## Efficient Simulation and Performance Stabilization for Time-Varying Single-Server Queues

Ni Ma

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Arts and Sciences

### COLUMBIA UNIVERSITY

2019

©2019

Ni Ma

All Rights Reserved

## ABSTRACT

### Efficient Simulation and Performance Stabilization for Time-Varying Single-Server Queues

### Ni Ma

This thesis develops techniques to evaluate and to improve the performance of single-server service systems with time-varying arrivals. The performance measures considered are the time-varying expected length of the queue and the expected customer waiting time. Time varying arrival rates are considered because they often occur in service systems. For example, arrival rates often vary significantly over the hours of each day and over the days of each week. Stochastic textbook methods do not apply to models with time-varying arrival rates. Hence new techniques are needed to provide high quality of service when stationary steady-state analysis is not appropriate. In contrast to the extensive recent literature on many-server queues with time-varying arrival rates, we focus on singleserver queues with time-varying arrival rates. Single-server queues arise in real applications where there is no flexibility in the number of service facilities (servers). Different analysis techniques are required for single-server queues, because the two kinds of models exhibit very different performance. Many-server models are more tractable because methods for highly tractable infinite-server models can be applied. In contrast, single-server models are more complicated because it takes a long time to respond to a build up of workload when there is only one server.

The thesis is divided into two parts: simulation algorithms for performance evaluation and service-rate controls for performance stabilization. The first part of the thesis develops algorithms to efficiently simulate the single-server time-varying queue. For the generality considered, no explicit mathematical formulas are available for calculating performance measures, so simulation experiments are needed to calculate and evaluate system performance. Efficient algorithms for both standard simulation and rare-event simulation are developed.

The second part of the thesis develops service-rate controls to stabilize performance in the timevarying single-server queue. The performance stabilization problem aims to minimize fluctuations in mean waiting times for customers coming at different times even though the arrival rate is time-varying. A new service rate control is developed, where the service rate at each time is a function of the arrival rate function. We show that a specific service rate control can be found to stabilize performance. In turn, that service rate control can be used to provide guidance for real applications on optimal changes in staffing, processing speed or machine power status over time. Both the simulation experiments to evaluate performance of alternative service-rate controls and the simulation search algorithm to find the best parameters for a damped time-lag service-rate control are based on efficient performance evaluation algorithms in the first part of the thesis.

In Chapter Two, we present an efficient algorithm to simulate a general non-Poisson nonstationary point process. The general point process can be represented as a time transformation of a rate-one base process and by exploiting a table of the inverse cumulative arrival rate function outside of simulation, we can efficiently convert the simulated rate-one process into the simulated general point process. The simulation experiments can be conducted in linear time subject to small error bounds. Then we can apply this efficient algorithm to generate the arrival process, the service process and thus to calculate performance measures for the  $G_t/G_t/1$  queues, which are single-server queues with time-varying arrival rates and service rates. Service models are constructed for this purpose where time-varying service rates are specified separately from the rate-one service requirement process, and service times are determined by equating service requirements with integrals of service rates over a time period equal to the service time.

In Chapter Three, we develop rare-event simulation algorithms in periodic  $GI_t/GI/1$  queues and further in  $GI_t/GI_t/1$  queues to estimate probabilities of rare but important events as a sanity check of the system, for example, estimating the probability that the waiting time is very long. Importance sampling, specifically exponential tilting, is required to estimate rare-event probabilities because in standard simulation, the number of experiments may blow up to achieve a targeted relative error and for each experiment, it may take a very long time to determine that the rare event does not happen. To extend the rare-event simulation algorithm to periodic queues, we derive a convenient expression for the periodic steady-state virtual waiting time. We apply this expression to establish bounds between the periodic workload and the steady-state workload in stationary queues, so that we can prove that the exponential tilting algorithm with the same parameter efficient in stationary queues is efficient in the periodic setting as well, which has a bounded relative error. We apply this algorithm to compute the periodic steady-state distribution of reflected periodic Brownian motion with support of a heavy-traffic limit theorem and to calculate the periodic steady-state distribution and moments of the virtual waiting time. This algorithm's advantage in calculating these distributions and moments is that it can directly estimate them at a specific position of the cycle without simulating the whole queueing process until steady state is reached for the whole cycle.

In Chapter Four, we conduct simulation experiments to validate performance of four service-rate controls: the rate-matching control, which is directly proportional to the arrival rate, two squareroot controls related to the square root staffing formula and the square-root control based on the mean stationary waiting time. Simulations show that the rate-matching control stabilizes the queue length distribution but not the virtual waiting time. This is consistent with established theoretical results, which follow from the observation that with rate-matching control, the queueing process becomes a time transformation of the stationary queueing process with constant arrival rates and service rates. Simulation results also show that the two square-root controls analogous to the server staffing formula are not effective in stabilizing performance. On the other hand, the alternative square-root service rate control based on the mean stationary waiting time approximately stabilizes the virtual waiting time when the cycle is long so that the arrival rate changes slowly enough.

In Chapter Five, since we are mostly interested in stabilizing waiting times in more common scenarios when the traffic intensity is not close to one or when the arrival rate does not change slowly, we develop a damped time-lag service-rate control that performs fairly well for this purpose. This control is a modification of the rate-matching control involving a time lag and a damping factor. To find the best parameters for this control, we search over reasonable intervals for the most timestable performance measures, which are computed by the extended rare-event simulation algorithm in  $GI_t/GI_t/1$  queue. We conduct simulation experiments to validate that this control is effective for stabilizing the expected steady-state virtual waiting time (and its distribution to a large extent). We also establish a heavy-traffic limit with periodicity in the fluid scale to provide theoretical support for this control. We also show that there is a time-varying Little's law in heavy-traffic, which implies that this control cannot stabilize the queue length and the waiting time at the same time.

## **Table of Contents**

Li	List of Figures			
List of Tables			x	
1	Intr	oduction	1	
	1.1	Rare-Event Simulation in Queues	4	
		1.1.1 Literature Review	9	
	1.2	Time-Varying Single-Server Queues	9	
		1.2.1 Literature Review	14	
	1.3	Main Contributions	15	
	1.4	Outline	17	
Ι	Per	rformance Evaluation for Time-Varying Single-Server Queues	19	
I 2	Per Effic	rformance Evaluation for Time-Varying Single-Server Queues cient Simulation of Time-Varying Queues	19 20	
I 2	Per Effic 2.1	rformance Evaluation for Time-Varying Single-Server Queues cient Simulation of Time-Varying Queues Introduction	<ul><li>19</li><li>20</li></ul>	
I 2	Per Effic 2.1 2.2	cformance Evaluation for Time-Varying Single-Server Queues         cient Simulation of Time-Varying Queues         Introduction       Introduction         The $G_t/G_t/1$ Model       Introduction	<ol> <li>19</li> <li>20</li> <li>21</li> </ol>	
I 2	Per Effic 2.1 2.2 2.3	cient Simulation of Time-Varying Queues         Introduction       Introduction         The $G_t/G_t/1$ Model       Introduction         Efficient Simulation       Introduction	<ol> <li>19</li> <li>20</li> <li>21</li> <li>23</li> </ol>	
I 2	Per Effic 2.1 2.2 2.3	cient Simulation of Time-Varying Queues         Introduction       Introduction         The $G_t/G_t/1$ Model       Introduction         2.3.1       Efficiency Through Re-Use	<ol> <li>19</li> <li>20</li> <li>21</li> <li>23</li> <li>24</li> </ol>	
I 2	Per Effic 2.1 2.2 2.3	cient Simulation of Time-Varying Queues         Introduction       Introduction         The $G_t/G_t/1$ Model       Introduction         2.3.1       Efficiency Through Re-Use         2.3.2       The Inverse Function	<ol> <li>19</li> <li>20</li> <li>21</li> <li>23</li> <li>24</li> <li>25</li> </ol>	
I 2	Per Effic 2.1 2.2 2.3	cient Simulation of Time-Varying Queues         Introduction	<ol> <li>19</li> <li>20</li> <li>21</li> <li>23</li> <li>24</li> <li>25</li> <li>29</li> </ol>	
I 2	Per Effic 2.1 2.2 2.3	cient Simulation of Time-Varying Queues         Introduction	<ol> <li>19</li> <li>20</li> <li>21</li> <li>23</li> <li>24</li> <li>25</li> <li>29</li> <li>32</li> </ol>	

3	Rar	e-Ever	nt Simulation for Periodic Queues	34
	3.1	Introd	uction	35
		3.1.1	Using Bounds to Connect to Stationary Methods	36
		3.1.2	A Convenient Representation	37
		3.1.3	Stylized Sinusoidal Examples	38
		3.1.4	Organization of the Chapter	38
	3.2	Revers	se-Time Representation of the Workload Process	39
	3.3	Bound	ls and Approximations	40
		3.3.1	Basic Bounds	41
		3.3.2	Tail Asymptotics	42
	3.4	Simula	ation Methodology	43
		3.4.1	Exponential Tilting for the $GI/GI/1$ Model	44
		3.4.2	Waiting Time in $GI/GI/1$ Model	45
		3.4.3	Workload in $GI/GI/1$ Model	46
		3.4.4	Applying the Bounds to the Periodic Case	48
		3.4.5	The Mean and Variance	50
		3.4.6	The Algorithm	51
		3.4.7	Computational Complexity and Running Times	53
	3.5	Simula	ation Examples	55
		3.5.1	Tail Probabilities	55
		3.5.2	Heavy-Traffic Scaling	56
		3.5.3	Hyperexponential Examples	58
		3.5.4	Estimating the Moments of $W_y$	63
	3.6	The S	upporting Heavy-Traffic FCLT	65
		3.6.1	The Heavy-Traffic FCLT	67
		3.6.2	Approximations for the Steady-State Workload	69
		3.6.3	Application to the Sinusoidal Arrival-Rate Function	70
		3.6.4	Approximations for the $G_t/G/1$ Model	71
	3.7	Conclu	usions	71

Π	Pe	erform	ance Stabilization in Time-Varying Single-Server Queues	73
4	The	Servi	ce-Rate Control Problem	<b>74</b>
	4.1	Introd	uction	74
	4.2	The Se	ervice-Rate Controls	76
		4.2.1	The Rate-Matching Control	76
		4.2.2	Two Square-Root Controls	78
		4.2.3	The PSA-Based Square-Root Control	79
	4.3	The E	xperiments	80
		4.3.1	Estimating Performance Measures	80
		4.3.2	The Study Cases	81
	4.4	Simula	ation Results	82
		4.4.1	The Rate-Matching Control	82
		4.4.2	Two Square-Root Controls	83
		4.4.3	The PSA Square-Root Control	86
	4.5	Conclu	usions	86
5	Dan	nped 7	Time-Lag Service-Rate Control	89
	5.1	Introd	uction	90
		5.1.1	A Nonstationary Stochastic Design Problem	90
		5.1.2	Related Literature	90
		5.1.3	The Rate-Matching Control	92
		5.1.4	The Open Problem	92
		5.1.5	Formulation of Optimal Control Problems	93
		5.1.6	Organization of the Chapter	94
	5.2	Simula	ation Examples	95
	5.3	The K	ey Technical Tools	97
		5.3.1	A Simulation Search Algorithm	97
		5.3.2	Heavy-Traffic Limits	98
	5.4	The M	Iodel	99
		5.4.1	The Arrival Process	99

	5.4.2	The General $G_t/G_t/1$ Model $\ldots$	100
	5.4.3	The Periodic $G_t/G_t/1$ Model $\ldots$	101
5.5	5 The	Simulation Search Algorithm	104
	5.5.1	The $GI_t/GI_t/1$ Model	104
	5.5.2	The Extended Algorithm	105
	5.5.3	Explicit Representations for the Sinusoidal Case	106
	5.5.4	The Search Algorithm	106
5.6	5 Supp	orting Heavy-Traffic Limits	106
	5.6.1	The Underlying Rate-One Processes	107
	5.6.2	A Family of Models	108
	5.6.3	The Scaled Queueing Processes	110
	5.6.4	The HT FWLLN	111
	5.6.5	The HT FCLT	113
5.7	7 Proo	f of Theorem 5.6.2	116
5.8	A H	$\Gamma$ FCLT with Periodicity $\ldots$	119
	5.8.1	An Alternative Family of Models	119
	5.8.2	An Associated Family of Service-Rate Controls	120
	5.8.3	The Scaled Queueing Processes	120
	5.8.4	The HT FCLT with Periodicity	121
5.9	9 Simu	lation Examples	123
	5.9.1	In the Setting of $\S5.6$	123
	5.9.2	In the Alternative Scaling of $\S5.8$	135
5.1	10 Conc	clusions	136
III	Biblio	graphy	139
Bibli	ograph	У	140
IV	Apper	ndices	150

A Supplement to Chapter Three

151

	A.1	Introd	uction	151
	A.2	More	Results on the Asymptotics	151
		A.2.1	Tail Asymptotics for RPBM	151
		A.2.2	Asymptotic Expansions for the Asymptotic Decay Rate	152
		A.2.3	Approximations for the Asymptotic Decay Rate	153
		A.2.4	Exact Values for the Asymptotic Decay Rate	154
		A.2.5	More Bounds	159
		A.2.6	Heuristic Approximations	160
	A.3	Simula	ation Results	160
		A.3.1	Tail Probabilities for the $M_t/M/1$ Periodic Queue with $\bar{\lambda} = 1$	162
		A.3.2	Tail Probabilities for the $M_t/M/1$ Periodic Queue with $\mu = 1$	177
		A.3.3	Tail Probabilities for the $(H_2)_t/M/1$ Periodic Queue $\ldots \ldots \ldots \ldots \ldots$	190
		A.3.4	Tail Probabilities for the $M_t/H_2/1$ Periodic Queue $\ldots \ldots \ldots \ldots \ldots \ldots$	201
		A.3.5	Estimates of the Mean and Standard Deviation	213
		A.3.6	The Impact of the Adjustment in $\S3.5.3$ and $\S3.5.4$	218
в	Mat	lab Co	ode for Simulation	<b>221</b>
	B.1	Chapt	er 2 Matlab Code	221
	B.2	Chapt	er 3 Matlab Code	230
	B.3	Chapt	er 5 Matlab Code	236

# List of Figures

1.1	Average hourly number of airplane arrivals to J.F.K (Terminal A in the plot) and	
	LaGuardia (Terminal B in the plot) airports.	2
1.2	Elements of a single-server queueing system with a waiting room and a service facility.	2
1.3	The arrival rate function $\lambda(t)$ in an example of a deterministic TV single-server queue	
	with a constant service rate $\mu$	11
1.4	This figure plots the expected steady-state virtual waiting time process as a function	
	of time in a cycle for a TV single-server queue. The arrival rate function is $\lambda(t) =$	
	$0.8 \times (1 + 0.2\sin(0.1t))$ and the arrival process is Markovian. The blue line shows	
	the waiting time when the service rate is a constant $\mu = 1$ , the red line shows the	
	waiting time with the rate-matching service-rate control and the yellow line shows	
	the waiting time with the damped time-lag service-rate control. Dashed black lines	
	show confidence intervals, which are quite narrow.	13
4.1	Estimated $E[Q(t)]$ for the rate-matching control in the $M_t/M_t/1$ system with differ-	
	ent $\gamma$ : 0.001 (left), 0.1 (middle) and 10 (right)	82
4.2	Comparison of estimated $E[W(t)]$ to its heavy traffic approximation in (4.15) under	
	the rate-matching control in $M_t/M_t/1$ system with different values of $\gamma$ : 0.001 (top	
	left), 0.01 (top right), 0.1 (bottom left) and 1 (bottom right). $\ldots$	83
4.3	Estimated $E[Q(t)]$ for the rate-matching control in three different $G_t/G_t/1$ models,	
	$(H_2/H_2)$ , $(H_2/E_2)$ and $(E_2/E_2)$ , and three different values of $\gamma$ : 0.001 (left), 0.1	
	(middle) and 10 (right).	84

4.4	Estimated $E[W(t)]$ for the rate-matching control in three different $G_t/G_t/1$ models,
	$(H_2/H_2)$ , $(H_2/E_2)$ and $(E_2/E_2)$ , and three different values of $\gamma$ : 0.001 (left), 0.1
	(middle) and 10 (right) $\ldots \ldots $
4.5	Estimated means $E[Q(t)]$ and $E[W(t)]$ for the square-root control in (4.17) for the
	$M_t/M_t/1$ system with different values of $\gamma:$ 0.001 (left), 0.1 (middle) and 1 (right) 85
4.6	Estimated means $E[Q(t)]$ and $E[W(t)]$ for the square-root control in (4.18) for the
	$M_t/M_t/1$ system with different values of $\gamma:$ 0.001 (left), 0.1 (middle) and 1 (right) 85
4.7	Estimated $E[Q(t)]$ and $E[W(t)]$ for the PSA square-root control in (4.20) in the
	$M_t/M_t/1$ model for two values of $\gamma$ : 0.001 (left) and 0.1 (right)
4.8	Estimated $E(W(t))$ (left column) and $E(Q(t))$ (right column) for the PSA square-
	root control in (4.20) in three $G_t/G_t/1$ models, $(H_2/H_2)$ , $(H_2/E_2)$ and $(E_2/E_2)$ , for
	two values of $\gamma$ : 0.001 (top) and 0.1 (bottom)
5.1	Estimates of the periodic steady-state values of $E[W_y]$ (blue solid line), $E[L_y]$ (red
	dashed line) and $\lambda(y)E[W_y]$ (green dotted line) for the optimal control $(\eta^*, \xi^*)$ for
	the sinusoidal example with parameter triples $(\rho, \beta, \gamma) = (0.8, 0.2, 0.1)$ (left) and
	$(0.95, 0.2, 0.1)$ (right), so that the cycle length is $C = 2\pi/\gamma = 62.8$ . The optimal
	controls are (5.84, 0.84) for $\rho = 0.8$ and (15.1, 2.13) for $\rho = 0.95$ 96
5.2	Three-dimensional plots of estimates of the maximum waiting time $\max_{0 \le y \le C} \{E[W_y]\}$
	for $(\eta, \xi)$ in $[0, 20] \times [0, 5]$ (left) $[3, 9] \times [0.6, 1.0]$ (right)
5.3	Plots over one cycle for the example in Table 5.1. The first row shows the expected
	periodic steady-state virtual waiting time (solid blue line), the expected steady-state
	workload (the dashed red line) and arrival rate multiplied by the mean waiting time
	(the dotted green line) for $\rho$ = 0.8 (left) and $\rho$ = 0.95 (right) in the base case
	$(\beta, \gamma) = (0.2, 2.5)$ . The second row shows arrival rate divided by $\rho$ and service rate
	for $\rho = 0.8$ (left) and $\rho = 0.95$ in the same base case. The optimal control parameters
	are $(\eta_{\rho}^*, \xi_{\rho}^*) = (5.80, 0.84)$ for $\rho = 0.8$ and $(27.7, 0.93)$ for $\rho = 0.95$ . The maximum
	minus minimum of $E[W_y]$ over a cycle equals 0.0321 for $\rho = 0.8$ and 0.1425 for $\rho = 0.95.125$

- 5.4Plots over one cycle for the example in Table 5.1. Estimates of the periodic steadystate standard deviation  $SD[W_y]$  for  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right), shown in red solid line. Also displayed are the fluid arrival rate  $\lambda_f = 1 + s(t)$  (blue dashed Plots over one cycle for the example in Table 5.1. Estimates of the periodic steady-5.5state probability of delay  $P(W_y > 0)$  for  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right), shown in red solid line. Also displayed are the fluid arrival rate  $\lambda_f = 1 + s(t)$  (blue dashed Plots over one cycle for the example in Table 5.1. Estimates of the periodic steady-5.6state ccdf  $P(W_y > x)$  for four values of y: 0, 1/4, 1/2, 3/4 for  $\rho = 0.8$  (left) and Plots over one cycle for the example in Figure 5.1 with  $\rho = 0.95$  but without the 5.7heavy-traffic scaling. Estimates of the ccdf (left) and pdf (right) for four values of
  - y: 0, 1/4, 1/2, 3/4... 128
- 5.8 Estimates of the expected waiting time  $E[W_y]$  for the one-parameter  $\eta$  control with  $\xi \equiv 1$ , for the sinusoidal example in Figure 5.1 with parameter triple  $(\rho, \beta, \gamma) = (0.8, 0.2, 0.1)$  and  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right). For  $\rho = 0.8$ ,  $\eta_{\rho}^* = 5.93$  the maximum deviation is 0.4109 and the relative error is about 10%; for  $\rho = 0.95$ ,  $\eta_{\rho}^* = 28.3$ , the maximum deviation is 3.034 and the relative error is about 14%. . . . 129

- 5.14 the expected periodic steady-state virtual waiting time (the blue line) and the expected steady-state workload (the red line) for  $\rho = 0.96$ ,  $\beta = 0.04$ ,  $\gamma = 0.004$ ,  $\eta_{\rho} = 24$ ,  $\xi_{\rho} = 0.96$ , yielding a maximum deviation 0.0228 (left) and  $\rho = 0.98$ ,  $\beta = 0.02$ ,  $\gamma = 0.001$ ,  $\eta_{\rho} = 49$ ,  $\xi_{\rho} = 0.98$ , yielding a maximum deviation 0.0070 (right). 136

# List of Tables

1.1	Estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model with sinusoidal	
	arrival-rate function $\lambda(t) = 1 + 0.2 \times \sin(\gamma t)$ as a function of $\gamma$ and b for: $\rho = 0.8, \mu =$	
	1.25, y = 0 based on 5000 replications	8
3.1	Comparison of the steady-state workload and waiting-time tail probabilities for $b =$	
	4,20 in the stationary $H_2/M/1$ queue with $\rho = 0.1$ . The exact values are calculated	
	from Theorem X.5.1 of Asmussen (2003). $\ldots$	48
3.2	Estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model with sinusoidal	
	arrival-rate function in (2.4) as a function of $\gamma$ and b for: $\rho = 0.8, \bar{\lambda} = 1, \mu = 1.25$	
	and $\beta = 0.2$ based on 5000 replications	56
3.3	Comparison of the ratios $P(W_y > b_{\rho})/P(W > b_{\rho})$ , where W is for the stationary	
	model, for 5 different values of $1-\rho$ and 8 different values of the position $y$ within the	
	cycle in the $M_t/M/1$ model with sinusoidal arrival-rate function in (3.38) with the	
	parameter 4-tuple in (3.39) using the base parameter triple $(\beta, \gamma, b) = (1.0, 2.5, 4.0)$ .	58
3.4	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model for $y = 0.0$	
	and $y = 0.5$ as a function of $1 - \rho$ with base parameter triple $(\beta, \gamma, b) = (1, 2.5, 4)$ in	
	(3.39) based on 40,000 replications	59
3.5	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/H_2/1$ model for $y = 0.0$	
	and $y = 0.5$ as a function of $1 - \rho$ with base parameter triple $(\beta, \gamma, b) = (1, 2.5, 4)$ in	
	(3.39) based on 40,000 replications	60
3.6	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $(H_2)_t / M / 1$ model for $y =$	
	0.0 and $y = 0.5$ as a function of $1 - \rho$ with base parameter triple $(\beta, \gamma, b) = (1, 2.5, 4)$	
	in $(3.39)$ based on 40,000 replications	61

