be described by a system of ordinary linear differential equations. In Chapter 1, mention was made of a number of alternative approaches proposed for non-exponential queue systems.

**TABLE 7. Summary of Serviceable Closure Methods for Computing the Time-Dependent Behaviour of the Single-Server Retrial Queue.**

METHOD 2C:  $\Pi_0(t), m(t)$  (Two Equations)

$$\begin{aligned}\Pi'_0(t) &= -\Pi_0(t) + \mu\Pi_1(t) - \nu m(t) + \nu m_1(t) \\ m'(t) &= \lambda - \mu\Pi_1(t)\end{aligned}\tag{2.3.1}$$

with

$$\begin{aligned}\Pi_1(t) &= 1 - \Pi_0(t) \\ m_1(t) &= \eta m(t) \frac{c\Pi_1(t)+1}{c+1} + (1-\eta) \frac{\Pi_1(t) \{c\Pi_1(t)+1\}}{\Pi_0(t)}\end{aligned}\tag{2.3.6}$$

(Calculations reported in this thesis used the arbitrary value  $\eta = 1/2$ , and no attempt was made to determine if this was an optimal value to use for  $\eta$ , nor was any search made to find values of  $\eta$  leading to greater accuracy.)

METHOD 2E:  $\Pi_0(t), m(t), v(t)$  (Two Equations)

$$m'(t) = \lambda - \mu\{1 - \Pi_0(t)\}\tag{2.4.1}$$

$$v'(t) = \lambda + \mu - \mu\Pi_0(t)\{2m(t) + 1\} + 2\mu m_0(t)$$

with

$$\Pi_0(t) = \begin{cases} \left(\frac{m}{v}\right)^{m^2/\{(v-m)(c+1)\}} & \text{if } v \neq m \\ e^{-m/(c+1)} & \text{if } v = m \end{cases}\tag{2.4.4}$$

$$m_0(t) = \eta c\{1 - \Pi_0(t)\} + (1-\eta)\{m(t)\Pi_0(t) - \Pi_1(t)\}\tag{2.4.9}$$

$$\Pi_1(t) = 1 - \Pi_0(t)$$

(when  $\lambda \approx \nu$ , choose  $\eta$  between 0.25 and 0.5,  
when  $\lambda \gg \nu$ , choose  $\eta = 0.1$ )

TABLE 7, continued.....

METHOD 3B:  $\Pi_0(t)$ ,  $m(t)$ ,  $m_0(t)$  or  $\gamma_0(t)$  (Three Equations)

$$\Pi'_0(t) = -\lambda \Pi_0(t) - \mu \Pi_1(t) - \nu \Pi_0(t) \gamma_0(t) \quad (2.5.1)$$

$$m'(t) = \lambda - \mu \Pi_1(t)$$

$$m'_0(t) = -\lambda \Pi_0(t) \gamma_0(t) - \nu \zeta_0(t) + \mu m_1(t) - \mu \Pi_1(t) \quad (2.5.2)$$

or

$$\gamma'_0(t) = \frac{m'_0(t) \Pi_0(t) - m_0(t) \Pi'_0(t)}{\{m_0(t)\}^2}$$

with

$$\Pi_1(t) = 1 - \Pi_0(t)$$

$$m_1(t) = m(t) - m_0(t)$$

$$\gamma_0(t) = m_0(t) / \Pi_0(t)$$

$$\zeta_0(t) = \frac{m_0(t) \{m_0(t) + 1\}}{\Pi_0(t)} \quad (2.5.4)$$

## APPENDIX A

### DETAILS OF CALCULATION OF EXACT QUEUE PROPERTIES

As described in Section 1.3, exact values of time dependent probabilities required to compute exact properties of the various queues considered were determined by solving truncations of the systems (1.1.3a,b) for the M/M/1/1 retrial queue and (1.4.1) for the simple M/M/1 queue using the RKF45 algorithm.

The dimension of the actual system solved numerically was set by specifying a maximum value of  $n$  (the number of customers in the system for the simple M/M/1 queue and the number of customers in orbit for the M/M/1/1 retrial queue). The RKF45 algorithm then extrapolated a numerical value for each of the probabilities corresponding to all values of  $n$  smaller than the specified maximum value, beginning with a fixed initial value at time zero. In all cases, solution estimates were obtained in steps of  $\Delta t = 0.5$  starting with  $t = 0$ .

At the end of each such step, all of the computed probabilities were summed, and the residual obtained by subtracting this sum from unity was a measure of the error caused by solving only the finite subset of the infinite system of differential equations describing the queue. The maximum value of  $n$  to be considered was generally chosen so that this residual probability never became greater than  $10^{-6}$ , though in some calculations used, this residual attained larger values over brief time intervals.

To partially compensate for this truncation error, the residual was added into the maximum index probability for the M/M/1 queue or pro rata into the maximum index probabilities for the M/M/1/1 retrial queue.

The actual maximum values of  $n$  used in calculation of exact probabilities for test calculations reported in this thesis and the maximum values attained by the probability residuals in each case are tabulated below. It should be noted that the quoted residuals are the maximum rather than typical values that occurred over the entire time range considered.

#### Simple M/M/1 Queue:

<u><math>\rho</math></u>	<u>Constant Parameter</u>		<u>Oscillatory Parameter</u>	
0.1	25	$<10^{-12}$	25	$<10^{-12}$
0.3	30	$<10^{-12}$	30	$1.0 \times 10^{-12}$
0.5	30	$4.6 \times 10^{-10}$	30	$5.6 \times 10^{-7}$
0.7	35	$8.7 \times 10^{-7}$	50	$2.6 \times 10^{-7}$
0.9	65	$5.3 \times 10^{-8}$	70	$2.6 \times 10^{-5}$
0.95	80	$1.4 \times 10^{-9}$		

**M/M/1/1 Retrial Queue ( $c = 1$ ):**

$\rho$	<u>Constant Parameter</u>		<u>Oscillatory Parameter</u>	
0.1	20	$<10^{-12}$	20	$6.0 \times 10^{-12}$
0.3	20	$2.4 \times 10^{-11}$	20	$8.1 \times 10^{-8}$
0.5	30	$3.0 \times 10^{-9}$	29	$3.9 \times 10^{-5}$
0.7	35	$6.1 \times 10^{-6}$	50	$2.2 \times 10^{-5}$
0.9	50	$5.0 \times 10^{-5}$	65	$5.4 \times 10^{-4}$
0.95	60	$1.1 \times 10^{-5}$		

**M/M/1/1 Retrial Queue ( $\rho = 0.5$ ):**

$c$	<u>Constant Parameter</u>		<u>Oscillatory Parameter</u>	
0.2	20	$4.1 \times 10^{-7}$	20	$2.5 \times 10^{-4}$
0.5	20	$8.4 \times 10^{-7}$	20	$4.4 \times 10^{-4}$
1.0	30	$3.0 \times 10^{-9}$	29	$3.9 \times 10^{-5}$
2.0	25	$4.7 \times 10^{-7}$	25	$4.1 \times 10^{-4}$
5.0	30	$4.1 \times 10^{-7}$	30	$5.0 \times 10^{-4}$
10.0	35	$2.3 \times 10^{-7}$	35	$4.0 \times 10^{-4}$