Estimated mean $E[W]$ and standard deviation $SD(W)$ as a function of $1 - \rho$ for five
cases of the stationary $M/M/1$ queue: $\mu = 1, \bar{\lambda} = \rho \dots \dots$
Estimated mean $E[W_y]$ and standard deviation $SD(W_y)$ as a function of $1 - \rho$ for
five cases of the $M_t/M/1$ queue at $y = 0.5$ : $\mu = 1, \bar{\lambda} = \rho$ and base parameter pair
$(\beta,\gamma) = (1,2.5)$
Estimated mean $E[W_y]$ and standard deviation $SD(W_y)$ as a function of $1 - \rho$ for
five cases of the $M_t/M/1$ queue at $y = 0.5$ : $\mu = 1, \overline{\lambda} = \rho$ and base parameter pair
$(\beta, \gamma) = (4, 2.5)$ having larger relative amplitude $\ldots \ldots \ldots$
Estimated mean $E[W_y]$ and standard deviation $SD(W_y)$ as a function of $1 - \rho$ for
five cases of the $(H_2)_t/M/1$ queue at $y = 0.5$ : $\mu = 1, \bar{\lambda} = \rho$ and base parameter pair
$(\beta, \gamma) = (1, 2.5)$
The (identical) solutions to the minimax and minimum-deviation optimization prob-
lems in $(5.7)$ and $(5.8)$ for the sinusoidal model in $(5.1)$ - $(5.4)$ with HT scaling in
(5.11) with parameters $(\rho, \beta_{\rho}, \gamma_{\rho}) = (\rho, 0.2, 2.5(1-\rho)^2)$ . The mean waiting times are
reported with and without space scaling
The solutions to the minimum-deviation optimization problem in $(5.8)$ for the si-
nusoidal model in Table 5.1 except $\beta$ has been increased to $\beta = 0.8$ from 0.2. The
reported average mean waiting times are reported with and without space scaling. $% \left( 120,120,120,120,120,120,120,120,120,120,$
The optimal $\eta$ and $\xi$ for $\rho = 0.8$ , $\beta = 0.5$ , cycle length $C = 60. \dots \dots \dots \dots \dots 132$
The optimal $\eta$ and $\xi$ for $\rho = 0.8$ , $\beta = 1$ , cycle length $C = 60. \dots \dots \dots \dots \dots 134$
A comparison of the 1-term and 2-term approximations for the asymptotic decay rate
$\theta_{ ho}^{*}$ from the asymptotic expansion in (A.3) with the exact values for the $M_t/H_2/1$
and $(H_2)_t/M/1$ models, where the $H_2$ distribution has $c^2 = 2.0$ and balanced means:
The scaled value $\theta_{\rho}^*/(1-\rho)$ is shown for 6 values of $1-\rho$
A comparison of the 1-term and 2-term approximations for the asymptotic decay rate
$\theta_{ ho}^{*}$ from the asymptotic expansion in (A.3) with the exact values for the $M_t/H_2/1$
and $(H_2)_t/M/1$ models, where the $H_2$ distribution has $c^2 = 2.0$ and balanced means:
The scaled value $\theta_{\rho}^*/(1-\rho)$ is shown for 6 values of $1-\rho$

A.3	Estimates of $\hat{p} \equiv P(W > b) \equiv Ae^{-\theta^* b}$ in the $M/M/1$ model with $\rho = 0.8, \bar{\lambda} = 1, \mu =$
	1.25 based on $n = 5000$ replications
A.4	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model for $y = 0.0$
	as a function of $\gamma$ and $b$ based on $n$ = 5,000 replications: $\rho$ = 0.8, $\bar{\lambda}$ = 1, $\mu$ =
	$1.25, \beta = 0.2 \ldots \ldots$
A.5	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model for $y =$
	$\pi/2\gamma$ as a function of $\gamma$ and $b$ based on $n = 5,000$ replications: $\rho = 0.8, \bar{\lambda} = 1, \mu =$
	$1.25, \beta = 0.2 \ldots \ldots$
A.6	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model for $y = \pi/\gamma$
	as a function of $\gamma$ and $b$ based on $n = 5,000$ replications: $\rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.8$
	0.2
A.7	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model for $y =$
	$3\pi/2\gamma$ as a function of $\gamma$ and $b$ based on $n = 5,000$ replications: $\rho = 0.8, \bar{\lambda} = 1, \mu =$
	$1.25, \beta = 0.2 \ldots \ldots$
A.8	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
	function of y based on $n = 5,000$ replications: $\gamma = 10, b = 20, \rho = 0.8, \overline{\lambda} = 1, \mu =$
	$1.25, \beta = 0.2 \dots \dots$
A.9	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
	function of y based on $n = 5,000$ replications: $\gamma = 0.1, b = 50, \rho = 0.8, \overline{\lambda} = 1, \mu = 0.1, b = 0.0, \overline{\lambda} = 1, \mu = 0.1, b = 0.0, \overline{\lambda} = 1, \mu = 0.1, b = 0.0, \overline{\lambda} = 0.0, \overline$
	$1.25, \beta = 0.2 \dots \dots$
A.10	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
	function of y based on $n = 5,000$ replications: $\gamma = 0.01, b = 300, \rho = 0.8, \overline{\lambda} = 1, \mu = 0.01, b = 0.00, \overline{\lambda} = 0.00, $
	$1.25, \beta = 0.2 \dots \dots$
A.11	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
	function of y based on $n = 5,000$ replications: $\gamma = 1, b = 20, \rho = 0.9, \overline{\lambda} = 1, \mu =$
	$1.11, \beta = 0.2 \dots \dots$
A.12	Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
	function of y based on $n = 5,000$ replications: $\gamma = 0.1, b = 50, \rho = 0.9, \overline{\lambda} = 1, \mu = 0.0, \overline{\lambda} = 1, \overline{\lambda} = 1, \mu = 0.0, \overline{\lambda} = 1, \overline$
	$1.11, \beta = 0.2 \dots \dots$

A.13 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y based on $n = 5,000$ replications: $\gamma = 10, b = 20, \rho = 0.8, \overline{\lambda} = 1, \mu =$
$1.25, \beta = 0.5 \dots \dots$
A.14 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y based on $n = 5,000$ replications: $\gamma = 1, b = 20, \rho = 0.8, \overline{\lambda} = 1, \mu =$
$1.25, \beta = 0.5 \dots \dots$
A.15 Summary of simulation results for a fixed $b$ and differing $y$ in a cycle: $\gamma=0.1,b=$
$100, \rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.5$
A.16 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y, for y in the small interval $[0.45, 0.55]$ based on $n = 5,000$ replications:
$\gamma = 0.1, b = 20, \rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.2$
A.17 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.1, b = 20, \rho = 0.8, \overline{\lambda} = 0.8, \mu =$
$1,\beta=0.2  \ldots  \ldots  \ldots  \ldots  \ldots  \ldots  \ldots  \ldots  \ldots  $
A.18 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y based on 5,000 replications: $\gamma = 1, b = 20, \rho = 0.8, \overline{\lambda} = 0.8, \mu = 1, \beta = 0.2179$
A.19 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.25, b = 40, \rho = 0.9, \overline{\lambda} = 0.9, \mu =$
$1, \beta = 0.1$
A.20 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y based on 5,000 replications: $\gamma = \frac{1}{16}, b = 80, \rho = 0.95, \overline{\lambda} = 0.95, \mu =$
$1,\beta=0.05\ldots181$
A.21 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.64, b = 25, \rho = 0.84, \bar{\lambda} = 0.84, \mu = 0.84, \bar{\lambda} = 0.84, \mu = 0.84, \bar{\lambda} = 0.84, \lambda$
$1,\beta=0.16\ldots182$
A.22 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.16, b = 50, \rho = 0.92, \bar{\lambda} = 0.92, \mu = 0.16, \bar{\lambda} = 0.92, $
$1,\beta=0.08\ldots$

A.23 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.04, b = 100, \rho = 0.96, \bar{\lambda} = 0.96, \mu =$
$1,\beta=0.04\ldots184$
A.24 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.01, b = 200, \rho = 0.98, \bar{\lambda} = 0.98, \mu = 0.98, \bar{\lambda} = 0.98,$
$1,\beta=0.02\ldots185$
A.25 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/M/1$ model as a func-
tion of y based on 5,000 replications: $\gamma = 0.0025, b = 400, \rho = 0.99, \overline{\lambda} = 0.99, \mu =$
$1,\beta=0.01\ldots186$
A.26 Comparison of ratio $P(W_y > b)/P(W > b) = A_y/\rho$ for different $\rho$ 's with base
parameter $(\beta, \gamma, b) = (1, 25, 4)$ using the scaling in (3.39) of the main thesis 187
A.27 Summary of simulation results for $M_t/M/1$ queue at $y = 0$ as a function of $1 - \rho$
with base parameter $(\beta, \gamma, b) = (1, 25, 4)$ using the scaling in (3.39) of the main thesis:188
A.28 Summary of simulation results for $M_t/M/1$ queue at $y = 0$ and at $y = 0.5$ as a
function of $1 - \rho$ with base parameter $(\beta, \gamma, b) = (1, 2.5, 4)$ using the scaling in (3.39)
of the main thesis $\ldots \ldots \ldots$
A.29 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $(H_2)_t / M / 1$ model as a
function of y based on 5,000 replications: $\gamma = 1, b = 20, \rho = 0.8, \overline{\lambda} = 0.8, \mu = 1, \beta = 0.8$
$0.2, \theta^* = 0.173 \dots \dots$
A.30 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $(H_2)_t / M / 1$ model as a
function of y based on 5,000 replications: $\gamma = 0.25, b = 40, \rho = 0.9, \overline{\lambda} = 0.9, \mu =$
$1,\beta=0.1,\theta^*=0.0761\ldots192$
A.31 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $(H_2)_t / M / 1$ model as a
function of y based on 5,000 replications: $\gamma = \frac{1}{16}, b = 80, \rho = 0.95, \overline{\lambda} = 0.95, \mu =$
$1, \beta = 0.05, \theta^* = 0.0356$ 193
A.32 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $(H_2)_t / M / 1$ model as a
function of y based on 5,000 replications: $\gamma = 0.64, b = 25, \rho = 0.84, \bar{\lambda} = 0.84, \mu = 0.84, \bar{\lambda} = 0.84, \mu = 0.84, \bar{\lambda} = 0.84, \lambda$
$1,\beta=0.16,\theta^*=0.131\ldots194$

A.33 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $(H_2)_t / M / 1$ model as a
function of y based on 5,000 replications: $\gamma = 0.16, b = 50, \rho = 0.92, \bar{\lambda} = 0.92, \mu = 0.000$
$1,\beta=0.08,\theta^*=0.0593\ldots\ldots\ldots195$
A.34 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $(H_2)_t / M / 1$ model as a
function of y based on 5,000 replications: $\gamma = 0.01, b = 200, \rho = 0.98, \bar{\lambda} = 0.98, \mu = 0.01, \lambda = 0.00, \lambda = 0.00,$
$1, \beta = 0.02, \theta^* = 0.0137196$
A.35 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $(H_2)_t / M / 1$ model as
a function of y based on 5,000 replications: $\gamma = 0.0025, b = 400, \rho = 0.99, \bar{\lambda} =$
$0.99, \mu = 1, \beta = 0.01, \theta^* = 0.00676$ 197
A.36 Comparison of ratio $P(W_y > b)/\rho$ in the $(H_2)_t/M/1$ queue as a function of $\rho$ with
base parameter $(\beta, \gamma, b) = (1, 25, 4)$ using the scaling in (3.39) of the main thesis 198
A.37 Summary of simulation results for the $(H_2)_t/M/1$ queue at $y = 0$ as a function of
$1-\rho$ with base parameter $(\beta,\gamma,b)=(1,25,4)$ using the scaling in (3.39) of the main
thesis. $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $\ldots$ $199$
A.38 Summary of simulation results for the $(H_2)_t/M/1$ queue at $y = 0$ and $y = 0.5$ as a
function of $1 - \rho$ with base parameter $(\beta, \gamma, b) = (1, 2.5, 4)$ using the scaling in (3.39)
of the main thesis
A.39 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/H_2/1$ model as a
function of y based on 5,000 replications: $\gamma = 1, b = 20, \rho = 0.8, \overline{\lambda} = 0.8, \mu = 1, \beta = 0.8$
$0.2, \theta^* = 0.124202$
A.40 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/H_2/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.25, b = 40, \rho = 0.9, \overline{\lambda} = 0.9, \mu =$
$1,\beta=0.1,\theta^*=0.0644\ \ldots\ \ldots\$
A.41 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/H_2/1$ model as a
function of y based on 5,000 replications: $\gamma = \frac{1}{16}, b = 80, \rho = 0.95, \overline{\lambda} = 0.95, \mu = 0.95, \overline{\lambda} = 0.95, \mu =$
$1, \beta = 0.05, \theta^* = 0.0328204$
A.42 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/H_2/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.64, b = 25, \rho = 0.84, \bar{\lambda} = 0.84, \mu = 0.84, \bar{\lambda} = 0.84, \mu = 0.84, \bar{\lambda} = 0.84, \lambda$
$1, \beta = 0.16, \theta^* = 0.101  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  \dots  $

A.43 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/H_2/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.16, b = 50, \rho = 0.92, \bar{\lambda} = 0.92, \mu =$
$1,\beta=0.08,\theta^*=0.0519\ldots$
A.44 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/H_2/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.04, b = 100, \rho = 0.96, \bar{\lambda} = 0.96, \mu =$
$1, \beta = 0.04, \theta^* = 0.0263207$
A.45 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/H_2/1$ model as a
function of y based on 5,000 replications: $\gamma = 0.01, b = 200, \rho = 0.98, \bar{\lambda} = 0.98, \mu =$
$1, \beta = 0.02, \theta^* = 0.0132 \dots \dots$
A.46 Simulation estimates of $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ in the $M_t/H_2/1$ model as a func-
tion of y based on 5,000 replications: $\gamma = 0.0025, b = 400, \rho = 0.99, \overline{\lambda} = 0.99, \mu =$
$1,\beta = 0.01, \theta^* = 0.00664 \dots 209$
A.47 Comparison of ratio $P(W_y > b)/\rho$ as a function of $\rho$ in $M_t/H_2/1$ queue with base
parameter $(\beta, \gamma, b) = (1, 25, 4)$ using the scaling in (3.39) of the main thesis 210
A.48 Summary of simulation results for $M_t/H_2/1$ queue at $y = 0$ as a function of $1 - \rho$
with base parameter $(\beta, \gamma, b) = (1, 25, 4)$ using the scaling in (3.39) of the main thesis 211
A.49 Summary of simulation results for $M_t/H_2/1$ queue at $y = 0$ and $y = 0.5$ as a function
of $1 - \rho$ with base parameter $(\beta, \gamma, b) = (1, 25, 4)$ using the scaling in (3.39) of the
main thesis $\ldots \ldots 212$
A.50 Estimated mean and standard deviation of the steady-state waiting time in $M/M/1$
queue as a function of $1 - \rho$ : $\mu = 1, \overline{\lambda} = \rho$
A.51 Estimated mean $E[W_y]$ and standard deviation $SD[W_y]$ as a function of $1-\rho$ for five
cases of the $M_t/M/1$ queue at $y = 0.0$ and $y = 0.5$ : $\mu = 1, \bar{\lambda} = \rho$ , base parameter
pair $(\beta, \gamma) = (1, 2.5)$ using the scaling in (3.39). $n = 40,000$ and $\delta = 0.001$ for all $\rho$ 's. 214
A.52 Estimated $E[W_y]$ and $SD[W_y]$ as a function of $1 - \rho$ for $M_t/M/1$ queue at $y = 0.0$
and 0.5: $\mu = 1, \bar{\lambda} = \rho$ , base parameter pair $(\beta, \gamma) = (4, 2.5)$ (with longer cycles than
in Table A.51) using the scaling in (3.39). $n=40,000$ and $\delta=0.001$ for all $\rho$ 's 215
A.53 Estimated $E[W_y]$ and $SD[W_y]$ as a function of $1 - \rho$ for $(H_2)_t/M/1$ at $y = 0.0$ and
0.5: $\mu = 1, \bar{\lambda} = \rho$ , base $(\beta, \gamma) = (1, 2.5)$ with scaling in (3.39). $n = 40,000$ for all $\rho$ 's. 216

- A.54 Estimated  $E[W_y]$  and  $SD[W_y]$  as a function of  $1 \rho$  for  $M_t/H_2/1$  at y = 0.0 and 0.5:  $\mu = 1, \bar{\lambda} = \rho$ , base  $(\beta, \gamma) = (1, 2.5)$  with scaling in (3.39). n = 40,000 for all  $\rho$ 's. 217 A.55 Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $(H_2)_t/M/1$  model without

## Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor Professor Ward Whitt for his continuous support through the five years of my PhD study and related research. Professor Whitt is a great researcher and an amazing mentor. I have always been impressed by his extensive knowledge, brilliant thoughts and deep passion for research. I appreciate his generous mentoring on both mathematical/engineering knowledge and oral/written presentation skills, his tremendous patience shown to me during the guidance, the lots of freedom he gave me to explore different research areas at my interest and his consistent caring for my study, career and life. It has been my greatest honor and privilege working closely with him and I would not have accomplished this without his kind support and guidance all the time.

I would like to thank the rest of my committee, Professor Karl Sigman and Professor Linda Green, Professor Ton Dieker and Professor Henry Lam, for their generous offering to serve on my dissertation committee and for their helpful and insightful suggestions. They are all great researchers in the area of this thesis and it is my honor to have them on my committee.

I am very grateful for all the amazing professors in both IEOR and DRO who have taught me lectures and seminars, and helped me with my research and career. Chapter Three of this thesis is motivated by the two seminar courses I took given by Professor Ton Dieker and Professor Jose Blanchet. They both devoted lots of time mentoring each of the student on his/her own research project. Professor Karl Sigman, Professor Garud Iyengar, Professor David Yao, Professor Paul Glasserman, Professor Tim Leung and Professor Martin Haugh all generously offered me invaluable help and suggestions in my application for this PhD program when I was a MSFE student here. I have attended excellent PhD classes taught by Professors Ward Whitt, David Yao, Donald Goldfarb, Cliff Stein, Ciamac Moallemi, Assaf Zeevi, Garud Iyengar, Karl Sigman, Ton Dieker, Jose Blanchet, Philip Protter and Shipra Agrawal. Their lectures have set a strong foundation in Operations Research for me and I owe them my heartfelt appreciation. My sincere thanks go to many senior PhD fellows who helped me along the journey: Professor Jing Dong, Professor Xinyun Chen, Professor Song-Hee Kim, Professor Yunan Liu, Juan Li, Xin Li, Chen Chen, Xingbo Xu, Chun Wang, Jing Guo, Brian Lu, Linan Yang, Bo Huang, Cun Mu, Zheng Wang, Wenqi Hu, Anran Li, Di Xiao and Liao Wang. I enjoyed taking classes and preparing qualification exams very much with PhD fellows in my year: Zhipeng Liu, Xinshang Wang, Yanan Pei, Fei He, Professor Mingxian Zhong, Professor Irene Lo, Franscois Fagan, Hal James Cooper, Jalaj Bhandari, Suraj Kumar Keshri and Octavio Ruiz Lacedelli. I shall never forget my wonderful office mates in Mudd 345 and all my other PhD friends: Wei You, Xu Sun, Xiaopei Zhang, Yixin Wang, Yang Kang, Professor Phillis Wan, Shuoguang Yang, Yunjie Sun, Yeqing Zhou, Fengpei Li, Fan Zhang, Jingtong Zhao, Yunhao Tang and many more.

I thank all the IEOR staff members for their support over the years and for keeping the IEOR community warm and pleasant.

Last but not the least, I am deeply indebted towards my parents, Xiaohong Guo and Zhouguang Ma for their unconditional love and support. They have been my constant source of courage, strength, persistence, confidence and inspiration. I would like to give my special thank to my husband for his trust, encouragement and caring. Without my family's love and support throughout these years, I could not have overcome the hard times. I would like to dedicate this work to them.

To My Family: Xiaohong Guo, Zhouguang Ma, Zhipeng Liu

## Chapter 1

## Introduction

Waiting lines or queues often appear wherever resources are needed to provide services. In response, queueing theory has been developed to improve operational efficiency. Queueing theory has many applications. For example, transportation systems like railway stations and traffic systems that optimize vehicle flow; service systems like banks and hospitals that serve customers; computing systems like computers and more complex cloud platforms that process computing jobs; market dynamics in finance where order arrivals and executions are modeled to predict trading volumes and prices.

In these applications, the arrival rates are mostly not constant, but varying over time. For example, Figure 1.1 taken from Koopman (1972) plots the average hourly number of airplane arrivals to J.F.K (Terminal A in the plot) and LaGuardia (Terminal B in the plot) airports, which displays obvious daily patterns. Data were gathered in 1968 during one month's operations. While the stationary queueing models studied in stochastic textbooks provide fundamental methodology to analyze performance of general queues, the time-varying (TV) arrival rates in real applications call for models to directly consider this TV property of the systems.

This thesis is a contribution to queueing theory as in the referenced books, such as Asmussen (2003), and focuses on TV single-server queues. Figure 1.2 illustrates a queueing system with a service facility consisting of a single server and a waiting room. In the multi-server queue case, there is more than one server in the service facility. Customers arrive at the system with an arrival rate function denoted as  $\lambda(t)$ , wait in the queue if the server is busy, get served in the service facility when previous customers are all gone and leave the system when the service is done. In the TV



**Figure 1.1:** Average hourly number of airplane arrivals to J.F.K (Terminal A in the plot) and LaGuardia (Terminal B in the plot) airports.

setting, the function  $\lambda(t)$  is not a constant. The principal system performance measures are the virtual waiting time W(t) to capture the waiting time for potential arrivals at each time and the queue length process Q(t). Single-server queues are basic components of more complex queueing networks and exhibit longer waiting times compared to multi-server queues. They are important to model queueing systems with limited service facilities like airplanes landing at an airport with a single runway (Koopman (1972)), trucks bringing cranberries to one cranberry-processing plant (Porteus (1989, 1993a,b)) and computing jobs sent to a single server (Chang (1970)).



Figure 1.2: Elements of a single-server queueing system with a waiting room and a service facility.

There is also a wide range of different techniques to study the dynamics of TV queues, which evolve as different subjects develop more advanced approaches: ordinary differential equations (ODE) approach based on continuous-time Markov chains (CTMC) for TV Markovian models to numerically solve the performance measures, as in Koopman (1972), Kolesar et al. (1975), Rothkopf and Oren (1979), Taaffe and Ong (1987), Ong and Taaffe (1989); steady-state analysis of the stochastic waiting time and queue length processes, as in Harrison and Lemoine (1977), Heyman and Whitt (1984), Lemoine (1981, 1989), Rolski (1981, 1989a,b); heavy traffic (HT) asymptotic models and their approximation models, as in Newell (1968a,b,c), Massey (1985), Mandelbaum and Massey (1995) and more recently in Whitt (2014, 2016a); Monte Carlo simulation to generate non-stationary processes, which applies to simulating TV queues to estimate distributions of performance measures, as in Lewis and Shedler (1979), Gerhardt and Nelson (2009) for thinning approach to non-homogenous Poisson processes, in Cinlar (1975), Chen and Schmeiser (1992), Nicol and Leemis (2014), Chen and Schmeiser (2015, 2017) for inversion approach and in Liu et al. (2018) for non-stationary non-Poisson processes; robust optimization as a relatively new approach to approximate the TV stochastic model by a more tractable optimization problem, as in Bertsimas et al. (2011), Ben-Tal et al. (2009), Beyer and Sendhoff (2007), Whitt and You (2018). See section 1.2.1 for a more extensive literature review on alternative methods to study TV queues.

In addition to the classic steady-state analysis, HT asymptotic analysis and standard Monte Carlo simulation, we make use of another important methodology in queueing theory, rare-event simulation, to study TV queues in this thesis. We extend the classic rare-event simulation algorithm for stationary queues to the TV setting, which not only estimates very small tail probabilities at different times, but also calculates distributions and moments of TV performance measures and HT limit distributions of TV queues. We show that rare-event simulation algorithms are actually very efficient in generating unbiased estimates of performance measures in complex TV queues.

This thesis makes contributions to both of the two areas:

- (i) rare-event simulation in queues and
- (ii) time-varying single-server queues.

We will motivate each of the topics separately below.

#### **1.1** Rare-Event Simulation in Queues

Rare-event simulation algorithms in queues aim to efficiently estimate the probability of certain rare but important events in queueing models. For example, we may want to estimate the probability that patient waiting times are extremely long in the Emergency Room to check the robustness of its operations. An exceptionally long waiting time is highly undesirable for both the patients and the medical staff and should be avoided. For another example, the manager of a finite waiting room service system may be interested in the probability that the waiting queue is longer than the buffer can hold to decide its buffer length. This probability would be the customer-loss probability for the system. In both of these situations, the goal is to assure that these probabilities are very small. Since general numerical techniques to calculate these probabilities for queueing models often are unavailable, simulation methods are an attractive effective alternative.

It is well known that standard simulation may not be feasible to estimate these small probabilities; see Heidelberger (1995). To illustrate the problem with standard simulation, consider a random variable X with probability density function (pdf) f(x) and assume we want to estimate the probability p that X is in the rare-event set A. With standard simulation, we generate n i.i.d. samples  $X_1, X_2, ..., X_n$  and form an unbiased estimator for p:

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \in A\}},\tag{1.1}$$

where  $I_A$  is the indicator function of the set A; i.e.,  $I_A$  is the random variable that assumes the value 1 on A and the value 0 on its complement. The expectation and variance of  $\hat{p}_n$  are as below:

$$E[\hat{p}_n] = p, \ Var[\hat{p}_n] = \frac{p(1-p)}{n}.$$
 (1.2)

The relative error of an estimator  $\hat{p}$  is calculated as:

$$re(\hat{p}) = \sqrt{Var(\hat{p})}/\hat{p},\tag{1.3}$$

therefore the relative error of  $\hat{p}_n$  follows:

$$re(\hat{p}_n) = \sqrt{\frac{p(1-p)}{n}} / \hat{p}_n \approx \sqrt{\frac{1}{pn}}.$$
(1.4)

If we want to achieve an accuracy of 0.01 relative error, then we can calculate the sample size needed:  $n = 1/(0.01^2 p)$ , a sample size inversely proportional to the level of the rare-event probability p. For example, if  $p = 10^{-k}$ , then we require a sample size as large as  $10^{k+4}$ , which becomes infeasible as k grows and the rare-event probability approaches 0. In contrast, rare-event simulation can develop methods that don't require an unbounded sample size as k grows to infinity.

An effective technique to address this estimation problem is importance sampling, which has been discussed, for example, in Hammersley and Handscomb (1964), Asmussen and Glynn (2007). It is a variance reduction technique that can be applied to construct an unbiased estimator that does not yield unbounded relative error as the rare event gets rarer. We now explain in the context of the previous simple example. In that setting we can write p as  $p = E_f[I_{\{X \in A\}}]$ , where X has pdf f(x). Consider another pdf function g(x) and p can also be expressed as the expectation of a function of X with pdf g(x), which is achieved by dividing and multiplying g(x) within the integrand:

$$p = E_f[I_{\{X \in A\}}] = \int I_{\{x \in A\}} f(x) dx$$

$$= \int I_{\{x \in A\}} \frac{f(x)}{g(x)} g(x) dx$$

$$= E_g[I_{\{X \in A\}} L(X)].$$
(1.5)

The ratio of the two density functions is called the likelihood ratio:

$$L(x) = \frac{f(x)}{g(x)}.$$
(1.6)

The importance sampling density function g(x) must satisfy the condition that g(x) > 0 for all x where f(x) > 0 to make the above transformation valid. The importance sampling estimation scheme is to generate n i.i.d. samples from pdf g(x) and then to calculate the estimator

$$\hat{p}_n(g) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \in A\}} L(X_i).$$
(1.7)

Equation (1.5) shows that the new estimator is also unbiased and the choice of the importance sampling distribution should aim to minimize the variance of the estimator. The second moment of the estimator  $I_{\{X \in A\}}L(X)$  equals:

$$E_g[(I_{\{X \in A\}}L(X))^2] = \int I_{\{x \in A\}}(\frac{f(x)}{g(x)})^2 g(x) dx$$
  
=  $\int I_{\{x \in A\}} \frac{f(x)}{g(x)} f(x) dx$   
=  $E_f[I_{\{X \in A\}}L(X)].$ 

Therefore we want to make f(x)/g(x) small on the rare-event set A or g(x) large on A, which means that under the importance sampling distribution, the event A happens much more frequently. For example, consider p(b) = P(W > b), where W is the steady-state waiting time. Suppose  $p(b) \sim c \cdot h(b)$  as  $b \to \infty$  ( $h(b) \to 0$ ). If the importance sampling distribution satisfies that  $E_g[(I_{\{X \in A\}}L(X))^2] \sim d \cdot h(b)$ , then we have bounded relative error as b goes to infinity:  $re(\hat{p}(b)) = \sqrt{(d-c^2)/N} < \infty$  as  $b \to \infty$ .

Importance sampling can be applied to more general cases than to a single random variable with a pdf function. The trick in the transformation equation (1.5) applies to other cases as well. Also the rare event A can be defined more generally, for example, in terms of stopping times of stochastic processes like discrete-time or continuous-time Markov chains, semi-Markov processes or more general stochastic processes; see for example Whitt (1980). Let  $\tau$  be the stopping time for a sequence of random variables  $\{X_n, n \ge 0\}$ , P be the original probability measure and P' be the importance sampling probability measure, H be an indicator function of X's. Then we have

$$E_P[H(X_0, ..., X_\tau)] = E_{P'}[H(X_0, ..., X_\tau)] \times L(X_0, ..., X_\tau),$$
(1.8)

where the likelihood ratio has the following equation if all X's are independent:

$$L(X_0, ..., X_{\tau}) = \prod_{i=1}^{\tau} \frac{p(X_i)}{p'(X_i)}.$$
(1.9)

There have been studies on importance sampling applied to estimating the rare-event probability of very large waiting times or queue lengths in single-server queues (which we review in the following paragraph), multi-server queues (e.g. Sadowsky (1991)) and queueing networks(e.g. Devetsikiotis and Townsend (1992a, 1993, 1992b)). The importance sampling results for single-server queues are closely related to the large deviations approach; see for example Bucklew (1990). This thesis focuses on waiting times in single-server queues using the exponential tilting method. In contrast to previous papers that concentrate on stationary queues, this thesis studies the rare-event simulation algorithms in the TV setting.

In order to apply rare-event simulation for the TV  $GI_t/GI_t/1$  queue, we apply previous rareevent simulation methods for the stationary GI/GI/1 queue. Hence we next review the standard approach there. The importance sampling algorithm in stationary single-server GI/GI/1 queues is based on the random walk expression for the waiting time sequence following the Lindley's recursion. It exploits the equivalence between the waiting times and the maximum of a random walk. Let  $V_n$  be service times,  $\rho^{-1}U_n$  be inter-arrival times. The waiting time distribution for the  $n^{th}$  customer is

$$W_n = \max_{0 \le k \le n} (\tilde{S}_k) \stackrel{d}{=} \max_{0 \le k \le n} (S_k), \tag{1.10}$$

where  $\tilde{S}_k = X_n + ... + X_{n-k+1}$  and  $S_k = X_1 + ... + X_k$ . The steady-state waiting time distribution follows:

$$W = \max_{k \ge 0} S_k,\tag{1.11}$$

which is the maximum of a negative-drifted random walk as  $\rho < 1$ . We note that the rare-event probability can be expressed as

$$P(W > b) = P(\max_{k \ge 0} S_k > b) = P(\tau_b^S < \infty)$$
(1.12)

for a large b, where  $\tau_b^S$  is the first time the random walk S hits b, which is called the hitting time of the random walk S at level b. The problem with standard simulation to estimate  $P(\tau_b^S < \infty)$  in this case is that (1) the sample size n would blow up and (2) it takes a long time to determine that the rare event will not happen for one experiment.

The exponential tilting approach can transform the random walk into one with a positive drift as developed in Asmussen (1985, 2003). Let F be the distribution function of X and assume existence of its moment generating function  $M(\theta) = E(e^{\theta X})$ . Define the tilted distribution function  $dF_{\theta}(x) = [e^{\theta x}/M(\theta)]dF(x)$  and the likelihood ratio for  $X_1, ..., X_n$  equals

$$L_n(\theta) = \frac{M(\theta)^n}{e^{\theta(X_1 + \dots + X_n)}} = M(\theta)^n e^{-\theta S_n}.$$
(1.13)

Let  $\theta^*$  be the asymptotically optimal  $\theta$ , which exists such that  $M(\theta^*) = 1$ . Because M(0) = 1, M'(0) < 0,  $M(\theta)$  is convex and continuous, at  $\theta^*$ ,  $M'(\theta^*) = E_{\theta^*}[X] > 0$ , so that the random walk becomes positively drifted and  $\tau < \infty$  with probability 1. In the exponential tilting simulation, we generate X from  $F_{\theta^*}$  and estimate the probability by

$$P(W > b) = P(\tau_b^S < \infty) = E_{\theta^*}[\exp(-\theta^* S_{\tau_b^S})] = e^{-\theta^* b} E_{\theta^*}(\exp(-\theta^* (S_{\tau_b^S} - b))).$$
(1.14)

More insights on the asymptotically optimal parameter  $\theta^*$  are discussed in Asmussen (1982), Anantharam (1988). In the TV setting, there are two problems with the basis for the rare-event algorithm. The first problem is how to define the steady-state waiting time in a TV queue, and since the reverse-time random walk and the forward-time random walk in 1.10 no longer have the same distribution due to the TV arrival rates, the second problem is that the equality in distribution no longer holds. As for the exponential tilting algorithm itself, one problem is to find the asymptotically optimal  $\theta^*$  and prove that the relative error is bounded under  $\theta^*$ . Another problem is to formulate the estimator involving the likelihood ratio. This thesis solves these problems to extend the algorithm to both  $GI_t/GI/1$  and  $GI_t/GI_t/1$  queues. Most importantly, we make use of this algorithm to study the dynamics of the queueing models and compute important distributions for the queueing system.

To illustrate the efficiency of our developed rare-event simulation algorithm in TV queues, we present simulation results for estimating  $P(W_0 > b)$  in the  $M_t/M/1$  model with an arrival rate function  $\lambda(t) = 1 + 0.2 \times \sin(\gamma t)$  and a fixed service rate  $\mu = 1.25$  in Table 1.1 below. We can see that with the same number of experiments, the relative error is approximately independent of bfor each  $\gamma$ , ranging from about 0.0029 for  $\gamma = 10$  to about 0.0055 for  $\gamma = 0.1$ , so the algorithm is efficient as b gets larger and the estimated probability gets smaller (from the level of 0.01 to the level of  $10^{-9}$ ). See Chapter 3 for details and for more simulation results.

**Table 1.1:** Estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model with sinusoidal arrivalrate function  $\lambda(t) = 1 + 0.2 \times \sin(\gamma t)$  as a function of  $\gamma$  and b for:  $\rho = 0.8, \mu = 1.25, y = 0$  based on 5000 replications.

	b	$\hat{p}$	$exp(-\theta^*b)$	$A_0(b)$	s.e.	95% CI (lb)	(ub)	r.e.
$\gamma = 10$	10	0.0654	0.0821	0.797	1.87E-04	0.0651	0.0658	0.00286
	20	0.00537	0.00674	0.797	1.55E-05	0.00534	0.00540	0.00289
	40	3.61E-05	4.54E-05	0.795	1.05E-07	3.59E-05	3.63E-05	0.00290
	80	1.64E-09	2.06E-09	0.796	4.82E-12	1.63E-09	1.65E-09	0.00294
$\gamma = 1$	10	0.0628	0.0821	0.765	1.87E-04	0.0624	0.0632	0.00298
	20	0.00516	0.00674	0.766	1.51E-05	0.00513	0.00519	0.00292
	40	3.49E-05	4.54E-05	0.769	1.00E-07	3.47E-05	3.51E-05	0.00287
	80	1.58E-09	2.06E-09	0.767	4.65E-12	1.57 E-09	1.59E-09	0.00294
$\gamma = 0.1$	10	0.0413	0.0821	0.503	2.33E-04	0.0409	0.0418	0.00565
	20	0.00360	0.00674	0.535	1.98E-05	0.00356	0.00364	0.00550
	40	2.50E-05	4.54E-05	0.551	1.37E-07	2.47E-05	2.53E-05	0.00548
	80	1.12E-09	2.06E-09	0.545	6.20E-12	1.11E-09	1.14E-09	0.00552

#### 1.1.1 Literature Review

Early studies on importance sampling, the basis for rare-event simulation include Hammersley and Handscomb (1964), Glynn and Iglehart (1989). For GI/GI/1 single-server queues, Asmussen (1985, 2003), Siegmund (1976), Lehtonen and Nyrhinen (1992) develop the exponential tilting approach and show that  $\theta^*$  is the unique asymptotically optimal value within some class of distributions. Further insight about the  $\theta^*$  can be found in Asmussen (1982), Anantharam (1988). The large deviations theory as the basis for efficient rare-event simulation has been studied in Cottrell et al. (1983), Donsker and Varadhan (1975a,b), Sadowsky and Bucklew (1990), Bucklew (1990), Siegmund (1976), Glasserman and Kou (1995), Glasserman and Wang (1997). Our thesis develops rare-event simulation algorithms for single-server queues but in the TV setting.

There is also work for the multi-server queues including the early paper Sadowsky (1991) and more recent papers including Blanchet and Lam (2014). On rare event simulation in queueing networks, Parekh and Walrand (1989) develops powerful heuristics for simulation overflows in Jackson networks and other works include Tsoucas (1989), Frater and Anderson (1989), Glasserman and Kou (1993, 1995), Frater et al. (1991), Anantharam et al. (1990), Weiss (1986).

The above literature mostly focuses on light-tailed input, while there have been efforts to develop rare-event simulation techniques for systems with heavy-tailed input including the initial work Asmussen et al. (2000) and more developments Dupuis and Wang (2004), Dupuis et al. (2006), Rojas-Nandayapa and Asmussen (2007), Blanchet and Glynn (2008). The development of state-dependent importance sampling estimators for heavy-tailed systems has been discussed in Bassamboo et al. (2006), Blanchet and Glynn (2008), Blanchet and Liu (2008), Blanchet et al. (2007). This thesis assumes light-tailed input and aims to understand the performance measures of a single-server TV queue using the rare-event simulation.

#### 1.2 Time-Varying Single-Server Queues

There are essentially three approaches to apply the stationary models in stochastic textbooks to the TV systems; see Whitt (2018) and references therein. First we can use the pointwisestationary approximation (PSA) when the arrival rate changes slowly compared to the service times. PSA assumes that performance of the queue at time t can be approximated by performance of a stationary queue with parameters taking effect at time t, which is a reasonable assumption when the arrival rate changes slowly enough. Second if PSA is not valid, we can use an appropriate worst case arrival rate to achieve high Quality-of-Service for customers at all times. High Qualityof-Service for customers can be quantified, for example, as low waiting time tail probabilities or low expected delay. Lastly we may use the long-run average arrival rate when we do not require high Quality-of-Service at all times. In this case, we ignore the local fluctuations and only consider the long-run average input. However, new TV models are needed to achieve high Quality-of-Service for more common cases.

The non-homogeneous Poisson process (NHPP) would be a natural model for the arrival process with a TV arrival rate, which is a reasonable approximation when independent decisions are made by many customers, who come to the queueing system with small probabilities. The supporting Poisson superposition theorem is discussed in stochastic books, for example, section 9.8 of Whitt (2002) and section 11.2 of Daley and Vere-Jones (2008). However there are applications where the arrival process is more variable (over-dispersion) or less variable (under-dispersion) than the Poisson process. For example, there may be forced separation between successive arrivals of the airplanes to the airport, which leads to under-dispersion in the arrival process. For another example, the queries about the latest stock prices to a computer server may display over-dispersion than a Poisson process, because there is uncertainty about the arrival rate stemming from the market news. Queries are likely to occur in clusters when there is big financial news coming out. In this thesis, we go beyond Markovian models to capture the non-Poisson input arising in real applications.

To understand the impact of TV arrival rates on performance measures, consider an example of a deterministic fluid model for a TV single-server queue, so that we can ignore the stochasticity and concentrate fully on the TV behavior of the queue. We consider a constant service rate  $\mu$ . As shown in 1.3, the arrival rate  $\lambda(t)$  increases from 0 at  $t_0$  to the level of the service rate  $\mu$  at  $t_1$ , keeps increasing to its peak value at  $t_2$ , drops to the service rate level at  $t_3$ , continues decreasing to 0 and at  $t_4$ , the queue first empties; see Example 1.1 in Whitt (2018). From  $t_0$  to  $t_1$ , the service rate is larger than the arrival rate, so no workload accumulates in the system. Starting from  $t_1$ , the arrival rate exceeds service rate and the queue starts to accumulate. At  $t_2$ , the arrival rate drops to be equal to  $\mu$  and the workload in queue reaches its maximum, which decreases as the arrival rate becomes larger than  $\mu$ . The queue finally empties at  $t_4$  when the extra service capacity finishes processing all workload in queue.

To derive the waiting time at time t, we need to look back prior to t, and since the extra service capacity is wasted, the furthest time in the past we need to consider is the one that maximizes the difference between the cumulative arrival rate and the cumulative service rate. Both integrals are taken by going reverse-time from t. For stationary models, the arrival and service rates are both constant over time, so going forward-time and reverse-time yield the same results (in the stochastic case, both give the same distributions), but for non-stationary models, we have to stick to the *reverse-time* direction. Another insight from this example is the *time lag* phenomenon between the time some arrival rate value takes effect and the time it impacts congestion. As shown in 1.3, the peak arrival rate occurs at  $t_2$ , while the maximum queue happens later at  $t_3$ ; the arrival rate drops to  $\mu$  at  $t_3$ , and the queue empties later at  $t_4$ . Changes in the arrival rate reveal their effects on the queue later in time because of the possible remaining workload in front of the new arrivals.



Figure 1.3: The arrival rate function  $\lambda(t)$  in an example of a deterministic TV single-server queue with a constant service rate  $\mu$ .

In TV queues, we no longer have stationary waiting times for customers and it makes more sense to consider the virtual waiting time or workload process as a function of time to capture the waiting time for potential arrivals at each time. The steady-state virtual waiting times are TV as well. The intuition of the reverse-time construction of the queue congestion forms the basis of a convenient representation of workload process in TV queues. The original representation was proposed by Lemoine (1981) for periodic  $M_t/G/1$  queues with a Markovian arrival process. We extend the representation to  $G_t/G/1$  queue in Chapter 3 to provide basis for rare-event algorithms
in  $GI_t/GI/1$  queues, as in equation 3.1. In Chapter 5, we further extend the representation to  $G_t/G_t/1$  queues based on which rare-event algorithms are also developed for  $GI_t/GI_t/1$  queues (virtual waiting time and workload processes become different in this model), as in equation 5.23. We go beyond exponential distributions and Markovian models in all the chapters of this thesis, and we can see, for example in the importance sampling estimator in Chapter 3 and Chapter 5, that some additional terms are needed to capture the characteristics of a general non-Markovian model; see equations 3.25, 3.28, 5.24 and Appendix A.3.6.

Customers will experience fluctuations in queue congestion due to the TV arrival rates, and service rate controls can be used to stabilize performance in service systems where there is no flexibility in the number of servers; see Whitt (2015). For the airplane landing example, the number of runways is fixed, but it may be possible to change the rate of airplane landings by controlling the required separation distance between airplanes. For the TSA inspection example, the number of security lines is fixed, but it may be possible to increase the inspection rate by relaxing inspection requirement. Note that we study an idealized case of what happens in these service operations: we consider a single server whose service rate is fully subject to control. Our study can help understand what are the desirable service-rate controls and what are the potential benefits of controlling service-rates.

We present simulation results in 1.4 for a TV single-server queue with a specific periodic arrival rate function  $\lambda(t) = 0.8 \times (1+0.2 \sin(0.1t))$ , exponential inter-arrival times and service requirements. The cycle length of the arrival rate function is 62.8. Figure 1.4 plots the expected steady-state virtual waiting time under different service rate functions for one cycle. Due to the time-lag property of the TV system, with a constant service rate, the virtual waiting time process peaks after the arrival rate peak as shown as the blue line. The rate-matching service rate is directly proportional to the arrival rate as defined in equation 4.1 and thus fails to adjust for the time lag. The red line plots the virtual waiting time under the rate-matching control, which is not stabilized (though more stabilized than with a constant service rate); see Chapter 4. Multiple alternative service rate controls are discussed in Chapter 4. We develop a damped time-lag service-rate control in Chapter 5 that stabilizes the expected waiting time fairly well, though not perfectly, as shown by the yellow line. The time lag intuition is taken into consideration in the formulation of this new control as defined in equation 5.4, which is the rate-matching control modified by a time lag and a

#### damping factor.



Figure 1.4: This figure plots the expected steady-state virtual waiting time process as a function of time in a cycle for a TV single-server queue. The arrival rate function is  $\lambda(t) = 0.8 \times (1+0.2 \sin(0.1t))$  and the arrival process is Markovian. The blue line shows the waiting time when the service rate is a constant  $\mu = 1$ , the red line shows the waiting time with the rate-matching service-rate control and the yellow line shows the waiting time with the damped time-lag service-rate control. Dashed black lines show confidence intervals, which are quite narrow.

To connect the TV model to the well studied stationary models, we can use the time-transformation composition construction for general non-Possion arrival and service processes, which was proposed by Massey and Whitt (1994), Gerhardt and Nelson (2009) and Nelson and Gerhardt (2011); see Chapter 2 for details. To simulate such a general non-stationary non-Poisson point process (NNPP), we can make use of this construction and convert generated stationary processes to non-stationary processes, which can be efficiently done by algorithms developed in Chapter 2 even though the cumulative arrival rate function is not directly invertible. By making use of this construction, TV model with the rate-matching control can be regarded as a time-transformation of the stationary model, so that the queue length process has a stationary steady-state and thus a stabilized distribution; see theorems 4.2.1 and 4.2.2.