## BIBLIOGRAPHY

- Alexsandrov, A. M., [1974], "A Queueing System With Repeated Orders", *Engineering Cybernetics Rev.*, 12, pages 1 - 4 .
- Burden, Faires, and Reynolds, [1978], *Numerical Analysis*, Prindle, Weber and Schmidt, Boston.
- Carter, G. M., and Cooper, R. B., [1972], "Queues with Service in Random Order", *Operations Research*, 20, pages 389 - 405 .
- Choo, Q. H., [1978], *The Interaction of Theory and Simulation in Queueing Analysis*, Ph. D. Thesis, Chelsea College, University of London.
- Choo, Q. H., and Conolly, B., [1979], "New Results in the Theory of Repeated Orders Queueing Systems", *J. Appl. Prob.*, 16, pages 631 - 640 .
- Clark, G. M., [1981], "Use of Polya Distributions in Approximate Solutions to Nonstationary M/M/s Queues", *Communications of the ACM*, 24, pages 206 - 217 .
- Cohen, J. W., [1957], "Basic Problems of Telephone Traffic Theory and the Influence of Repeated Calls", *Phillips Telecomm. Rev.* 18, pages 49 - 100 .
- Cooper, Robert B., [1981], *Introduction to Queueing Theory*, Second Edition, North Holland, New York.
- Falin, G. I., [1979], "A Single-Line System With Secondary Orders", *Engineering Cybernetics Rev.*, 17, pages 76 - 83 .
- Felberg, E., [1970], "Klassische Runge-Kutta Formeln Vierter und Neidrigerer Ordnung mit Schrittweiten-Kontrolle und ihre Anwendung auf Wärmeleitungs Probleme.", *Computing*, 6, pages 61 - 71 .
- Gross, Donald, and Harris, Carl M., [1974], *Fundamentals of Queueing Theory*, John Wiley & Sons, New York.
- Hashida, O., and Kodaira, K., [1975], "Repeated Calls Resulting from Data Terminal Busy and Delayed Delivery Service", *Electronics and Communications in Japan*, 58A, pages 1 - 9 .
- Keilson, J., Cozzolini, J., and Young, H., [1968], "A Service System With Unfilled Requests Repeated", *Operations Research*, 6, pages 1126 - 1137 .
- Keller, J. B., [1982], "Time-Dependent Queues", *SIAM Review*, 24, pages 401 - 412.
- Kelton, W. D., and Law, A. M., [1985], "The Transient Behaviour of the M/M/s Queue, With Implications for Steady-State Simulation", *Operations Research*, 33, pages 378 - 396 .

- Kendall, D. G., [1953], "Stochastic Processes Occurring in the Theory of Queues and Their Analysis by the Method of Imbedded Markov Chains", *Ann. Math. Statist.*, 24, pages 338 - 354 .
- Koopman, B. O., [1972], "Air-Terminal Queues Under Time-Dependent Conditions", *Operations Research*, 20, pages 1089 - 1114 .
- Kulkarni, V. G., [1982], Letter to the Editor, *J. Appl. Prob.*, 19, pages 901 - 904 .
- Kulkarni, V. G., [1983], "On Queueing Systems With Retrials", *J. Appl. Prob.*, 20, pages 380 - 389 .
- Lyashenko, N. N., and Nikulin, M. S., [1986], *Computer and Some Statistical Inference*, Technical Report #32, Department of Statistics, University of British Columbia.
- Massey, W. A., [1985], "Asymptotic Analysis of the Time Dependent M/M/1 Queue", *Mathematics of Operations Research*, 10, pages 305 - 327 .
- Morse, P. M., [1958], *Queues, Inventories, and Maintenance*, John Wiley & Sons, New York.
- Newell, G. F., [1968a], "Queues With Time-Dependent Arrival Rates I -- The Transition Through Saturation", *J. Appl. Prob.*, 5, pages 436 - 451 .
- Newell, G. F., [1968b], "Queues With Time-Dependent Arrival Rates II -- The Maximum Queue and The Return to Equilibrium", *J. Appl. Prob.*, 5, pages 579 - 590 .
- Newell, G. F., [1968c], "Queues With Time-Dependent Arrival Rates III -- A Mild Rush Hour", *J. Appl. Prob.*, 5, pages 591 - 606 .
- Odoni, A. R., and Roth, E., [1983], "An Empirical Investigation of the Transient Behaviour of Stationary Queueing Systems", *Operations Research*, 34, pages 432 - 455 .
- Ross, Sheldon M., [1983], *Stochastic Processes*, John Wiley & Sons, New York.
- Rothkopf, M. H., and Oren, S. S., [1979], "A Closure Approximation for the Nonstationary M/M/s Queue", *Management Science*, 25, pages 522 - 534 .
- Saaty, T. L., [1961], *Elements of Queueing Theory with Applications*, McGraw-Hill, New York.
- Shanks, E. B., [1966], "Solutions of Differential Equations by Evaluations of Functions", *Math. Comp.*, 20, pages 21 - 38 .
- Shiryayev, A. N., [1984], *Probability*, Springer-Verlag, New York.

Stidham, S., [1972], " $L = \lambda W$ : A Discounted Analogue and a New Proof", *Operations Research*, 20, pages 1115 - 1126 .

Stidham, S., [1974], "A Last Word on  $L = \lambda W$ ", *Operations Research*, 22, pages 417 - 421 .

Zachmanoglou, E. C. and Thoe, D. W., [1976], *Introduction to Partial Differential Equations With Applications*, Williams & Wilkins, Baltimore.