#### 1.2.1 Literature Review

On the structural results for the steady-state processes for TV single-server queues, Harrison and Lemoine (1977), Heyman and Whitt (1984), Lemoine (1981, 1989), Rolski (1981, 1989a,b) mainly establish limiting theorems for the virtual and actual waiting time processes in  $M_t/G/1$  queues with periodic Poisson arrivals. We base our analysis on the extension of the representation for periodic steady-state workload processes in Lemoine (1981) and go beyond to consider general non-Poisson input.

On the numerical algorithms for calculating performance measures in TV queues, the ODE approach based on CTMC to the  $M_t/M_t/s_t$  queues has become the accepted approach: Koopman (1972), Kolesar et al. (1975), Rothkopf and Oren (1979), Taaffe and Ong (1987), Ong and Taaffe (1989) mainly study the ODE approach and closure approximations for ODE equations to reduce their size. The single-server setting can be regarded as a special case of the multi-server system. A useful approximation for stochastic TV models is the deterministic fluid model and Edie (1954), Oliver and Samuel (1962), May and Keller (1967), Newell (1982), Porteus (1989, 1993a,b) discuss the single-server TV fluid model. Robust optimization approach was proposed more recently in Bertsimas et al. (2011), Ben-Tal et al. (2009), Beyer and Sendhoff (2007) and Whitt and You (2018) develops TV robust queueing techniques to derive tractable optimization approximations to the mean steady-state workload in TV single-server queues.

On the asymptotic HT methods and their approximations, Newell (1968a,b,c), Massey (1985), Mandelbaum and Massey (1995) study diffusion approximations through HT limits for TV singleserver queues and obtain direct HT approximations for the  $M_t/M_t/1$  queues. More recently, Whitt (2014) introduces a new HT scaling to expose the TV behavior for  $G_t/GI/1$  queue length processes and Whitt (2016a) provides a new perspective on more possibilities for the scaling. We base our HT analysis on these scalings and study HT limits for workload and virtual waiting time processes in  $G_t/G/1$  and  $G_t/G_t/1$  queues.

On the Monte Carlo simulation techniques to generate arrival processes for the queues, Lewis and Shedler (1979), Gerhardt and Nelson (2009) discuss the thinning method and Çinlar (1975), Chen and Schmeiser (1992), Nicol and Leemis (2014), Chen and Schmeiser (2015, 2017) discuss the inversion approach to to generate a NHPP from a stationary Poisson process. He et al. (2016) generates NNPP arrival processes by inversion from the base renewal process, where the inversion is done by searching the x-axis for a given y value. We develop a more efficient algorithm to simulate a general NNPP arrival process and generate each arrival in O(1) time, which is achieved by tabling the piecewise-constant approximate inverse function outside of simulation; see Chapter 2 and Ma and Whitt (2015a). A new paper Liu et al. (2018) proposes a combined inversion-andthinning approach for simulating NNPPs, which applies thinning to the generated NNPPs with a piecewise-constant approximate arrival rate function. This approach generates the NNPP with the exact arrival rate function without error by adding the thinning step, while our approach is more efficient in running time but subject to a small error bound; see Theorem 2.3.1. We apply our algorithm to simulating TV queues to estimate performance measures, where a large number of i.i.d. experiments are conducted with a large number of arrivals generated for each experiment (say 40,000 experiments with 16,000 arrivals each), so a high time-efficiency is required.

While we study the stabilization problem in single-server queues proposed by Whitt (2015), there has been substantial literature on the server staffing problem in multi-server systems; for example, Liu and Whitt (2012b), Defraeye and van Nieuwenhuyse (2013), Yom-Tov and Mandelbaum (2014), He et al. (2016). The service rate controls analogous to the square-root staffing formula do not perform well in single-server queues and new controls are needed to achieve good performance; see Chapter 4 and 5 for details. The service-rate control problem allocates capacity to a single queue at different times, which is similar to the capacity allocation problem studied in Kleinrock (1964), Wein (1989) for open Jackson networks. Since performance measures are generally not mathematically computable, simulation experiments are needed to evaluate performance of different service-rate controls (or server-staffing formulas for multi-server queues). We develop efficient algorithms to simulate a general NNPP arrival process, which we show can be applied to simulate general nonstationary non-Markovian queues including  $G_t/G_t/1$  queues; also see He et al. (2016), Li et al. (2016), Whitt and Zhao (2017) for applications of the algorithm.

## **1.3** Main Contributions

This thesis makes the following key contributions:

1. We develop new methods to simulate general NNPPs exploiting the inverse tabling method (§2.3). We use this new algorithm to directly generate general arrival processes in non-

stationary non-Markovian queueing models. As for generating the service processes, we construct the service models specifying deterministic service rates separately from stochastic service requirement processes and still make use of the new algorithm ( $\S2.4$ ).

- 2. We base our analysis of single-server TV queues on the Lemoine representation of steadystate workload process for periodic  $M_t/G/1$  queues. We extend the expression first to  $G_t/G/1$ queues and then to  $G_t/G_t/1$  queues to study the TV behavior of steady-state virtual waiting times (§3.2, §5.5). We extend the classic rare-event simulation algorithm for GI/GI/1 queues to  $GI_t/GI/1$  and further to  $GI_t/GI_t/1$  queues. This is achieved by extending the Lemoine representation and establishing bounds between the periodic workload and the stationary workload with the average arrival rate, so that the relative error for estimators of  $P(W_y > b)$ can be proved to be uniformly bounded in b (§3.4, §5.5). Based on the extended rare-event simulation algorithm, an efficient algorithm is developed to calculate the periodic steadystate distribution and moments of the virtual waiting time  $W_y$  at time yC within a cycle of length C in periodic single-server queues  $GI_t/GI/1$  and  $GI_t/GI_t/1$ . We use this algorithm to understand the dynamics of the waiting time and workload processes as the arrival function changes (§3.4.5).
- 3. We develop the HT limit with periodicity for the workload process in  $G_t/G/1$  queues and also for workload and virtual waiting time processes in  $G_t/G_t/1$  queues making use of two scalings (§3.6, §5.6 - §5.8). With the aid of the heavy-traffic limit theorem, the developed rareevent algorithm also applies to compute the periodic steady-state distribution of (i) reflected periodic Brownian motion by considering appropriately scaled  $GI_t/GI/1$  models and (ii) a large class of general  $G_t/G/1$  queues by approximating by  $GI_t/GI/1$  models with the same heavy-traffic limit (§3.6).
- 4. We study the stabilization problem in single-server queues by formulating alternative service rate controls (§4.2). We evaluate the performance of four controls: the rate-matching control, which is directly proportional to the arrival rate, stabilizes the queue length distribution; the two controls analogous to the multi-server staffing formula are not effective in stabilizing performance measures; the other control based on the mean stationary waiting time stabilizes the expected virtual waiting time when the arrival rate changes slowly (§4.4).

- 5. We formulate a new damped time-lag control that stabilizes the virtual waiting time fairly well even if the arrival rate does not change slowly (§5.1.4). This new control is based on a simulation search algorithm we develop (§5.5). The HT limits provide insights into the service-rate controls. The state space collapse in our theorem shows that there is a TV Little's law in heavy-traffic, implying that the queue length and waiting time cannot be simultaneously stabilized in this limit with the damped time-lag control (§5.6).
- 6. We conduct extensive simulation experiments to show the accuracy and efficiency of both our standard simulation and our rare-event simulation algorithms for both  $GI_t/GI/1$  queues and  $GI_t/GI_t/1$  queues and to evaluate the performance of alternative service-rate controls (§3.5, §4.4, §5.9).

## 1.4 Outline

This thesis consists of two parts: *performance evaluation* and *performance stabilization*. Part I studies simulation techniques to evaluate performance of a TV single-server queue, which provides the basis for studies on improving performance in the same queueing system in Part II. It includes Chapter 2 and 3. In Chapter 2, which is based on section 2 of Ma and Whitt (2015b) and Ma and Whitt (2015a), an efficient algorithm to simulate a general NNPP is proposed and this algorithm can be applied to simulate a non-stationary non-Markovian queueing system, for example, a TV single-server model with TV service-rates. In Chapter 3, we develop the rare-event simulation algorithm in TV single-server queues to estimate the tail probabilities and moments of the steady-state workload process in TV queues; this is based on Ma and Whitt (2018b).

Part II works on the performance stabilization problem in TV single-server queues, which is achieved by service-rate controls as a function of the arrival rates. This part makes use of the simulation tools in Part I to find optimal parameters for service-rate controls and evaluate their performance. Chapter 4 is based on Ma and Whitt (2015b). It studies the rate-matching control, two square-root controls analogous to the server-staffing formula and a PSA-based control. The rate-matching control can perfectly stabilize the queue length process and the PSA-based control can stabilize the expected waiting time when PSA is appropriate. Chapter 5, based on Ma and Whitt (2018a), develops a damped time-lag control, which requires the simulation search algorithm based on Chapter 3 to find the optimal damping and time-lag parameters. This control performs fairly well in stabilizing the waiting time distribution when PSA is not appropriate.

# Part I

# Performance Evaluation for Time-Varying Single-Server Queues

# Chapter 2

# Efficient Simulation of Time-Varying Queues

Efficient simulation algorithms are developed to evaluate the performance, e.g. expected waiting time, expected queue length, in a queue with time-varying arrival rates, service in order of arrival and unlimited waiting space. Both Markovian and non-Markovian models are considered. Customer service requirements are specified separately from the service rate, which is time-varying as well. New versions of the inverse method exploiting tables constructed outside the simulation are developed to efficiently simulate a general non-Poisson non-stationary point processes for queueing approximations, which can be applied to generate both the arrival times and service times. This chapter is based on Ma and Whitt (2015b) and Ma and Whitt (2015a).

# 2.1 Introduction

In this chapter we study efficient simulation to evaluate performance in a single-server queue with unlimited waiting space, service provided in order of arrival, time-varying arrival rate and independent and identically distributed (i.i.d.) service requirements specified separately from the service rate actually provided. This simulation approach will be applied to the performance stabilization problem later in Chapter 4.

Specifically, we consider a class of general  $G_t/G_t/1$  single-server queues with unlimited waiting space, service in order of arrival, a time-varying arrival rate, and a time-varying service rate. Our

methods apply to general arrival rate functions, but as in previous work we use stylized sinusoidal arrival rate functions with a range of parameters. We consider arrival processes that are time-transformed stationary renewal processes, with the specified arrival rate function. We assume that the service requirements are i.i.d. random variables with a general distribution, specified independently of the service rate control. The general GI arrival and service processes allow different levels of stochastic variability to go with the predictable deterministic variability of the time-varying rates.

We develop new methods to simulate these non-stationary non-Markovian queueing models. As in §7 of Massey and Whitt (1994), Gerhardt and Nelson (2009) and He et al. (2016), we represent the arrival process as the composition of a rate-1 stationary point process and the deterministic cumulative arrival rate function. For this study we use renewal processes for the base rate-1 process, but the method is more general. We efficiently generate both the service times and the arrival times by exploiting tabled inverse functions, as can be done in generating non-uniform random numbers; see §11.2 and §III.2 of Devroye (1986) and §3.8 of L'Ecuyer (2012).

The remainder of this chapter is organized as follows. In §2.2 we define the  $G_t/G_t/1$  model and then we discuss the simulation methodology for generating the non-stationary non-Markovian models. We describe in detail the construction for non-stationary non-Poisson processes, the inverse function tabling algorithm in §2.3 and the service time model in §2.4, after which the simulation is elementary.

# **2.2** The $G_t/G_t/1$ Model

We construct the arrival and service processes by using deterministic time-transformations of general rate-1 processes. We first consider the arrival counting process A, where A(t) counts the number of arrivals occurring in the time interval [0, t]. We define A using a cumulative arrival rate function

$$\Lambda(t) = \int_0^t \lambda(s) \, ds, \quad t \ge 0, \quad \text{where} \quad 0 < \lambda_L \le \lambda(t) \le \lambda_U < \infty, \tag{2.1}$$

and a general rate-1 counting process N with unit jumps. We define A by the composition

$$A(t) \equiv N(\Lambda(t)), \quad t \ge 0.$$
(2.2)

Given that E[N(t)] = t,  $t \ge 0$  (the rate-1 property), A defined by (2.2) has the specified rate:  $E[A(t)] = E[N(\Lambda(t))] = \Lambda(t)$ . The deterministic function  $\Lambda(t)$  specifies the predictable variability, while all the unpredictable stochastic variability is specified by the base counting process N. This construction is without loss of generality, because given any A with unit jumps and  $E[A(t)] = \Lambda(t)$ , we can let  $N = A(\Lambda^{-1}(t)), t \ge 0$ , where  $\Lambda^{-1}$  is the inverse of  $\Lambda$ , which is well defined. Hence, (2.2) holds with  $E[N(t)] = t, t \ge 0$ .

We now turn to the service process. Paralleling our model of the arrival process, we assume that the service requirements are generated by a counting process  $N_s$  with unit jumps, which is independent of N. We define the evolution of the queueing model, given the arrival process A, the service requirement process  $N_s$  and the time-varying service-rate control  $\mu(t)$ , by jointly defining the number in system Q(t) and the departure counting process D(t). In particular, we require that these processes satisfy the two equations

$$Q(t) = A(t) - D(t) \quad \text{and} \quad D(t) \equiv N_s(\int_0^t \mu(s) \mathbb{1}_{\{Q(s)>0\}} \, ds), \quad t \ge 0,$$
(2.3)

The representation (2.3) can be justified by applying mathematical induction to the successive event changes in Q(t); see §2.1 of Pang et al. (2007). Note that the process D has the service rate  $\mu(t)$  whenever the system is not empty:  $E[D(t)] = \int_0^t \mu(s) \mathbb{1}_{\{Q(s)>0\}} ds, t \ge 0.$ 

In this chapter we consider the special case of the model above in which the service requirements  $S_k$  are i.i.d random variables with a general cdf G having mean 1 and finite second moment. If the mean were not actually 1 initially, we could rescale both these service requirements and the service-rate control to make it so, so that is without loss of generality. The associated rate-1 counting process is the equilibrium version of the renewal counting process, which differs from the ordinary renewal counting process only by having the first interval having the stationary-excess cdf  $G_e(t) = \int_0^t [1 - G(s)] ds, t \ge 0$ , instead of the cdf G of all other intervals. The same holds for the arrival process. We will generate N using i.i.d. random variables with mean 1; then the associated rate-1 process is the equilibrium renewal process.

Often an exceptional first interval is not too important, and can be considered part of the initial conditions, along with starting the queueing system empty. We then can generate both the arrival process and the service process using ordinary renewal processes with mean-1 inter-renewal times. Then the arrival rate is asymptotically correct as  $t \to \infty$ .

To simulate the model, we first generate the successive arrival times and then the successive service times. It is then straightforward to construct the associated queueing processes. We next describe an efficient simulation algorithm for general non-Poisson non-stationary point processes which can be directly applied to generate the arrival process. After that, we explain the steps for generating service times.

# 2.3 Efficient Simulation

In simulation experiments to evaluate queueing performance, it has been accepted practice to use stylized arrival rate functions that capture essential features of arrival rate functions that can be estimated from data. In particular, it has been standard to use the sinusoidal arrival rate function

$$\lambda(t) \equiv \lambda(t; \bar{\lambda}, \beta, \gamma) \equiv \bar{\lambda}(1 + \beta \sin(\gamma t)) \quad \text{for} \quad 0 < \beta < 1 \quad \text{and} \quad \gamma > 0, \tag{2.4}$$

where  $\bar{\lambda}$  is the average arrival rate (the spatial scale),  $\beta$  is the relative amplitude and  $\gamma$  is the time scaling factor, determining the associated cycle length  $C = 2\pi/\gamma$ .

Our main idea for simulating non-Poisson non-stationary arrival processes is to exploit the inverse method, as often used in generating non-uniform random numbers; see §II.2 and §III.2 of Devroye (1986) and §3.8 of L'Ecuyer (2012). The inverse method can be used for NHPP's, but it is even more appealing here because it allows us to efficiently simulate a large class of non-Poisson non-stationary arrival processes, not just one.

Since the arrival times of A and N, denoted by  $A_k$  and  $N_k$  respectively, are related by

$$A_k = \Lambda^{-1}(N_k), \quad k \ge 1, \tag{2.5}$$

the first step in this approach to construct a large class of non-Poisson non-stationary arrival process models is by using the inverse  $\Lambda^{-1}$  of the cumulative arrival rate function  $\Lambda$  provided that an efficient algorithm is available for generating the rate-one process N. The cumulative arrival rate function  $\Lambda$  for the sinusoidal arrival rate function in (2.4) is

$$\Lambda(t) \equiv \int_0^t \lambda(s) \, ds = \bar{\lambda} [t + (\beta/\gamma)(1 - \cos(\gamma t))], \quad t \ge 0.$$
(2.6)

The associated inverse function  $\Lambda^{-1}$  is well defined for (2.6) and any arrival rate function for which

$$0 < \lambda_L \le \lambda(t) \le \lambda_U < \infty \quad \text{for all} \quad 0 \le t \le C < \infty;$$
(2.7)

e.g., we could apply basic properties of inverse functions, as in §13.6 of Whitt (2002).

Since the inverse function  $\Lambda^{-1}$  is often unavailable explicitly, we construct a suitably accurate approximation of it and apply it by table lookup. In §2.3.1 we explain how the possibility of reuse provides remarkable efficiency; in §2.3.2 we develop an algorithm to efficiently construct the approximate inverse function with specified accuracy; and in §2.3.3 we discuss additional application issues.

#### 2.3.1 Efficiency Through Re-Use

The main advantage of the inverse function approach is the possibility of re-use. Since the inverse function satisfies a fixed point equation, an alternative way to calculate the inverse is to solve the fixed point equation for each arrival time, perhaps by search, exploiting the monotonicity. That is done in Chen and Schmeiser (1992). However, that search has to be performed at each arrival time. The search has the advantage that there should usually be many fewer arrivals in a fixed interval [0, C] than arguments in a tabled inverse function, but the inverse function has the advantage that the simulation and re-used. Moreover, the calculation from the table can be very fast, because it is possible to proceed forward through the table only once.

#### 2.3.1.1 One Cycle for Periodic Arrival Rate Functions

The algorithm can be accelerated if the arrival rate function is periodic, because it suffices to calculate the inverse only for a single cycle. For example, with the sinusoidal arrival rate function in (2.4),  $\Lambda(2k\pi/\gamma) = \bar{\lambda}2k\pi/\gamma$  for all integers  $k \ge 0$ , so that  $\Lambda^{-1}(2k\bar{\lambda}\pi/\gamma) = 2k\pi/\gamma$  for all integers  $k \ge 0$ . Hence, it suffices to construct the inverse for  $0 \le t < 2\pi/\gamma$ . Overall, we get

$$\Lambda^{-1}((2k\bar{\lambda}\pi/\gamma)+t) = (2k\pi/\gamma) + \Lambda^{-1}(t), \quad 0 \le t \le 2\bar{\lambda}\pi/\gamma, \tag{2.8}$$

so that it suffices to calculate  $\Lambda^{-1}$  on the interval  $[0, 2\bar{\lambda}\pi/\gamma]$ .

#### 2.3.1.2 Different Scaling of Time and Space

We also can use one constructed inverse function  $\Lambda^{-1}$  to obtain inverse functions for scaled versions of the original function  $\Lambda$ . This commonly occurs with sinusoidal arrival rate functions  $\lambda(t; \bar{\lambda}, \beta, \gamma)$  in (2.4). We are often interested in different spatial and temporal scale parameters  $\bar{\lambda}$  and  $\gamma$ . Since

$$\Lambda(t;\bar{\lambda},\beta,\gamma) = \bar{\lambda}\Lambda(\gamma t;1,\beta,1)/\gamma, \qquad (2.9)$$

we can apply Lemma 13.6.6 of Whitt (2002) to express the inverse as

$$\Lambda^{-1}(t;\lambda,\beta,\gamma) = \Lambda^{-1}(\gamma t/\lambda;1,\beta,1)/\gamma.$$
(2.10)

Hence, we can use the constructed inverse function  $\Lambda^{-1}(t; 1, \beta, 1)$  for  $\Lambda(t; 1, \beta, 1)$  to construct the inverse function  $\Lambda^{-1}(t; \bar{\lambda}, \beta, \gamma)$  for  $\Lambda(t; \bar{\lambda}, \beta, \gamma)$ ; i.e., we can reduce the three parameters to just one.

#### 2.3.1.3 Multiple Non-Poisson Non-Stationary Arrival Process Models

In order to evaluate performance approximations and system controls such as staffing algorithms, we need to consider a variety of models to ensure that the methods are successful for a large class of models. It is thus significant that a constructed inverse function  $\Lambda^{-1}$  can be re-used with different rate-1 stochastic counting processes N. For any rate-1 counting process N that we can simulate, we can generate the corresponding non-stationary arrival process with the same arrival rate function  $\lambda$  simply by applying the tabled inverse function to the arrival times of that rate-1 process, as in (5.11). Methods for simulating stationary counting processes are well established.

#### 2.3.1.4 Multiple Replications to Obtain Accurate Performance Estimates

The tabled inverse function can be re-used in each replication when many replications are performed to obtain accurate performance estimates. For example, we might use  $10^4$  or more i.i.d. replications.

#### 2.3.2 The Inverse Function

By (5.11), if we can simulate the arrival times  $N_k$  of the designated rate-1 process, then to simulate the desired arrival times  $A_k$  of the non-stationary point process A, it only remains to compute  $\Lambda^{-1}(N_k)$  for each k. This is straightforward if the inverse function is available explicitly. If we use data to estimate the cumulative arrival rate function, then we can fit a convenient invertible function  $\Lambda$ . Indeed, there seems to be no reason not to use an invertible function. For example, it could be a piecewise-linear function as in Gerhardt and Nelson (2009), Leemis (1991), Massey et al. (1996), Nelson and Gerhardt (2011). However, starting with an explicit non-invertible function  $\Lambda$ , as in (2.6), we want to efficiently construct an approximation of  $\Lambda^{-1}$  that is (i) easy to implement, (ii) fast in its implication and has (iii) suitably small specified accuracy. We could act just as if we had data, and fit a convenient invertible function, but then it remains to substantiate that the three goals have been met. To achieve these three goals, we contend that a good approach is to construct a piecewise-constant approximation. Of course, this construction can yield multiple points when that is not possible in the counting process A, but that is easily eliminated if it is deemed important; see §2.3.3.4. At some extra work, we could convert the piecewise-constant approximation to a piecewise-linear approximation, paralleling §Leemis (1991). For all these modifications, our error bound still applies. For the queueing applications, this last refinement step should usually not be necessary.

We assume that a cumulative arrival rate function  $\Lambda$  associated with an arrival rate function  $\lambda$  satisfying (5.1) is given over a finite interval [0, C]. By (5.1), there exists a function r such that  $\Lambda^{-1}(t) = \int_0^t r(s) \, ds, \ 0 \le t \le \Lambda(C)$ , and

$$0 < 1/\lambda_U \le r(t) \le 1/\lambda_L < \infty, \quad 0 \le t \le \Lambda(C).$$
(2.11)

Our goal is to efficiently construct an approximation J to the inverse function  $\Lambda^{-1}$  mapping the interval  $[0, \Lambda(C)]$  into [0, C] with specified accuracy

$$\|J - \Lambda^{-1}\| \equiv \sup_{0 \le t \le \Lambda(C)} \{|J(t) - \Lambda^{-1}(t)|\} \le \epsilon$$

$$(2.12)$$

for some suitably small target  $\epsilon > 0$ . This is a natural way to quantify the error, because  $\epsilon$  specified the maximum error in the arrival times.

Our general strategy is to partition the two intervals [0, C] and  $[0, \Lambda(C)]$  into  $n_x$  and  $n_y$  evenly spaced subintervals of width  $\delta_x$  and  $\delta_y$ , respectively, and then define J at  $i\delta_y$  to be an appropriate  $j\delta_x$ , for each  $i, 0 \le i \le n_y$ . We extend J to  $[0, \Lambda(C)]$  by making J a right-continuous step function, assuming these constant values specified at  $i\delta_y$ .

Key parameters for our algorithm are

$$\omega \equiv \omega_{\Lambda} \equiv \frac{\lambda_U}{\lambda_L}, \quad \delta_x = \frac{\epsilon}{1+\omega} \quad \text{and} \quad \delta_y = \lambda_U \delta_x = \frac{\lambda_U \epsilon}{1+\omega}, \tag{2.13}$$

where  $\lambda_L$  and  $\lambda_U$  are the lower and upper bounds on the arrival rate function  $\lambda$  given in (5.1) and  $\epsilon$  the desired error bound in (2.12). Thus  $\omega$  is the slope ratio with  $1 \leq \omega < \infty$ , while  $\delta_y$  and  $\delta_x$  are spacings used to achieve the target error bound  $\epsilon$  in (2.12).

27

To construct J, we first calculate  $\Lambda(m)$  for each of the  $n_x + 1$  points x' in [0, C] by letting

$$a(j) \equiv \Lambda(j\delta_x), \ 0 \le j \le n_x.$$
(2.14)

Then we approximate the  $\Lambda^{-1}(y')$  value of each of the  $n_y + 1$  points y' in  $[0, \Lambda(C)]$  by a suitable point within the  $n_x$  points in [0, C], i.e.,

$$b(i) \equiv \inf \{j \ge 0 : a(j) \ge i\delta_y\}, \quad 0 \le i \le n_y.$$

$$(2.15)$$

Then  $J(i\delta_y) = b(i)\delta_x$  for all  $i, 0 \le i \le n_y$ . The simple vector representations in (2.14) and (2.15) are the basis for the implementation efficiency.

Algorithm 1 Constructing the approximation J of the inverse function  $\Lambda^{-1}$  for given time C, function  $\Lambda : [0, C] \to [0, \Lambda(C)]$  and error bound  $\epsilon$ 

- 1: Set  $\omega \leftarrow \lambda_U / \lambda_L$ ,  $\delta_x \leftarrow \epsilon / (1 + \omega)$ ,  $\delta_y \leftarrow \overline{\lambda_U \epsilon / (1 + \omega)}$ ,  $n_x \leftarrow C(1 + \omega) / \epsilon$ ,  $n_y \leftarrow \Lambda(C) / \delta_y / /$  (five constant parameters)
- 2: Set  $x' \leftarrow (0 : \delta_x : C), y' \leftarrow (0 : \delta_y : \Lambda(C)) //(two equally spaced vectors of length <math>n_x + 1$  and  $n_y + 1)$
- 3: Set  $a \leftarrow \Lambda(x), b \leftarrow [] //(two new vectors of length <math>n_x + 1$  and  $n_y + 1$  with b zero vector)
- 4: Set  $i \leftarrow 1, j \leftarrow 1$  //(initialize for  $n_x + n_y$  operations)
- 5: While  $j < n_x + 1$  &&  $i < n_y + 1$  do
- 6: If y(i) > a(j) Then
- 7:  $j \leftarrow j + 1$
- 8: Else
- 9:  $b(i) \leftarrow j, \quad i \leftarrow i+1$
- 10: End if
- 11: End While

12:  $//(For \ 0 \le i \le n_y, \ J(i\delta_y) = b(i)\delta_x; \ J \ extended \ to \ [0, \Lambda(C)] \ by \ right-continuity.)$ 

We can finally get the value of J at any time y' in  $[0, \Lambda(C)]$  by

$$J(y') = J(|y'/\delta_y|\delta_y), \ 0 \le y' \le \Lambda(C), \tag{2.16}$$

where  $\lfloor y' \rfloor$  is the floor function, yielding the greatest integer less than or equal to y. However, this extension is not used directly because we start by changing  $N_k$  to  $\lfloor N_k/\delta_y \rfloor \delta_y$ , so we only use J defined on the finite subset  $\{i\delta_y : 0 \le i \le n_y\}$ . The function J is constructed to be one-to-one on the finite subset  $\{i\delta_y : 0 \le i \le n_y\}$ .

**Theorem 2.3.1.** (error bound and computational complexity) Algorithm 1 above constructs a nondecreasing function J on  $[0, \Lambda(C)]$  approximating  $\Lambda^{-1}$  with the error upper bound  $\epsilon$  prescribed in (2.12) using of order  $O(n_x + n_y) = O(2C(1 + \omega)/\epsilon)$  storage (two vectors each of size  $n_x$  and  $n_y$ ) with computational complexity of order  $O(n_x + n_y) = O(2C(1 + \omega)/\epsilon)$ .

**Proof** For any  $\delta_y > 0$  and  $\delta_x > 0$ , a bound on the error in J is

$$\|J - \Lambda^{-1}\| \equiv \sup_{0 \le t \le \Lambda(C)} |J(t) - \Lambda^{-1}(t)| = \sup_{0 \le i \le n_y} \sup_{t \in [i\delta_y, (i+1)\delta_y)} |J(i\delta_y) - \Lambda^{-1}(t)|$$
  

$$= \sup_{0 \le i \le n_y} \sup_{t \in [i\delta_y, (i+1)\delta_y)} |b(i)\delta_x - \Lambda^{-1}(i\delta_y) + \Lambda^{-1}(i\delta_y) - \Lambda^{-1}(t)|$$
  

$$\leq \sup_{0 \le i \le n_y} (|b(i)\delta_x - \Lambda^{-1}(i\delta_y)| + |\Lambda^{-1}(i\delta_y) - \Lambda^{-1}((i+1)\delta_y)|)$$
  

$$\leq \delta_x + \delta_y / \lambda_L, \qquad (2.17)$$

where the fourth line follows because the point  $\Lambda^{-1}(i\delta_y)$  lies in the interval  $(b(i)\delta_x, b(i+1)\delta_x]$ .

Next observe that the function J will be one-to-one (have distinct values) on the set  $\{i\delta_y : 0 \le i \le n_y\}$  if  $\delta_y \ge \lambda_U \delta_x$ . Now we choose  $\delta_y$  such that

$$\delta_y = \lambda_U \delta_x. \tag{2.18}$$

Then J is one-to-one on  $\{i\delta_y : 0 \le i \le n_y\}$  and, by (2.17) and (2.18),

$$\|J - \Lambda^{-1}\| \le \delta_x + \delta_y / \lambda_L \le \frac{\epsilon}{1 + \omega} + \frac{\omega\epsilon}{1 + \omega} = \epsilon.$$
(2.19)

Turning to the computational complexity, we see that four vectors need to be stored: x', y', a and b, which is of total length  $2(n_x + n_y + 2)$ . To construct the table of J, the while loop in algorithm 1 searches for b(i) for each  $0 \le i \le n_y$ , which checks each of the  $(n_x + n_y)$  points only once and takes time  $O(n_x + n_y)$ . Finally, by (2.13) again,

$$n_x + n_y = \frac{C}{\delta_x} + \frac{\Lambda(C)}{\delta_y} = \frac{C(1+\omega)}{\epsilon} + \frac{\Lambda(C)(1+\omega)}{\lambda_U \epsilon} \le \frac{2C(1+\omega)}{\epsilon}.$$
 (2.20)

#### 2.3.3 Application Issues

#### 2.3.3.1 Generating the Arrival Times

Given Algorithm 1, the algorithm to construct the actual arrival times  $A_k = \Lambda^{-1}(N_k)$  given all the rate-1 arrival times  $N_k$  can be very simple. If we apply the floor function and the inverse function in Algorithm 1 in a single vector operation to all components of the vector of rate-1 arrival times, then the code can be expressed in a single line.

Algorithm 2 constructing the vector  $A \equiv \{A_k\}$  of arrival times in [0, C] given Algorithm 1 specified in terms of the triple  $(\delta_y, \delta_x, b)$  depending on the error bound  $\epsilon$  in (2.12) and the associated nondecreasing vector of nonnegative rate-1 arrival times  $N \equiv \{N_k : 1 \le k \le n\}$  with  $N_n \le \Lambda(C)$ 1: Set  $A \leftarrow b(\lfloor N/\delta_y \rfloor) \delta_x //$  (vector application of the floor function and Algorithm 1 term by term)

In the single line of Algorithm 2 we have used (2.16) and line 12 of Algorithm 1, i.e.,

$$J(\lfloor t/\delta_y \rfloor \delta_y) = b(\lfloor t/\delta_y \rfloor) \delta_x \quad \text{or} \quad J(i\delta_y) = b(i)\delta_x, \quad 0 \le i \le n_y.$$
(2.21)

This is important for implementation efficiency, because we make only one pass through the table to generate all the arrival times  $A_k$ .

#### 2.3.3.2 Partitioning Into Subintervals

For difficult arrival rate functions, it might be preferable to modify the representation of the inverse function, e.g., moving closer to a piecewise-linear approximation. In particular, if the slope ratio  $\omega$ in (2.13) is large, then it may be easy to accelerate the algorithm by dividing the original interval [0, C] into subintervals. A simple example is a piecewise linear function with two pieces, one having a flat slope and the other having a steep slope, so that the ratio  $\omega$  might be very large. If we divide the interval into the two parts where  $\Lambda$  is linear, then  $\omega$  is reduced to 1 on one subinterval. Given that we divide [0, C] into the two intervals  $[0, C_1]$  and  $[C_1, C]$ , we can calculate  $\Lambda^{-1}$  separately on the two intervals  $[0, \Lambda(C_1)]$  and  $[\Lambda(C_1), \Lambda(C)]$ .

#### 2.3.3.3 Choosing the Error Bound

It is natural to ask how the error bound  $\epsilon$  should be chosen in practice. We think it should usually be possible to choose  $\epsilon$  relatively small compared to an expected interarrival time of A, which has a time-varying value exceeding  $1/\lambda_U$  for  $\lambda_U$  in (5.1). However, for queueing applications that might be smaller than necessary, because the relevant time scale in a queueing system is typically of order equal to a mean service time, which depends on the units used to measure time. Suppose, without loss of generality, we choose the time units so that the mean service time is 1. Then we think it usually should suffice to let  $\epsilon$  be small compared to the maximum of these, e.g.,  $\epsilon \approx \max\{1, 1/\lambda_U\}/100$ .

To illustrate, consider an example of a moderately large call center in which the mean service time is about 5 minutes, while the arrival rate is 600 per hour or 1/6 per second, as in §3.1 of Kim and Whitt (2014), which makes  $\lambda_U = 600/12 = 50$  in units of mean service times. In this context, it directly seems reasonable to let  $\epsilon$  be one second. The rough guideline above yields  $\epsilon = 300/100 = 3$ seconds.

Assuming that time is measured in mean service times and  $\lambda_U \ge 1$  in that scale, the computational complexity from Theorem 2.3.1 becomes  $2C(1 + \omega) \times 10^2$ . In the call center example, if we let  $C = 24 \times 12 = 288$  corresponding to one 24-hour day measured in units of 5 minute-calls, then the computational complexity of the algorithm to calculate the inverse function is 57,  $600(1 + \omega)$ .

#### 2.3.3.4 Breaking Ties: Ensuring an Orderly Point Process

We have constructed the approximate inverse function J to be one-to-one in the finite subset  $\{i\delta_y : 0 \le i \le n_y\}$ . However, that does not present multiple points in A, because all points from the rate-1 process N in the interval  $[i\delta_y, (i+1)\delta_y)$  are mapped into the same point  $b(i)\delta_x$ , for each  $i, 0 \le i \le n_y - 1$ .

First, we can easily identify multiple points by looking for the zeros in the vector  $\Delta A$ , where  $\Delta A_k \equiv A_k - A_{k-1}$ . Then we can easily remove them if we want. Suppose that  $A_{k-1} < A_k = A_{k+j} < A_{k+j+1}$  for some  $k \ge 1$  and  $j \ge 1$ . Then replace  $A_{k+i}$  by  $A_k + i\epsilon/(j+1)$ ,  $1 \le i \le j$ . We could further randomize by using  $A_k + (i + U_{k+i})\epsilon/(j+1) + 1 \le i \le j$ , where  $\{U_k : k \ge 1\}$  is a sequence of i.i.d. uniform random variables on [0, 1]. However, these adjustments should not be required for queueing applications if we are satisfied with the "measurement error" of  $\epsilon$ , as discussed in §2.3.3.3.

#### 2.3.3.5 Selecting the Rate-One Stochastic Process N

In applications, a key remaining problem is actually identifying an appropriate non-Poisson nonstationary arrival process. Assuming that ample data are available to estimate the cumulative arrival rate function, the question about choosing A is roughly equivalent to the question about choosing the rate-1 process N for given cumulative arrival rate function  $\Lambda$ .

As discussed in Massey and Whitt (1994), He et al. (2016), it is natural to specify the functional central limit theorem behavior of N, by the asymptotic index of dispersion for the arrival process A, i.e., we use measurements of A to estimate

$$c_a^2 \equiv \lim_{t \to \infty} \frac{Var(A(t))}{E[A(t)]} = \lim_{t \to \infty} \frac{Var(N(t))}{E[N(t)]}.$$
(2.22)

It is then easy to choose stationary renewal processes N with this  $c_a^2$  Whitt (1982). However, while this should yield an appropriate  $c_a^2$ , this does not nearly specify the processes N and A fully. However, heavy-traffic limit theorems indicate that this may be sufficient; see §4 of He et al. (2016).

#### 2.3.3.6 Random-Rate Arrival Processes

As discussed in Whitt (1999), Kim et al. (2015) and references therein, it may be desirable to represent the arrival rate over each day as random. For example, the model of the arrival process on one day of length T might be

$$A(t) = N(X\Lambda(t)), \quad 0 \le t \le T, \tag{2.23}$$

where N is a rate-1 stochastic processes, perhaps Poisson, while  $\Lambda$  is a deterministic cumulative arrival rate function and X is a positive random variable. The overall cumulative arrival rate of A is

$$E[A(t)] = E[N(X\Lambda(t))] = E[X]\Lambda(t), \quad 0 \le t \le T.$$

$$(2.24)$$

With this structure, we can exploit the scaling properties in §2.3.1.2 to accelerate simulations. In particular, the representation (2.24) can be viewed as a variant of our model in which the cumulative arrival rate function is the random function  $\Lambda_X(t) \equiv X\Lambda(t)$ . Fortunately, the inverse of  $\Lambda_X$  can be expressed directly in terms of the inverse  $\Lambda^{-1}$  and the random variable X by

$$\Lambda_X^{-1}(t) = \Lambda^{-1}(t/X), \quad 0 \le t \le X\Lambda(T)$$
(2.25)

For any single realization of the random variable X above, we can simulate the stochastic process A in the manner described in previous sections. However, to assess the system performance, we would need to consider the values of X over successive days, but these random variables  $X_k$  over successive days k are likely to be dependent with distributions depending on the day of the week and the week of the year. Nevertheless, the inverse in (4.10) can be efficiently calculated for each of these these days using the single inverse function  $\Lambda^{-1}$ . However, by sampling sufficiently many days, we may capture the impact of this random variable X.

## 2.4 Generating the Service Times

We use a similar inverse function method to generate the service times, but the method is more complicated, because to apply (2.3) we need to keep track of when the server is busy. Thus, we start by developing a recursion.

Let  $B_k$ ,  $D_k$ ,  $V_k$  and  $W_k$  be the times that arrival k who arrives at  $A_k$  begins service, departs, spends in service and waits before starting service, respectively. Then we have the basic recursion:  $B_k = \max \{D_{k-1}, A_k\}, D_k = B_k + V_k$  and  $W_k = B_k - A_k$ , where the arrival times  $A_k$  have been generated already. Given that the system starts empty, we can initialize the recursion with  $D_0 = 0$ and  $B_1 = A_1$ , so that the only variable not formulated in the recursion is the service time  $V_k$ .

Since the service requirement  $S_k$  is completed by the server busy working from time  $B_k$  to time  $B_k + V_k$ , the service time  $V_k$  satisfies the equation

$$S_k = \int_{B_k}^{B_k + V_k} \mu(s) \, ds, \quad k \ge 1.$$
(2.26)

We can solve for service times explicitly by

$$V_k = M^{-1}(S_k + M(B_k)) - B_k$$
, where  $M(t) \equiv \int_0^t \mu(s) \, ds$  (2.27)

and  $M^{-1}$  is the inverse of M, which is well defined providing that  $0 < \mu_L \leq \mu(t) \leq \mu_U < \infty$ , paralleling (2.1), which we assume to be the case.

Again we work to reduce the computational burden. Just as for the arrival rate function  $\Lambda$ , we see that the function M is typically periodic, so that we only need to compute  $M^{-1}$  over a single cycle. We avoid performing the integration in the direct definition of M and approximate the function M by the piecewise constant function  $M(x'(i)) = \int_0^{x'(i)} \mu(s) ds \approx \sum_{j=1}^i \mu(x'(j))\tau$ , implemented with the recursion  $M(x'(i+1)) = M(x'(i)) + \mu(x'(i+1))\tau$  for suitably small  $\tau$ , starting with M(x'(0)) = 0. To obtain the  $M^{-1}$  value for each customer, we table the inverse function much as we did for  $\Lambda^{-1}$ .

# 2.5 Conclusions

In this chapter we have developed an efficient algorithm for simulating the model of a time-varying single-server queue with a time-varying service-rate and we have described the efficient algorithm for simulating a general non-Poisson non-stationary point processes.

The model is a single-server queue with service in order of arrival, unlimited waiting space and a time-varying arrival rate function. The simulation algorithm applies to arbitrary arrival rate functions, but as examples we used the sinusoidal periodic arrival rate function in (2.4) with average arrival rate  $\bar{\lambda}$ , relative amplitude  $\beta$  and time-scaling factors  $\gamma$ . The service requirements were i.i.d. random variables specified separately from the time-varying service-rate. The arrival processes were mostly nonhomogeneous Poisson processes, but the method applies to very general arrival processes that can be represented as a deterministic time transformation of a stationary point process as in (2.2). Experiments can be conducted for stationary processes constructed from renewal processes with non-exponential as well as exponential distributions. This allows representing different levels of stochastic variability.

Conducting the simulations for these non-stationary queues turned out to be quite challenging. An important component of the efficient simulation was constructing a table of the inverse cumulative arrival rate function when it is not explicitly available and exploiting table lookup to calculate the arrival times and service times. The use of tables for a periodic arrival rate function is appealing because the table for one cycle can be used for other cycles and for scaled versions of the original arrival rate function, as shown in §2.3. Later in Chapter 4 we will use this simulation algorithm to conduct simulation experiments for the performance stabilization problem.

# Chapter 3

# Rare-Event Simulation for Periodic Queues

In this chapter, an efficient algorithm is developed to calculate the periodic steady-state distribution and moments of the virtual waiting time in a single-server queue with a periodic arrival-rate function. The virtual waiting time at time t is the waiting time of a potential arrival at that time. We use  $W_y$  to denote the steady-state virtual waiting time at time yC within a cycle of length C, 0 < y < 1. The algorithm applies exactly to the  $GI_t/GI/1$  model, where the arrival process is a time-transformation of a renewal process, with the first inter-arrival time of the renewal process having the exceptional equilibrium distribution. A new representation of  $W_y$  shown in this chapter makes it possible to apply a modification of the classic rare-event simulation for the stationary GI/GI/1 model exploiting importance sampling with an exponential change of measure. We establish bounds between the periodic workload and the stationary workload with the average arrival rate that enable us to prove that the relative error in estimates of  $P(W_y > b)$  using the extended rare-event simulation algorithm is uniformly bounded in b as b goes to infinity. With the aid of a recent heavy-traffic limit theorem, Theorem 3.2 of Whitt (2014), the algorithm also applies to compute the periodic steady-state distribution of (i) reflected periodic Brownian motion (RPBM) by considering appropriately scaled  $GI_t/GI/1$  models and (ii) a large class of general  $G_t/G/1$ queues by approximating by  $GI_t/GI/1$  models with the same heavy-traffic limit. As is shown in the theorem, the heavy-traffic limits of both the steady-state queue length and the steady-state virtual waiting time follow the RPBM distribution, whose explicit expressions are not available. Our work contributes by efficiently calculating its steady-state distribution. Simulation examples demonstrate the accuracy and efficiency of the algorithm for estimating both tail probabilities and moments of  $W_y$  in  $GI_t/GI/1$  queues, and steady-state distributions of RPBM. This chapter is an edited version of Ma and Whitt (2018b).

### 3.1 Introduction

For the steady-state performance of the stationary GI/GI/1 single-server queue with unlimited waiting room and service in order of arrival, we have effective algorithms, e.g., Abate et al. (1993), Asmussen (2003). We also have exact formulas in special cases and useful general approximation formulas in heavy traffic, e.g., Asmussen (2003), Whitt (2002). For the periodic steady-state performance of associated single-server queues, having periodic arrival-rate functions, there is much less available. There is supporting theory in Harrison and Lemoine (1977), Lemoine (1981, 1989), Rolski (1981, 1989a). On the algorithm side, there is a recent contribution on perfect sampling in Xiong et al. (2015). Of particular note is the paper on the periodic  $M_t/GI/1$  queue by Asmussen and Rolski (1994) that provides a theoretical basis for a rare-event simulation algorithm (although no algorithm is discussed there); also see  $\S$ VII.6 of Asmussen and Albrecher (2010) and Morales (2004). The goal there was to calculate ruin probabilities, but those are known to be equivalent to waiting-time and workload tail probabilities. A heavy-traffic limit for the periodic  $G_t/G/1$  queue was also established recently by Whitt (2014), which shows that the basic processes can be approximated by reflected periodic Brownian motion (RPBM), but so far there are no algorithms or simple formulas for RPBM.

In this chapter, we provide an effective algorithm to calculate the periodic steady-state distribution and moments of the remaining workload  $W_y$  at time yC within a cycle of length C,  $0 \le y < 1$ , in a single-server queue with a periodic arrival-rate function. The algorithm applies exactly to the  $M_t/GI/1$  model, where the arrival process is a nonhomogeneous Poisson process (NHPP), and any  $GI_t/GI/1$  model, where the arrival process is a time-transformation of an equilibrium renewal process. A new representation of  $W_y$  (in (3.1) below) makes it possible to apply a modification of the classic rare-event simulation for the stationary GI/GI/1 model exploiting importance sampling using an exponential change of measure, as in Ch. XIII of Asmussen (2003) and Ch. VI of Asmussen and Glynn (2007). We show that the algorithm is effective for estimating the mean and variance as well as small tail probabilities of the periodic steady-state workload.

The main example is the periodic  $M_t/GI/1$  queue, but our results go well beyond the periodic  $M_t/GI/1$  queue. By also treating the more general  $GI_t/GI/1$  queue, we are able to apply the algorithm to compute the steady-state distribution of the limiting RPBM in Whitt (2014). To cover the full range of parameters of the RPBM, we need the generalization to  $GI_t/GI/1$ . (In particular, this enables us to calculate the periodic steady-state distribution of the limiting RPBM for the  $GI_t/GI/1$  model in (3.48) and (3.52) for any variability parameter  $c_x$ .) As we will explain in §3.6.4, the algorithm for the  $GI_t/GI/1$  models, but we do not report simulation results for that extension here.

We report results from extensive simulation experiments for  $GI_t/GI/1$  models to demonstrate the effectiveness of the algorithm. Both the convergence to RPBM and the effectiveness of the algorithm for RPBM are demonstrated by displaying the results for a range of traffic intensities  $\rho$ approaching 1. This unity in the numerical results requires the nonstandard heavy-traffic scaling in Whitt (2014), which we review in §3.6. (In particular, the deterministic arrival-rate function is scaled as well as space and time; see (3.38).) The unity in the numerical results provided by the heavy-traffic scaling is in the same spirit as the scaling in the numerical results in Abate and Whitt (1998), Choudhury et al. (1997).

#### 3.1.1 Using Bounds to Connect to Stationary Methods

We are able to apply the familiar rare-event simulation for the GI/GI/1 model to the periodic  $GI_t/GI/1$  model because we can make strong connections between the given periodic  $GI_t/GI/1$  model and the associated GI/GI/1 model with the constant average arrival rate. In fact, this connection is largely achieved directly by construction, because we represent the periodic arrival counting process A as a deterministic time transformation of an underlying rate-1 counting process N by (2.1) and (2.2). This is a common representation when N is a rate-1 Poisson process; then A is an NHPP. For the  $G_t/G/1$  model, N is understood to be a rate-1 stationary point process. Hence, for the  $GI_t/GI/1$  model, N is an equilibrium renewal process with time between renewals having mean 1, which is a renewal process except the first inter-renewal time having the equilibrium

distribution. The representation in (2.2) also has been used for processes N more general than NHPP's by Massey and Whitt (1994), Gerhardt and Nelson (2009), Nelson and Gerhardt (2011), He et al. (2016), Ma and Whitt (2015a), Whitt (2015) and Li et al. (2016).

Given that we use representation (2.2), we show that it is possible to uniformly bound the difference between the cumulative arrival-rate function  $\Lambda$  and the associated linear cumulative arrival-rate function  $\bar{\lambda}e$  of the stationary model, where  $\bar{\lambda}$  is the average arrival rate and e is the identity function,  $e(t) \equiv t, t \geq 0$ . Consequently, we are able to bound the difference between the steady-state workloads W in the stationary G/G/1 model and  $W_y$  in the periodic  $G_t/G/1$  model.

#### 3.1.2 A Convenient Representation

We exploit the arrival process construction in (2.2) to obtain a convenient representation of the stationary workload  $W_y$  in terms of the underlying stationary process  $N \equiv \{N(t) : t \ge 0\}$  in (2.2) and the associated sequence of service times  $V \equiv \{V_k : k \ge 1\}$  via

$$W_y \stackrel{d}{=} \sup_{s \ge 0} \left\{ \sum_{k=1}^{N(s)} V_k - \tilde{\Lambda}_y^{-1}(s) \right\}, \quad 0 \le y < 1,$$
(3.1)

where

$$\tilde{\Lambda}_y(t) \equiv \Lambda(yC) - \Lambda(yC - t), \quad t \ge 0,$$
(3.2)

is the reverse-time cumulative arrival-rate function starting at time yC within the periodic cycle  $[0, C], 0 \le y < 1$ , and  $\tilde{\Lambda}_y^{-1}$  is its inverse function, which is well defined because  $\tilde{\Lambda}_y(t)$  is continuous and strictly increasing. Representation (3.1) is convenient because all stochastic dependence is captured by the first term within the supremum, while all deterministic time dependence is captured by the second term.

From the representation in (3.1), it is evident that from each sample path of the underlying stochastic process (N, V), we can generate a realization of  $W_y$  in (3.1) for any y in [0,1) by just changing the deterministic function  $\tilde{\Lambda}_y^{-1}$ . Moreover, from the rare-event construction in §3.4, we can simultaneously obtain an estimate of  $P(W_y > b)$  for all b in the bounded interval  $[0, b_0]$  with the same time complexity as for applying the estimation for the single value  $b_0$ . Thus, we can essentially obtain estimates for all *performance parameter pairs*  $(y, b) \in [0, 1) \times [0, b_0]$  in the process of doing the estimation for only one pair. This efficiency is very useful to conduct simulation studies to expose the way that  $P(W_y > b)$  and the other performance measures depend on (y, b).

#### 3.1.3 Stylized Sinusoidal Examples

We illustrate the rare-event simulation by showing simulation results for  $GI_t/GI/1$  queues with sinusoidal arrival-rate function (2.4) where  $\beta$ ,  $0 < \beta < 1$ , is the relative amplitude and the cycle length is  $C = 2\pi/\gamma$ . We let the mean service time be  $\mu^{-1} = 1$ , so that the average arrival rate is the traffic intensity, i.e.,  $\bar{\lambda} = \rho$ . With this scaling, we see that there is the fundamental *model parameter triple* ( $\rho, \beta, \gamma$ ) or, equivalently, ( $\rho, \beta, C$ ). The associated cumulative arrival-rate function is (2.6) and the associated reverse-time cumulative arrival-rate function defined in (3.2) is

$$\tilde{\Lambda}_{y}(t) = \rho \left( t + (\beta/\gamma) \left( \cos \left( \gamma(yC - t) \right) - \cos \left( \gamma yC \right) \right) \right), \quad t \ge 0.$$
(3.3)

We only consider the case  $\rho < 1$ , under which a proper steady-state exists under regularity conditions (which we do not discuss here). Behavior differs for short cycles and long cycles. There are two important cases for the relative amplitude: (i)  $0 < \beta < \rho^{-1} - 1$  and (ii)  $\rho^{-1} - 1 \le \beta \le 1$ . In the first case, we have  $\rho(t) < 1$  for all t, where  $\rho(t) \equiv \lambda(t)$  is the instantaneous traffic intensity, but in the second case we have intervals with  $\rho(t) \ge 1$ , where significant congestion can build up. If there is a long cycle as well, the system may be better understood from fluid and diffusion limits, as in Choudhury et al. (1997). (Tables 3.8 and 3.9 illustrate the significant performance difference for the mean  $E[W_y]$ .)

#### 3.1.4 Organization of the Chapter

We start in §3.2 by reviewing the reverse-time representation of the workload process, which leads to representation (3.1). In §3.3 we establish the bounds and associated asymptotic and approximations connecting the periodic model to the associated stationary model with the average arrival rate. In §3.4 we develop the simulation algorithm for the  $GI_t/GI/1$  model and establish theoretical results on its efficiency. We also discuss the computational complexity and running times. In §3.5 we present simulation examples. In §3.6 we review and extend the heavy-traffic FCLT in Theorem 3.2 of Whitt (2014), which explains the scaling that unifies our numerical results in the simulation experiments. in §3.6.4 we discuss the approximation for general periodic  $G_t/G/1$  models. In §3.7 we draw conclusions. Additional material is presented in the supplement Appendix A, including approximations for the important asymptotic decay rate and more simulation examples.

### **3.2** Reverse-Time Representation of the Workload Process

We consider the standard single-server queue with unlimited waiting space where customers are served in order of arrival. Let  $\{(\Delta A_k, V_k)\}$  be a sequence of ordered pairs of interarrival times and service times and  $\{A_k\}$  be the arrival times. (in §3.2 and in §3.3 we do not need to impose any *GI* conditions.) Let an arrival counting process be defined on the positive half line by  $A(t) \equiv$ max  $\{k \ge 1 : A_k \le t\}$  for  $t \ge A_1$  and  $A(t) \equiv 0$  for  $0 \le t < A_1$ , and let the total input of work over the interval [0, t] be the random sum

$$Y(t) \equiv \sum_{k=1}^{A(t)} V_k, \quad t \ge 0.$$
(3.4)

Then we can apply the reflection map to the net input process Y(t) - t to represent the workload (the remaining work in service time) at time t, starting empty at time 0, as

$$W(t) = Y(t) - t - \inf \{Y(s) - s : 0 \le s \le t\} = \sup \{Y(t) - Y(s) - (t - s) : 0 \le s \le t\}, \quad t \ge 0.$$

We now convert this standard representation to a simple supremum by using a reverse-time construction, as in Loynes (1962) and Chapter 6 in Sigman (1995). This is achieved by letting the interarrival times and service times be ordered in reverse time going backwards from time 0. Then  $\tilde{A}(t)$  counts the number of arrivals and  $\tilde{Y}(t)$  is the total input over the interval [-t, 0] for  $t \ge 0$ . With this reverse-time construction (interpretation), we can write

$$W(t) = \sup \{ \dot{Y}(s) - s : 0 \le s \le t \}, \quad t \ge 0,$$
(3.5)

and we have W(t) increasing to  $W(\infty) \equiv W$  with probability 1 (w.p.1) as  $t \uparrow \infty$ . In a stable stationary setting, under regularity conditions, we have  $P(W < \infty) = 1$ ; see §6.3 of Sigman (1995).

We now consider the periodic arrival-rate function  $\lambda(t)$  with cycle length C, average arrival rate  $\bar{\lambda} = \rho < 1$  and bounds  $0 < \lambda_L \leq \lambda(t) \leq \lambda_U < \infty$  for  $0 \leq t \leq C$ . As in (2.2), we can construct the arrival process A by transforming a general rate-1 stationary process N by the cumulative arrival-rate function. We let the service times  $V_k$  be a general stationary sequence with  $E[V_k] = 1$ .

We now exploit (3.5) in our more specific periodic  $G_t/G/1$  context. The workload at time yC

in the system starting empty at time yC - t can be represented as

$$W_{y}(t) = \sup_{0 \le s \le t} \{ \tilde{Y}_{y}(s) - s \}$$
  
$$\stackrel{d}{=} \sup_{0 \le s \le t} \{ \sum_{k=1}^{N(\tilde{\Lambda}_{y}(s))} V_{k} - s \}$$
  
$$= \sup_{0 \le s \le \tilde{\Lambda}_{y}(t)} \{ \sum_{k=1}^{N(s)} V_{k} - \tilde{\Lambda}_{y}^{-1}(s) \}, \qquad (3.6)$$

where  $\tilde{Y}_y$  is the reverse-time total input of work starting at time yC within the cycle of length C,  $\Lambda(t)$  is the cumulative arrival-rate function in (2.2),  $\tilde{\Lambda}_y(t)$  is the reverse-time cumulative arrivalrate function in (3.2) and  $\tilde{\Lambda}_y^{-1}$  is its inverse function. The second line equality in distribution holds when N is a stationary point process, which is a point process with stationary increments and a constant rate. In the  $GI_t/GI/1$  setting, N is an equilibrium renewal process and thus this regularity condition is satisfied. Let  $\{U_k, k \ge 1\}$  be interarrival times for the base process N. Note that in this specific setting,  $V_k$ 's are i.i.d. with distribution V, but  $U_1$  has equilibrium distribution  $U_e$ , which may be different from the i.i.d. distributions of  $U_k$ ,  $k \ge 2$  in (3.6). Just as  $W(t) \uparrow W$ w.p.1 as  $t \to \infty$ , so  $W_y(t) \uparrow W_y$  w.p.1 as  $t \to \infty$ , for  $W_y$  in (3.1).

Even though (3.6) is valid for all t, we think of the system starting empty at times -kC, for  $k \ge 0$ , so that we let yC - t = -kC or, equivalently, we stipulate that t = C(k+y),  $0 \le y < 1$ , and consider successive values of k and let  $k \to \infty$  to get (3.1). That makes (3.6) valid to describe the distribution of W(C(k+y)) for all  $k \ge 0$ . We think that (3.6) and (3.1) are new representations, but they can be related to various special cases in the literature.

## **3.3** Bounds and Approximations

We first bound the periodic system above and below by modifications of the corresponding stationary system with an arrival process that has the average arrival rate. Then we establish limits and introduce approximations. In doing so, we extend results in Asmussen and Rolski (1994).

#### 3.3.1 Basic Bounds

We now compare the periodic steady-state workload  $W_y$  in (3.1) and the associated stationary workload W defined as in (3.1) with  $\rho^{-1}s$  replacing  $\tilde{\Lambda}_y^{-1}(s)$ :

$$W \stackrel{d}{=} \sup_{s \ge 0} \Big\{ \sum_{k=1}^{N(s)} V_k - \rho^{-1} s \Big\},$$
(3.7)

Note that in both (3.1) and (3.7), N is understood to be a stationary point process. In particular, for the  $GI_t/GI/1$  model, N is an equilibrium renewal process with the first inter-renewal time having the equilibrum distribution, therefore W is the stationary workload in the associated GI/GI/1model, which may differ from the stationary waiting time in the same model. We now show that we can bound  $W_y$  above and below by a constant difference from the stationary workload W by rewriting (3.1) as

$$W_y = \sup_{s \ge 0} \left\{ \sum_{k=1}^{N(s)} V_k - \rho^{-1} s - (\tilde{\Lambda}_y^{-1}(s) - \rho^{-1} s) \right\}.$$
 (3.8)

From (3.8), we immediately obtain the following lemma.

**Lemma 3.3.1.** (upper and lower bounds on  $W_y$ ) For  $W_y$  in (3.1) and W in (3.7),

$$W_y^- \equiv W - \zeta_y^- \le W_y \le W - \zeta_y^+ \equiv W_y^+ \tag{3.9}$$

where

$$\zeta_y^- \equiv \sup_{0 \le s \le \rho C} \{ \tilde{\Lambda}_y^{-1}(s) - \rho^{-1}s \} \ge 0 \quad and \quad \zeta_y^+ \equiv \inf_{0 \le s \le \rho C} \{ \tilde{\Lambda}_y^{-1}(s) - \rho^{-1}s \} \le 0.$$
(3.10)

Note that the supremum and infimum in (3.10) are over the interval  $[0, \rho C]$ . Because the average arrival rate is  $\rho$ ,  $\tilde{\Lambda}_y(C) = \Lambda(C) = \rho C$  and thus  $\tilde{\Lambda}_y^{-1}(\rho C) = C$ . Given that  $\Lambda$  is continuous and strictly increasing, we can use properties of the inverse function as in §13.6 of Whitt (2002) to determine an alternative representation of the bounds in terms of the reverse-time cumulative arrival-rate function  $\tilde{\Lambda}_y$ . We emphasize that these bounds depend on y.

**Lemma 3.3.2.** (alternative representation of the bounds) The constants  $\zeta_y^-$  and  $\zeta_y^+$  can also be expressed as

$$\zeta_{y}^{-} = -\rho^{-1} \inf_{0 \le s \le C} \{ \tilde{\Lambda}_{y}(s) - \rho s \} \ge 0 \quad and \quad \zeta_{y}^{+} = -\rho^{-1} \sup_{0 \le s \le C} \{ \tilde{\Lambda}_{y}(s) - \rho s \} \le 0.$$
(3.11)

**Proof.** We use basic properties of inverse functions, as in §13.6 of Whitt (2002). First, note that, for any homeomorphism  $\phi$  on the interval [0, C],

$$\sup_{0 \le s \le C} \{\phi(s) - s\} = \sup_{0 \le s \le C} \{\phi(\phi^{-1}(s)) - \phi^{-1}(s)\} = \sup_{0 \le s \le C} \{s - \phi^{-1}(s)\} = -\inf_{0 \le s \le C} \{\phi^{-1}(s) - s\}.$$
(3.12)

To treat  $\zeta_y^-$  in (3.10), we apply (3.12) to  $\tilde{\Lambda}_y^{-1}$  after rescaling time to get

$$\sup_{0 \le s \le \rho C} \{ \tilde{\Lambda}_y^{-1}(s) - \rho^{-1} s \} = \sup_{0 \le u \le C} \{ \tilde{\Lambda}_y^{-1}(\rho u) - u \} = -\inf_{0 \le u \le C} \{ \rho^{-1} \tilde{\Lambda}_y(u) - u \}$$
$$= -\rho^{-1} \inf_{0 \le s \le C} \{ \tilde{\Lambda}_y(s) - \rho s \}.$$
(3.13)

In (3.13), the first equality is by making the change of variables  $u = \rho^{-1}s$ ; the second equality is by (3.12) plus Lemma 13.6.6 of Whitt (2002), i.e.,  $(\tilde{\Lambda}_y^{-1} \circ \rho e)^{-1} = (\rho^{-1}e \circ \tilde{\Lambda}_y) = \rho^{-1}\tilde{\Lambda}_y$ ; the third equality is obtained by multiplying and dividing by  $\rho$ .

We now combine the one-sided extrema into an expression for the absolute value.

Corollary 3.3.1. (single bound) As a consequence,

$$|W_{y} - W| \leq \zeta_{y} \equiv \max\{\zeta_{y}^{-}, -\zeta_{y}^{+}\} = \rho^{-1} \|\tilde{\Lambda}_{y} - \rho e\|_{C} \equiv \rho^{-1} \sup_{0 \leq s \leq C} \{|\tilde{\Lambda}_{y}(s) - \rho s|\} < \infty.$$
(3.14)

**Corollary 3.3.2.** (bounds in the sinusoidal case) For the sinusoidal case in (2.4), the bounds can be expressed explicitly as

$$\zeta_y^- = \frac{\beta(\cos\left(\gamma Cy\right) + 1)}{\gamma} \quad and \quad \zeta_y^+ = \frac{\beta(\cos\left(\gamma Cy\right) - 1)}{\gamma}.$$
(3.15)

**Proof.** By (3.3),

$$\tilde{\Lambda}_{y}(t) - \rho t = (\rho \beta / \gamma) \left( \cos \left( \gamma (Cy - t) \right) - \cos \left( \gamma Cy \right) \right), \quad t \ge 0,$$
(3.16)

from which (3.15) follows by choosing t to make  $\cos(\gamma(Cy-t)) = \pm 1$ .

#### 3.3.2 Tail Asymptotics

For many models, it is possible to obtain an approximation for the tail probability of W of the form

$$P(W > b) \approx A e^{-\theta^* b}, \quad b \ge 0, \tag{3.17}$$

based on the limit

$$\lim_{b \to \infty} e^{\theta^* b} P(W > b) = A.$$
(3.18)

For the GI/GI/1 model, the limit (3.18) is discussed in §XIII.5 of Asmussen (2003), where the random variable  $X_k \equiv V_k - \rho U_k$  is required to have a nonlattice distribution. However, the limit (3.18) also has been established for much more general models, allowing dependence among the interarrival times and service times; see Abate et al. (1994), Choudhury et al. (1996) and references therein. If indeed, the limit (3.18) holds for W, then we easily get corresponding bounds for  $W_y$ .

We remark that logarithmic asymptotics from Glynn and Whitt (1994) supports the weaker approximation

$$P(W_y > b) \approx P(W > b) \approx e^{-\theta^* b}, \quad b \ge 0.$$
(3.19)

The following corollary draws implications from the limit (3.17), from the bounds we have established, assuming that the limit (3.17) is valid.

**Corollary 3.3.3.** (tail-limit bounds) If  $e^{\theta^* b} P(W > b) \to A$  as  $b \to \infty$  for some  $\theta^* > 0$ , then

$$\limsup_{b \to \infty} e^{\theta^* b} P(W_y > b) \leq \lim_{b \to \infty} e^{\theta^* b} P(W > b + \zeta_y^+) = A_y^+ \equiv A e^{-\zeta_y^+ \theta^*} \quad and$$
$$\liminf_{b \to \infty} e^{\theta^* b} P(W_y > b) \geq \lim_{b \to \infty} e^{\theta^* b} P(W > b + \zeta_y^-) = A_y^- \equiv A e^{-\zeta_y^- \theta^*}.$$
(3.20)

as  $b \to \infty$ . If  $e^{\theta^* b} P(W_y > b) \to A_y$  as  $b \to \infty$ , then

$$A_y^- \le A_y \le A_y^+ \quad and \quad A_y^- \le A \le A_y^+. \tag{3.21}$$

For the GI/GI/1 model, we have the Cramer-Lundberg inequality for W in Theorem XIII.5.1 of Asmussen (2003), yielding  $P(W > b) \le e^{-\theta^* b}$  for all b.

**Corollary 3.3.4.** (periodic Cramer-Lundberg bound) For the periodic  $GI_t/GI/1$  model,

$$P(W_y > b) \le e^{-\theta^*(b + \zeta_y^+)} \quad for \ all \quad b > 0.$$

# 3.4 Simulation Methodology

We now apply the representation in (3.1) and the bounds in §3.3 to obtain an effective rare-event simulation method for the periodic  $GI_t/GI/1$  queueing model. Our approach is to first generate

exponentially tilted interarrival times and service times until a process involving them hits a given level b and then to calculate an estimate of tail probability using these generated values for each simulation replication. Hence, the algorithm is primarily deterministic calculations. We obtain estimates of statistical precision by performing a large number of independent replications.

#### 3.4.1 Exponential Tilting for the GI/GI/1 Model

We apply the familiar rare-event simulation method for the stationary GI/GI/1 model using importance sampling with an exponential change of measure, as in §XIII of Asmussen (2003) and §§V and VI of Asmussen and Glynn (2007). For the discrete-time waiting times in the GI/GI/1model based on  $\{(\rho^{-1}U_k, V_k)\}$ , where  $\{U_k\}$  and  $\{V_k\}$  are independent sequences of i.i.d. nonnegative mean-1 random variables, the key random variables are  $X_k(\rho) \equiv V_k - \rho^{-1}U_k$ . We assume that  $U_k$ ,  $V_k$  and thus  $X_k(\rho)$  have finite moment generating functions (mgf's)  $m_U(\theta)$ ,  $m_V(\theta)$ , and  $m_X(\theta) \equiv m_{X(\rho)}(\theta)$ , e.g.,  $m_V(\theta) \equiv E[e^{\theta V_k}]$ , and probability density functions (pdf's)  $f_U$ ,  $f_V$  and  $f_X \equiv f_{X(\rho)}$ . As usual, we define the exponential tilting pdf  $f_{X,\theta}(x) = e^{\theta x} f_X(x)/m_X(\theta)$  and for our simulation use the "optimal value"  $\theta^*$  such that  $m_X(\theta^*) = 1$ . That optimal tilting parameter coincides with the asymptotic decay rate  $\theta^*$  in Corollary 3.3.3.

There are several simplifications that facilitate implementation. First, as in Example XIII.1.4 of Asmussen (2003), we can construct the tilted pdf  $f_{X,\theta}(x)$  by constructing associated tilted pdf's of  $f_U$  and  $f_V$ , in particular, because  $X_k(\rho) \equiv V_k - \rho^{-1}U_k$ , it suffices to let  $f_{V,\theta}(x) = e^{\theta x} f_V(x)/m_V(\theta)$  and

$$f_{-U/\rho,\theta}(x) = \frac{e^{\theta x} f_{-U/\rho}(x)}{m_{-U/\rho}(\theta)} \quad \text{or} \quad \frac{e^{-\theta y/\rho} \rho f_U(y)}{m_U(-\theta/\rho)}$$
(3.22)

with the second expression obtained after making a change of variables, so that  $m_X(\theta) = m_V(\theta)m_U(-\theta/\rho)$ . We thus obtain the i.i.d. tilted random variables with pdf  $f_{X,\theta^*}(x)$  by simulating independent sequences of i.i.d. random variables with the pdf's  $f_{V,\theta^*}(x)$  and  $f_{-U/\rho,\theta^*}(x)$ .

Second, for all our examples, we consider common distributions that produce twisted pdf's having the same form as the original pdf's; it is only necessary to change the parameters. In particular, this property holds for the  $M, H_2, E_k$  and M+D distributions that we propose to exploit in §3.6.4. In particular, if V is a rate- $\mu$  exponential (M) random variable with pdf  $f_V(x) = \mu e^{-\mu x}$ , then  $f_{V,\theta}(x)$  is again an exponential random variable with parameter  $\mu - \theta$ , where we are required to have  $\mu > \theta > 0$ . Moreover, for the M/M/1 queue with arrival rate  $\lambda$  and service rate  $\mu$ , the associated optimal tilted parameters are  $\lambda_{\theta^*} = \mu$  and  $\mu_{\theta^*} = \lambda$ ; i.e., the optimal tilting just switches the arrival and service rates; see Example XIII.1.5 of Asmussen (2003).

If V has an  $H_2$  distribution with pdf  $f_V(x) = p\mu_1 e^{-\mu_1 x} + (1-p)\mu_2 e^{-\mu_2 x}$ , having parameter triple  $(p, \mu_1, \mu_2)$ , then  $f_{V,\theta}(x)$  again has an  $H_2$  distribution, but with a new parameter triple  $(p_{\theta}, \mu_{1,\theta}, \mu_{2,\theta})$ , where  $\mu_{j,\theta} = \mu_j - \theta$  and  $p_{\theta} = [p\mu_1/(\mu_1 - \theta)/\{[p\mu_1/(\mu_1 - \theta)] + [(1-p)\mu_2/(\mu_2 - \theta)]\}$ . We remark that the twisted  $H_2$  pdf does not inherit the balanced-means property of the original  $H_2$  pdf and has a different squared coefficient of variation (scv, variance divided by the square of the mean), but still  $c^2 > 1$ .

We now turn to the pdf's with scv  $c^2 < 1$ . First, a twisted  $E_k$  distribution is again  $E_k$ . More generally (because  $E_k$  is a special gamma distribution), if V has a gamma pdf  $f_V(x; \alpha, \mu) = \mu^{\alpha} x^{\alpha-1} e^{-\mu x} / \Gamma(\alpha)$ , then  $f_{V,\theta}(x)$  has a gamma pdf with parameter pair  $(\alpha_{\theta}, \mu_{\theta}) = (\alpha, \mu - \theta)$ ; see §V.1.b of Asmussen and Glynn (2007). Finally, if V is an M + D distribution with parameter pair  $(d, \mu - \theta)$ .

As a consequence, we can generate the tilted random variables in the standard way given underlying uniform random variables; e.g., we can apply the function  $h(x) = -\log(1-x)/\mu$  to a vector of uniform random variables to obtain the corresponding vector of exponential random variable with mean  $1/\mu$ . For each  $H_2$  random variable we can use two uniforms, one to select the exponential component and the other to generate the appropriate exponential; i.e., a random variable X with the  $H_2$  distribution having parameter triple  $(p, \mu_1, \mu_2)$  can be expressed in terms of the pair of i.i.d. uniforms  $(Z_1, Z_2)$  as

$$X = -((1/\mu_1)1_{\{Z_1 \le p\}} + (1/\mu_2)1_{\{Z_1 > p\}})\log(Z_2),$$
(3.23)

where  $1_A$  is the indicator variable with  $1_A = 1$  on the event A.

#### **3.4.2** Waiting Time in *GI/GI/1* Model

Let  $W^*$  denote the steady-state discrete-time waiting time, which coincides with the steady-state continuous-time workload W in the GI/GI/1 model for Poisson arrivals, but not otherwise. The heavy-traffic limits coincide, as can be seen from Theorem 9.3.4 of Whitt (2002).

The standard simulation for rare-event probability of large waiting times in the GI/GI/1 model is achieved by performing the change of measure using the tilted interarrival times and service times, as indicated in §3.4.1, where the tilting parameter  $\theta^*$  coincides with the asymptotic decay rate in §3.3.2, as described in Ch. XIII of Asmussen (2003) and §VI.2a of Asmussen and Glynn (2007).

To implement the simulation, we generate the random variables  $U_k$  and  $V_k$  from their tilted distributions with  $\theta^*$ . We estimate the tail probability of stationary waiting time  $P(W^* > b)$ by its representation as  $P(\tau_b^S < \infty)$ , where  $\tau_b^S$  is the first hitting time of  $S_n$  at level b, with  $S_n \equiv \sum_{k=1}^n X_k(\rho)$ . The tail probability can be expressed in terms of the stopped sum  $S_{\tau_b^S}$  using the underlying probability measure  $P_{\theta^*}$ . Note that  $S_{\tau_b^S} = b + Y(b)$ , where Y(b) is the overshoot of bby  $\{S_n\}$ , all under  $P_{\theta^*}$ . Under the new probability measure  $P_{\theta^*}$ ,  $S_n$  hits b with probability 1, so we only need to estimate the likelihood ratio. Thus the tail probability of the GI/GI/1 steady-state waiting time  $W^*$  can be expressed as

$$P(W^* > b) = P(\tau_b^S < \infty) = E_{\theta^*}[I\{\tau_b^S < \infty\}L_{\tau_b^S}(\theta^*)] = E_{\theta^*}[L_{\tau_b^S}(\theta^*)]$$
  
=  $E_{\theta^*}[m_X(\theta^*)^{\tau_b^S}e^{-\theta^*S_{\tau_b^S}}] = E_{\theta^*}[e^{-\theta^*S_{\tau_b^S}}] = e^{-\theta^*b}E_{\theta^*}[e^{-\theta^*Y_S(b)}],$  (3.24)

where  $L_{\tau_b^S}(\theta^*)$  is the likelihood ratio of  $\{X_k(\rho)\}_{1 \le k \le \tau_b^S}$  with respect to  $P_{\theta^*}$ . The second moment of this estimator is  $E_{\theta^*}[L_{\tau_b^S}(\theta^*)^2] = E_{\theta^*}[e^{-2\theta^*S_{\tau_b^S}}]$ . Theorem XIII.7.1 of Asmussen (2003) shows that the rare-event estimator of P(W > b) has relative error that is uniformly bounded in b as  $b \to \infty$ . (The proof of Theorem XIII.7.1 relies on Theorems XIII.5.1-3; the pdf assumption implies that X has a nonlattice distribution.)

#### **3.4.3** Workload in *GI/GI/1* Model

We are interested in the rare-event probability of large stationary workload W as in (3.7), where arrival process N is an equilibrium renewal process, because this is the process that we used to develop bounds of  $W_y$  in section 3.3. The classical exponential tilting method applies to simulating the rare-event probability of stationary waiting time  $W^*$  as reviewed in §3.4.2. The stationary waiting time is as in (3.7) with N being the renewal process without the exceptional first interrenewal time. To apply this exponential tilting method to stationary workload W, we need to make a slight modification of the algorithm above.

Now the equilibrium renewal process N has the exceptional first interarrival time and a constant rate  $\rho$ . We still use the usual partial sum process  $S_n \equiv \sum_{k=1}^n (V_k - \rho^{-1}U_k)$ , where  $V_k$  are still i.i.d with distribution V, but  $U_1$  has the equilibrium distribution of  $U_e$  and  $U_k$ ,  $k \ge 2$  are i.i.d with distribution U. We do the same tilting for all  $X_k(\rho)$ 's still using  $P_{\theta^*}$ , with  $dP_{\theta^*}(x) = [e^{\theta^* x}/m_X(\theta^*)]dP(x)$ . Note that  $\theta^*$  is solved from  $m_{X_k}(\theta^*) = 1$ , where  $k \ge 2$  and when k = 1, this equation may not hold. Now the likelihood ratio becomes

$$L_{\tau_{b}^{S}}(\theta^{*}) = m_{X_{1}}(\theta^{*}) \times m_{X_{2}}(\theta^{*}) \times \dots \times m_{X_{\tau_{b}^{S}}}(\theta^{*}) / (e^{\theta(X_{1}+X_{2}+\dots+X_{\tau_{b}^{S}})})$$
  
=  $m_{X_{1}}(\theta^{*})e^{-\theta S_{\tau_{b}^{S}}},$ 

where the second line follows because  $m_{X_k}(\theta^*) = 1, k \ge 2$ .

Therefore we need to add a constant multiplier  $m_{X_1}(\theta^*)$  to equation (3.24):

$$P(W > b) = P(\tau_{b}^{S} < \infty)$$
  
=  $E_{\theta^{*}}[L_{\tau_{b}^{S}}(\theta^{*})]$   
=  $E_{\theta^{*}}[m_{X_{1}}(\theta^{*})m_{X}(\theta^{*})^{\tau_{b}^{S}-1}e^{-\theta^{*}S_{\tau_{b}^{S}}}]$   
=  $E_{\theta^{*}}[m_{X_{1}}(\theta^{*})e^{-\theta^{*}S_{\tau_{b}^{S}}}]$   
=  $m_{X_{1}}(\theta^{*})e^{-\theta^{*}b}E_{\theta^{*}}[e^{-\theta^{*}Y_{S}(b)}].$  (3.25)

Note that (3.25) is also different from (3.24) in that the first  $X_1(\rho)$  in the partial sum  $S_{\tau_b^S}$  may have a different distribution from  $\{X_k(\rho), k \ge 2\}$ . The exact form of  $m_{X_1}(\theta^*)$  is as below

$$m_{X_1}(\theta^*) = E\{\exp\{\theta^* V - \theta^* \rho^{-1} U_e\}\}$$
  
=  $E\{\exp\{-\theta^* \rho^{-1} U_e\}\}/E\{\exp\{-\theta^* \rho^{-1} U\}\}.$ 

where the second line still follows from  $m_{X_k}(\theta^*) = 1$  and thus  $E\{\exp\{\theta^*V\}\} = 1/E\{\exp\{-\theta^*\rho^{-1}U\}\}$ .

Given that the estimator in (3.24) has bounded relative error as b goes to infinity, the estimator in (3.25) has bounded relative error as b goes to infinity as well. This is because when b is large, the first  $X_1$  does not influence the distribution of the overshoot  $Y_S(b)$  and thus  $Y_S(b)$  has the same distribution under  $P_{\theta^*}$  in both estimators.

Table 3.1 shows simulation estimates for the workload tail probabilities P(W > b) and the associated waiting-time tail probabilities  $P(W^* > b)$  using the algorithms in §3.4.3 and §3.4.2 respectively. In both cases, we refer to the estimates as  $P(W > b) \equiv \hat{p} = Ae^{-\theta^* b}$ , where  $\theta^*$  is common to both. We use a very small  $\rho = 0.1$  here so that workload and waiting time probabilities are very different. These numerical results match the exact values of  $\hat{p}$  and A calculated from Theorem X.5.1 of Asmussen (2003).
**Table 3.1:** Comparison of the steady-state workload and waiting-time tail probabilities for b = 4,20 in the stationary  $H_2/M/1$  queue with  $\rho = 0.1$ . The exact values are calculated from Theorem X.5.1 of Asmussen (2003).

	workload	waiting time	workload	waiting time
ρ	0.1	0.1	0.1	0.1
$\theta^*$	0.8690	0.8690	0.8690	0.8690
exact $A$	0.1	0.1310	0.1	0.1310
exact $p$	0.003093	0.004050	2.83E-09	3.70E-09
b	4	4	20	20
$\hat{p}$	0.003104	0.004055	2.84E-09	3.69E-09
$e^{-\theta^*b}$	0.0309	0.0309	2.83E-08	2.83E-08
A	0.1004	0.1311	0.1004	0.1305
s.e.	2.73E-05	3.55E-05	2.49E-11	3.25E-11
$\%95~{ m CI}~{ m lb}$	0.003050	0.003985	2.79E-09	3.63E-09
$\%95~{\rm CI}$ ub	0.003157	0.004125	2.89E-09	3.76E-09
r.e.	0.008788	0.008765	0.008771	0.008792

### 3.4.4 Applying the Bounds to the Periodic Case

From (3.1), we see that any positive b must be hit for the first time at an arrival time. Thus, we have the alternative discrete-time representation

$$W_y = \sup_{n \ge 0} \Big\{ \sum_{k=1}^n V_k - \tilde{\Lambda}_y^{-1}(N^{-1}(n)) \Big\} = \sup_{n \ge 0} \Big\{ \sum_{k=1}^n V_k - \tilde{\Lambda}_y^{-1}(\sum_{k=1}^n U_k) \Big\},$$
(3.26)

where  $U_k$  is the  $k^{\text{th}}$  interarrival time in the equilibrium renewal process N, i.e.  $U_1$  assumes the equilibrium distribution  $U_e$  while  $\{U_k, k \ge 2\}$  are i.i.d. with distribution U.

For the periodic  $GI_t/GI/1$  model with  $\bar{\lambda} = \rho$ , we can apply a variant of the exponential change of measure for the waiting times in the GI/GI/1 model in §3.4.1 above. We use the underlying measure  $P_{\theta^*}$  determined for GI/GI/1. we use the usual partial sum process  $S_n \equiv \sum_{k=1}^n X_k(\rho)$  for GI/GI/1 and the associated process

$$R_n \equiv \sum_{k=1}^n V_k - \tilde{\Lambda}_y^{-1} (\sum_{k=1}^n U_k).$$
 (3.27)

We estimate the tail probability  $P(W_y > b)$  by its representation as  $P(\tau_b^R < \infty)$ , where  $\tau_b^R$  is the first hitting time of  $R_n$  at level b. Under the new probability measure,  $R_n$  hits b with probability 1, so we only need to estimate the likelihood ratio. We still twist  $X_k(\rho) = V_k - \rho^{-1}U_k$  in the same way,

which is equivalent to twisting  $V_k$  and  $\rho^{-1}U_k$  separately, as discussed in §3.4.1. Then the likelihood ratio for  $\{X_k(\rho) : 1 \le k \le n\}$  is the same as before, i.e.,  $L_n(\theta) = m_{X_1}(\theta)m_X(\theta)^{(n-1)}e^{-S_n}$ . As a consequence, we obtain the representation

$$P(W_{y} > b) = P(\tau_{b}^{R} < \infty) = E_{\theta^{*}}[L_{\tau_{b}^{R}}(\theta^{*})]$$
  
=  $E_{\theta^{*}}[m_{X_{1}}(\theta^{*})m_{X}(\theta^{*})^{(\tau_{b}^{R}-1)}e^{-\theta^{*}S_{\tau_{b}^{R}}}] = m_{X_{1}}(\theta^{*})E_{\theta^{*}}[e^{-\theta^{*}S_{\tau_{b}^{R}}}].$  (3.28)

Still note that the first  $X_1(\rho)$  in the partial sum  $S_{\tau_k^R}$  has a different distribution from  $\{X_k, k \ge 2\}$ .

At first glance, (3.28) does not look so useful, because the random sum  $S_{\tau_b^R}$  involves the hitting time  $\tau_b^R$  for  $\{R_n\}$  instead of  $\{S_n\}$ , but we can shift the focus to  $R_{\tau_b^R}$  because we can bound the difference between  $S_{\tau_b^R}$  and  $R_{\tau_b^R}$ .

Lemma 3.4.1. (bound on difference of random sums) Under the assumptions above,

$$|S_{\tau_b^R} - R_{\tau_b^R}| \le \zeta_y \equiv \max\{|\zeta_y^+|, \zeta_y^-\},\tag{3.29}$$

where  $\zeta_y^+$  and  $\zeta_y^-$  are the one-sided bounds in (3.10) and (3.11). In addition,  $\tau_{b-\zeta}^S \leq \tau_b^R \leq \tau_{b+\zeta}^S$ .

**Proof.** The bound in (3.29) follows immediately from (3.10) and (3.11), because

$$|R_n - S_n| = \left| \left( \sum_{k=1}^n V_k - \tilde{\Lambda}_y^{-1} \sum_{k=1}^n U_k \right) - \left( \sum_{k=1}^n V_k - \sum_{k=1}^n \rho^{-1} U_k \right) \right| \le \zeta_y \equiv \max\{|\zeta_y^+|, \zeta_y^-\}$$
(3.30)

for all  $n \ge 1$ , where  $\zeta_y^+$  and  $\zeta_y^-$  are the one-sided bounds in (3.10) and (3.11).

Lemma 3.4.1 allows us to focus on  $R_{\tau_b^R}$ , where  $\tau_b^R$  is the hitting time for  $\{R_n\}$ . To do so, we impose an additional regularity condition. The regularity condition requires the excess service-time distribution in probability measure  $P_{\theta^*}$  be bounded above in stochastic order by a proper cdf, i.e.,

$$P_{\theta^*}(V > t + x | V > t) \equiv \frac{P_{\theta^*}(V > t + x)}{P_{\theta^*}(V > t)} \le G^c(x) \quad \text{for all} \quad t \ge 0,$$
(3.31)

where  $G^c(x) \equiv 1 - G(x) \to 0$  as  $x \to \infty$ . For example, it suffices for the service time to be bounded. It also suffices for the service-time distribution to have an exponential tail, which holds if there is a constant  $\eta > 0$  such that

$$e^{\eta x} P_{\theta^*}(V > x) \to L, \quad 0 < L < \infty \quad \text{as} \quad x \to \infty.$$
 (3.32)

If (3.32) holds, then

$$\frac{e^{\eta(t+x)}P_{\theta^*}(V>t+x)}{e^{\eta t}P_{\theta^*}(V>t)} \to 1 \quad \text{as} \quad t \to \infty,$$
(3.33)

so that (3.31) holds asymptotically with  $G^c(x) \equiv e^{-\eta x}$ . It holds over any bounded interval because the ratio is continuous and bounded, given (3.32). Of course, condition (3.31) would not hold if  $x^p P_{\theta^*}(V > x) \to L$  as  $x \to \infty$  for  $0 < L < \infty$  and p > 0.

**Theorem 3.4.1.** (bounded relative error) The rare-event simulation algorithm for the tail probability  $P(W_y > b)$  in the periodic  $GI_t/GI/1$  queue is unbiased and, if the service-time distribution satisfies condition (3.31), then the rare-event simulation algorithm produces relative error that is uniformly bounded in b, just as for the stationary GI/GI/1 model, provided that the conditions for the rare-event simulation in the GI/GI/1 model are imposed so that the estimates are unbiased with bounded relative error.

**Proof.** The unbiasedness follows from (3.28). Lemma 3.4.1 allows us to focus on  $R_{\tau_b^R}$ . The remaining result parallels Theorem XIII.7.1 in Asmussen (2003) for the GI/GI/1 model, which draws on Theorems XIII.5.1-3. Just as  $S_{\tau_b^S} = b + Y_S(b)$ , where  $Y_S(b)$  is the overshoot of b upon first passage to b in the random walk  $\{S_n\}$ , so is  $R_{\tau_b^R} = b + Y_R(b)$ , where  $Y_R(b)$  is the overshoot of b upon first passage to b in the sequence  $\{R_n\}$ . The results for the stationary case are based on the well developed theory for that overshoot, which depend on the random walk structure. In contrast, less is known for  $\{R_n\}$ . However, we do see from (3.26) that the overshoot can be regarded as an excess-distribution of the last service time. Thus, under the extra condition (3.31), we can again apply the proof in Asmussen (2003), using

$$e^{-k\theta^*b} \ge E_{\theta^*}[e^{-k\theta^*R_{\tau_b^R}}] \ge e^{-k\theta^*b}E_{\theta^*}[e^{-k\theta^*Y_R(b)}] \ge ce^{-k\theta^*b}$$

for 0 < c < 1, where  $c = E[e^{-k\theta^*Z}]$ ,  $P(Z > x) = G^c(x)$ ,  $x \ge 0$ , and k is a positive integer.

### 3.4.5 The Mean and Variance

We now show how tail-integral representations of the mean and higher moments on p. 150 of Feller (1971) can be exploited to obtain corresponding simulations of these related quantities using our rare-event simulation algorithm. Recall that, for any nonnegative random variable X, the mean can be expressed as

$$E[X] = \int_0^\infty P(X > t) \, dt,$$
 (3.34)

while the corresponding representation of the  $p^{\text{th}}$  moment for any p > 1 is

$$E[X^{p}] = \int_{0}^{\infty} pt^{p-1} P(X > t) dt.$$
(3.35)

To obtain a finite algorithm, it is natural to approximate the integrals for the mean and the second moment by finite sums plus a tail approximation, i.e.,

$$E[W_y] \approx \sum_{k=0}^n (P(W_y > k\delta)\delta) + \frac{P(W_y > n\delta)}{\theta^*}$$
$$E[W_y^2] \approx \sum_{k=0}^n (2P(W_y > k\delta)k\delta) + 2P(W_y > n\delta)(\frac{n\delta}{\theta^*} + \frac{1}{(\theta^*)^2}).$$
(3.36)

In each case, the second term is based on applying the tail integral formula over  $[n\delta, \infty)$  with the approximation

$$P(W_y > n\delta + x) \approx P(W_y > n\delta)e^{-\theta^* x}$$
(3.37)

and integrating.

To understand how to choose the discretization parameter  $\delta$  in (3.36), suppose that  $P(W > t) = ae^{-\theta^* t}$ . In that case, the infinite sum for the mean can be expressed as

$$\sum_{k=0}^{\infty} a\delta e^{-\theta^*k\delta} = \frac{a}{\theta^*} \left( 1 + \theta^* \frac{\delta}{2} + O(\delta^2) \right) \quad \text{as} \quad \delta \downarrow 0,$$

so that the relative error for the mean is  $\theta^*(\delta/2) + O(\delta^2)$ . Similarly, the corresponding calculation for the second moment indicates an asymptotic relative error proportional to  $\theta^*\delta$ . The subsequent truncation approximations involving *n* imposes no additional error, provided that the tail is exponential, which is likely to hold in view of §3.3.2. Thus, the truncation is good provided that approximation (3.37) is good, which can be checked with the algorithm.

In closing, we remark that because  $\theta^*(\rho)$  tends to be of order  $1-\rho$  as  $\rho \uparrow 1$ , as explained in §A.2.2 of Appendix A, we can maintain fixed relative error in the discretization if we let  $\delta$  be inversely proportional to  $1-\rho$  or  $\theta^*(\rho)$  as  $\rho \uparrow 1$ . That can be useful because otherwise the computational complexity increases as  $\rho$  increases, as we show in the next sections. We illustrate letting  $\delta$  increase with increasing  $\rho$  in Table 3.10.

### 3.4.6 The Algorithm

This exponential tilting algorithm to estimate tail probabilities  $P(W_y > b)$  in the  $GI_t/GI/1$  queue is based on equation (3.28) with the following steps. (We elaborate on Steps 4 and 5 in refapp.) Without loss of generality, we assume service rate is  $\mu = 1$  and thus  $\overline{\lambda} = \rho$ .

Step 1. Before we conduct the simulation, we first construct a table of the inverse cumulative arrival-rate function  $\rho \tilde{\Lambda}_y^{-1}$ , i.e., the inverse of the reverse-time cumulative arrival-rate function  $\tilde{\Lambda}_y$  in (3.2) scaled by  $\rho$ , for any time yC in the cycle to be considered. For that purpose, we use Algorithm 1 developed in Chapter 2. That algorithm constructs an approximation  $J_y$  to the inverse function  $\rho \tilde{\Lambda}_y^{-1}$  for one cycle from the interval [0, C] to the interval [0, C]. This table is the same for a fixed y no matter what value  $\rho$  takes, which will be used for efficiently calculating  $\tilde{\Lambda}_y^{-1}$  later. The computational complexity has shown to be of order  $O(C/\epsilon)$ , where C is the length of a cycle of the periodic arrival-rate function and  $\epsilon$  is an allowed error tolerance.

Step 2. Again, before we conduct the simulation, we determine the required number of **partial sums** needed in each replication, which we denote by  $n_s$ . Note that we need this step because Matlab is much faster in vector operations than in loops. However, if another software is used to implement this algorithm, we can skip this step and generate exponentially twisted service times and interarrival times one by one in a loop until the hitting time  $\tau_b^R$  is reached. Given the largest b under consideration, we estimate the expected number by  $m_s \equiv b/E_{\theta^*}[V_k - \rho^{-1}U_k]$  by approximating the sum by Brownian motion which is asymptotically correct as b gets large, e.g. by  $\S5.7.5$  of Whitt (2002). If we use a Brownian motion approximation for the random walk, then we can get the approximate mean and variance by applying Theorems 5.7.13 and 5.7.9 of Whitt (2002). For the canonical Brownian motion in Theorem 5.7.13, the variance of the first passage time is equal to the mean, but in general the ratio of the variance to the mean is proportional to the scv  $c_X^2 \equiv Var(X)/E[X]^2$ . Hence, we use  $n_s = \max\{C, Lm_s\}$ , where C is a minimum number like 100 and L is a safety-factor multiplier to account for the stochastic variability, which might be taken to be simply 10, but could be constructed more carefully. The largest value of b will depend on the case. If we want to treat multiple cases at once for simulation efficiency, we need to determine the largest required value of  $n_s$ . If  $m_s$  is large, then it is natural to use  $n_s = m_s + 5\sqrt{c_X^2 m_s}$  instead of  $n_s = 10m_s$ , because then  $5\sqrt{c_X^2 m_s}$  is about 5 standard deviations, which should be sufficient, and beneficial if  $5\sqrt{c_X^2 m_s} \ll (L-1)m_s$ .

Step 3. As the first part of the actual stochastic simulation, for each replication we now generate the required random vectors of tilted interarrival times and service times; For each replication, generate  $\tilde{V} \equiv (V_1, ..., V_n)$  and  $\rho^{-1}\tilde{U} \equiv (\rho^{-1}U_1, ..., \rho^{-1}U_n)$  where  $n = n_s$  from step 2 above,  $V_k$  are i.i.d. random variables from  $F_V^{\theta^*}$ , the exponentially tilted distribution of  $V_k$  with parameter  $\theta^*$  and  $\rho^{-1}U_k$  i.i.d. from  $F_{\rho^{-1}U}^{-\theta^*}$ , the exponentially tilted distribution of  $\rho^{-1}U_k$  with parameter  $-\theta^*$ . The distributions of  $V_k$  and  $U_k$  under the tilted probability measure  $P_{\theta^*}$  were discussed in §3.4.1.

**Step 4.** Using vector operations, we calculate the associated vectors of partial sums and transformed partial sums. Use Algorithm 2 in Chapter 2 to calculate the time-transformed arrival times.

Step 5. Use (3.28) to calculate the tail probability  $\mathbf{P}(\mathbf{W}_{\mathbf{y}} > \mathbf{b})$ . If  $n_s$  is not large enough to reach hitting times  $\tau_b^R$ , we repeat Step 3 to generate additional vectors of  $\tilde{V}$  and  $\rho^{-1}\tilde{U}$  and repeat Step 4 to calculate additional partial sums and transformed partial sums. We treat the cases of the tail probability for a single value of b differently from multiple values of b, as required when we estimate moments. For multiple values of b, we use one loop to find all stopping times at each element of the vector b.

Step 6. We run the algorithm for N i.i.d. replications. Estimate  $P(W_y > b)$ ,  $EW_y$  and  $EW_y^2$  by the sample averages over the N replications. We estimate the associated confidence intervals in the usual way, using the Gaussian distribution if N is large enough and the Student-t distribution otherwise.

In conclusion, we point out that there is flexibility in the order of the steps specified above. We can re-use random variables if we generate the random vectors in an early step. We can avoid storage problems if we perform calculations for each replication separately. As usual, there is a tradeoff in storage requirements and computation efficiency.

### 3.4.7 Computational Complexity and Running Times

We implemented the algorithm using matlab on a desktop computer. All examples were for the sinusoidal arrival-rate function  $\lambda$  in (2.4) with associated reverse-time cumulative arrival-rate func-

We now specify the **computational complexity of the algorithm** above. Given the inverse function table for  $\tilde{\Lambda}_y^{-1}$  computed in advance using the algorithm in Chapter 2, the remaining algorithm has an approximate linear computational complexity of  $O(b/E_{\theta^*}[V_k - \rho^{-1}U_k])$ , Specifically for the  $M_t/M/1$  model, the computational complexity is  $O(b\rho/(1-\rho))$ , being directly proportional to b and inversely proportional to  $1 - \rho$ . This can be made precise as  $b \uparrow \infty$  or as  $\rho \uparrow 1$ , and presumably in some joint limit as  $b/(1-\rho)\uparrow$ , but we do not do that here. For b large or for  $\rho$  large, we can perform asymptotics to make the following approximations valid.

The hitting time  $\tau_b$  of the random walk  $S_n$  as defined in (3.27) has expectation  $E(\tau_b) = b/(E_{\theta^*}(V_k - \rho^{-1}U_k))$  by approximating  $S_n$  by a Brownian motion, for b that is very large compared to the step size of the random walk. Now consider the hitting time  $\tau_b$  of  $R_n$  as defined in (3.27). Since the average arrival rate  $\bar{\lambda} = \rho$ , the expected value of this hitting time is approximately the same as that for  $S_n$ .

When both  $V_k$  and  $\rho^{-1}U_k$  are exponential random variables with rates 1 and  $\rho$  respectively, under the new measure  $\theta^*$ , they are still exponential with rates  $\rho$  and 1 respectively. Thus  $b/E_{\theta^*}(V_k - \rho^{-1}U_k) = b/(1/\rho - 1) = b\rho/(1 - \rho).$ 

It can be advantageous to estimate the tail probabilities  $P(W_y > b)$  for multiple values of b simultaneously. This can be done for each b by keeping track of the passage times for them while considering the largest value of b. This is very useful when we want to plot the cdf or its probability density function (pdf), or when we want to calculate the mean.

We now describe our experiments with **running times** on a desktop computer. Before conducting the simulation, we did step 1, constructing the table of the inverse function  $\rho \tilde{\Lambda}_y^{-1}$  in one cycle, which takes computational time of  $O(C/\epsilon) = O(1/\gamma\epsilon)$  by Theorem 3.6.1, where C is the cycle length of the arrival rate function,  $\gamma$  is the parameter in the sinusoidal arrival-rate function and  $\epsilon$  is the error bound we choose for the inverse function table. The longest cycle we consider has  $\gamma = 0.00025$  (for (3.39) with  $\rho = 0.99$ ), or C = 25, 120. For  $\epsilon = 10^{-4}$ , it took 0.08 seconds to form the table needed for a single value of y.

In each replication, we can quickly determine the required length of the random variable vector, generate the vectors of random variables and calculate the partial sums, which are steps 2 to 4. The

most time is required for step 5, searching for the stopping time for one b, or for all stopping times for a long vector of b. When we do the search for one b, the computational time is  $O(b/(E_{\theta^*}[V_k - \rho^{-1}U_k]))$ , which is the approximate expected stopping time. When we do this for a long vector of b, we use a big loop which takes time linear in the maximum stopping time and the length of vector b, i.e.,  $O(\max(b)/(E_{\theta^*}[V_k - \rho^{-1}U_k] + n_b))$ , where  $n_b$  is the length of vector b. Specifically, for the  $M_t/M/1$ queue, the computational times are  $O(b\rho/(1 - \rho))$  and  $O(\max(b)\rho/(1 - \rho) + n_b))$  respectively. For example, in  $M_t/M/1$  queue, when  $\rho = 0.8$ , we choose  $\max(b) = \log(1000)/\theta^* = \log(1000)/(1 - \rho)$ ,  $\delta = 0.0002/(1 - \rho)$ , then maximum stopping time  $O(\max(b)\rho/(1 - \rho)))$  is negligible compared to the length of the vector b. The first part of time increases as  $\rho$  increases while the second part does not depend on  $\rho$  as both the largest b and  $\delta$  are inversely proportional to  $(1 - \rho)$ . In this case, when we did 40,000 replications, the run time was 127 seconds on the desktop to find all stopping times, whereas it took about 10 seconds to find one stopping time for the largest b.

### 3.5 Simulation Examples

We now give examples to illustrate the new simulation algorithm. All our examples are for the sinusoidal arrival-rate function in (2.4) with parameter triple  $(\bar{\lambda}, \beta, \gamma)$ . More results appear in the online supplement.

### 3.5.1 Tail Probabilities

We start by illustrating the efficiency of the rare-event simulation estimator of the tail probability  $P(W_y > b)$ , which gets exponentially small as b increases, and thus is prohibitively hard to estimate accurately by direct simulation. Table 3.2 shows that the relative errors of simulation estimates of  $P(W_y > b)$  for the  $M_t/M/1$  model in several cases are approximately independent of b. That property is held in all models considered.

In particular, Table 3.2 shows estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  and the components  $A_y$  and  $e^{-\theta^* b}$  for the special case y = 0.0 based on 5000 i.i.d. replications. Table 3.2 also shows estimates of the standard error (s.e.) of  $\hat{p}$ , the upper and lower bounds of the 95% confidence interval (CI), and the relative error (r.e.), which is the s.e. divided by the estimate of the mean. For Table 3.2, we used the arrival-rate function (2.4) with  $\bar{\lambda} = 1$ , and  $E[V_1] = 0.8$ , so that  $\rho = 0.8$ . We let

 $\beta = 0.2$  and consider three values of  $\gamma$ : 10, 1 and 0.1, making cycle lengths of 0.628, 6.28 and 62.8. The rapid fluctuation with  $\gamma = 10$  makes the arrival process very similar to a homogeneous Poisson process, because the cumulative arrival-rate function approaches a linear function; see Theorem VIII.4.10 in Jacod and Shiryaev (1987), Problem 1 on p. 360 of Ethier and Kurtz (1986) and Whitt (2016b). We also simulated the M/M/1 model with  $\beta = 0$  to verify simulation correctness.

Table 3.2 shows that the algorithm is effective for estimating  $P(W_0 > b)$ , because the relative error is approximately independent of b for each  $\gamma$ , ranging from about 0.0029 for  $\gamma = 10$  to about 0.0055 for  $\gamma = 0.1$ .

**Table 3.2:** Estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model with sinusoidal arrivalrate function in (2.4) as a function of  $\gamma$  and b for:  $\rho = 0.8$ ,  $\bar{\lambda} = 1$ ,  $\mu = 1.25$  and  $\beta = 0.2$  based on 5000 replications.

	b	$\hat{p}$	$exp(-\theta^*b)$	$A_0(b)$	s.e.	95% CI (lb)	(ub)	r.e.
$\gamma = 10$	10	0.0654	0.0821	0.797	1.87E-04	0.0651	0.0658	0.00286
	20	0.00537	0.00674	0.797	1.55E-05	0.00534	0.00540	0.00289
	40	3.61E-05	4.54E-05	0.795	1.05E-07	3.59E-05	3.63E-05	0.00290
	80	1.64E-09	2.06E-09	0.796	4.82E-12	1.63E-09	1.65E-09	0.00294
$\gamma = 1$	10	0.0628	0.0821	0.765	1.87E-04	0.0624	0.0632	0.00298
	20	0.00516	0.00674	0.766	1.51E-05	0.00513	0.00519	0.00292
	40	3.49E-05	4.54E-05	0.769	1.00E-07	3.47E-05	3.51E-05	0.00287
	80	1.58E-09	2.06E-09	0.767	4.65 E- 12	1.57 E-09	1.59E-09	0.00294
$\gamma = 0.1$	10	0.0413	0.0821	0.503	2.33E-04	0.0409	0.0418	0.00565
	20	0.00360	0.00674	0.535	1.98E-05	0.00356	0.00364	0.00550
	40	2.50E-05	4.54E-05	0.551	1.37E-07	2.47 E-05	2.53E-05	0.00548
	80	1.12E-09	2.06E-09	0.545	6.20E-12	1.11E-09	1.14E-09	0.00552

### 3.5.2 Heavy-Traffic Scaling

We produce unified numerical results by exploiting heavy-traffic scaling. In particular, we scale the arrival rate function so that the performance measures have heavy-traffic limits as  $\rho \uparrow 1$ , which we explain in §3.6. In the special case of (2.4), we consider an arrival-rate function scaled by the overall traffic intensity  $\rho$ , specifically,

$$\lambda_{\rho}(t) = \rho + (1 - \rho)\rho\beta \sin{(\gamma(1 - \rho)^2 t)}, \quad t \ge 0,$$
(3.38)

so that the cycle length in model  $\rho$  is  $C_{\rho} = C^*(1-\rho)^{-2} = 2\pi/(\gamma(1-\rho)^2)$ . Before scaling, the cycle length is  $C^* = 2\pi/\gamma$ .

When we consider the periodic steady-state workload, we include spatial scaling by  $1 - \rho$  Hence, to have asymptotically convergent models, we should choose parameter four-tuples  $(\bar{\lambda}_{\rho}, \beta_{\rho}, \gamma_{\rho}, b_{\rho})$ indexed by  $\rho$ , where

$$(\bar{\lambda}_{\rho}, \beta_{\rho}, \gamma_{\rho}, b_{\rho}) = (\rho, (1-\rho)\beta, (1-\rho)^{2}\gamma, (1-\rho)^{-1}b),$$
(3.39)

where  $(\beta, \gamma, b)$  is a feasible base triple of positive constants with  $\beta < 1$ . (We must constrain  $\beta_{\rho} \leq 1$ so that  $\lambda_{\rho}(t) \geq 0$  for all t.) Hence, we have the  $\rho$ -dependent constraint  $\rho_b = (1 - \rho)\beta \leq 1$ . There is no problem if  $\beta \leq 1$ , but we may want to consider  $\beta > 1$ . In that case,  $\beta_{\rho}$  is only well defined for  $\rho \geq 1 - (1/\beta)$ . For example, if  $\beta = 5.0$ , then we require that  $\rho \geq 0.8$ .

**Example 3.5.1.** (Using  $M_t/M/1$  to estimate the performance of RPBM)

To illustrate how we can apply simulations of the  $M_t/M/1$  model with increasing traffic intensities, let the base parameter triple be  $(\beta, \gamma, b) = (1.0, 2.5, 4.0)$ . Then the parameter 4-tuple for  $\rho = 0.8$  is

$$(\bar{\lambda}_{\rho}, \beta_{\rho}, \gamma_{\rho}, b_{\rho}) = (0.8, (1 - 0.8)\beta, (1 - 0.8)^2\gamma, (1 - 0.8)^{-1}b) = (0.8, 0.2, 0.1, 20.0).$$
(3.40)

The associated parameter 4-tuple for  $\rho = 0.9$  is (0.90, 0.10, 0.025, 40.00).

Let W be the steady-state workload in the stationary M/M/1 model with the same scaling, which has an exponential distribution except for an atom  $1 - \rho$  at the origin. Table 3.3 shows estimates of the ratio  $P(W_y > b_{\rho})/P(W > b_{\rho})$  for 5 different values of  $1 - \rho$ , where we successively divide  $1 - \rho$ by 2, and 8 different values of the position y within the cycle in the  $M_t/M/1$  model with sinusoidal arrival-rate function in (3.38) with the parameter 4-tuple in (3.39) using the base parameter triple  $(\beta, \gamma, b) = (1.0, 2.5, 4.0)$ . (The parameter 4-tuples for  $\rho = 0.8$  and  $\rho = 0.9$  are shown above.)

Table 3.3 shows that, for each fixed y, all estimates as a function of  $\rho$  serve as reasonable practical approximations for the others as well as for the RPBM limit developed in §3.6. The convergence in Table 3.3 is summarized by showing the average difference, average absolute difference and root mean square error (rmse) of the entry with the corresponding estimate for  $\rho = 0.99$  in the final column, taken over 40 evenly spaced values of y in the interval [0, 1).

**Table 3.3:** Comparison of the ratios  $P(W_y > b_\rho)/P(W > b_\rho)$ , where W is for the stationary model, for 5 different values of  $1 - \rho$  and 8 different values of the position y within the cycle in the  $M_t/M/1$  model with sinusoidal arrival-rate function in (3.38) with the parameter 4-tuple in (3.39) using the base parameter triple ( $\beta, \gamma, b$ ) = (1.0, 2.5, 4.0).

У	$1 - \rho = 0.16$	$1-\rho=0.08$	$1-\rho=0.04$	$1-\rho=0.02$	$1-\rho=0.01$
0.000	0.96364	0.96523	0.96424	0.96357	0.96344
0.125	0.97619	0.97686	0.97504	0.97493	0.97482
0.250	1.00456	1.00450	1.00255	1.00251	1.00305
0.375	1.03278	1.03264	1.03035	1.03152	1.03152
0.500	1.04565	1.04470	1.04278	1.04346	1.04405
0.625	1.03213	1.03096	1.03230	1.03150	1.03204
0.750	1.00225	1.00404	1.00425	1.00277	1.00241
0.875	0.97371	0.97696	0.97629	0.97457	0.97545
avg diff	0.00037	0.00112	0.00015	-0.00019	
avg. abs. dif	0.00099	0.00121	0.00081	0.00039	
rmse	0.00116	0.00134	0.00096	0.00049	

### 3.5.3 Hyperexponential Examples

We now present results from simulation experiments with nonexponential service times and interarrival times in the base process N. In particular, we work with hyperponential  $(H_2)$  examples.

Tables 3.4, 3.5 and 3.6 show estimates of  $P(W_y > b)$  for the  $M_t/M/1$ ,  $M_t/H_2/1$  and  $(H_2)_t/M/1$ models, respectively. All three tables show results for y = 0.0 and y = 0.5 as a function of  $1 - \rho$ with base parameter triple  $(\beta, \gamma, b) = (1, 2.5, 4)$  in (3.39) based on 40,000 replications. The mean service time is fixed at  $\mu^{-1} = 1$ , so that  $\bar{\lambda} = \rho$  in all cases. The scv of the  $H_2$  cdf is always  $c^2 = 2$ . The scaling in (3.39) is performed as a function of  $\rho$  in order to produce nearly stable results in each row.

We start by showing the estimate of the tail probability  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$ , and then we show the corresponding estimates for the components  $e^{-\theta^* b}$  and  $A_y \equiv e^{\theta^* b} \hat{p}$ , which is followed by the lower and upper bounds in (3.20) of Corollary 3.3.3. We finally show the s.e., the associated 95% CI bounds (lb and ub), and the r.e. In all cases the relative error is less than 0.0015 or 0.15%.

For the two cases y = 0.0 and y = 0.5, we also display estimates of scaled tail probabilities,  $P(W_y > b)/P(W > b)$ , where P(W > b) is the corresponding estimate for the stationary model. We do this because we seek estimates that are more stable as functions of  $1 - \rho$ , and thus support approximations for the limiting RPBM tail probability, which is the scaled limit as  $\rho \uparrow 1$ . In Tables 3.5 and 3.6 for the  $M_t/H_2/1$  and  $(H_2)_t/M/1$  models we also show the alternative ratios  $P(W_y > b)/\rho$ ; we do not show that for  $M_t/M/1$  in Table 3.4 because the ratios are proportional, given that  $P(W > b) = \rho e^{-\theta^* b}$  for M/M/1 and  $\theta^*(\rho) = 1 - \rho$ . Tables 3.5 and 3.6 show that greater stability is achieved with the ratio  $P(W_y > b)/(W > b)$ .

**Table 3.4:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model for y = 0.0and y = 0.5 as a function of  $1 - \rho$  with base parameter triple  $(\beta, \gamma, b) = (1, 2.5, 4)$  in (3.39) based on 40,000 replications.

$1-\rho$	0.16	0.08	0.04	0.02	0.01
$\hat{p}$ for $y = 0.0$	0.011053	0.012192	0.012814	0.013122	0.013263
$e^{-\theta^* b}$	0.0183	0.0183	0.0183	0.0183	0.0183
$A_y$	0.604	0.666	0.700	0.716	0.724
$A_y^-$ LB in (3.20)	0.377	0.413	0.431	0.440	0.445
$A_y^+$ UB in (3.20)	0.840	0.920	0.960	0.980	0.990
s.e.	1.75 E-05	1.69E-05	1.71E-05	1.73E-05	1.74E-05
95% CI (lb)	0.01102	0.01216	0.01278	0.01309	0.01323
(ub)	0.01109	0.01223	0.01285	0.01316	0.01330
r.e.	0.001582	0.001387	0.001333	0.001319	0.001313
$P(W_y > b) / P(W > b)$	0.71845	0.72356	0.72879	0.73103	0.73144
diff w.r.t. last column	0.01298	0.00788	0.00264	0.00041	0.00000
abs diff	0.01298	0.00788	0.00264	0.00041	0.00000
$\hat{p}$ for $y = 0.5$	0.025888	0.028396	0.029551	0.030110	0.030430
$e^{-\theta^* b}$	0.0183	0.0183	0.0183	0.0183	0.0183
$A_y$	1.413	1.550	1.613	1.644	1.661
$A_y^-$ LB in (3.20)	0.840	0.920	0.960	0.980	0.990
$A_y^+$ UB in (3.20)	1.869	2.047	2.137	2.181	2.203
s.e.	$3.87 \text{E}{-}05$	3.74E-05	3.80E-05	3.86E-05	3.89E-05
95% CI (lb)	0.02581	0.02832	0.02948	0.03003	0.03035
(ub)	0.02596	0.02847	0.02963	0.03019	0.03051
r.e.	0.001496	0.001318	0.001286	0.001281	0.001279
$P(W_y > b) / P(W > b)$	1.68266	1.68517	1.68068	1.67751	1.67821
diff w.r.t. last column	-0.00445	-0.00696	-0.00247	0.00071	0.00000
abs diff	0.00445	0.00696	0.00247	0.00071	0.00000

Tables 3.4, 3.5 and 3.6 strongly support the heavy-traffic limit in Theorem 3.6.1, establishing convergence to RPBM as  $\rho \uparrow 1$ . The stability of the scaled quantities is especially clear through the ratios  $P(W_y > b)/P(W > b)$ . For the ratios at the bottom of the tables, we also show the

**Table 3.5:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/H_2/1$  model for y = 0.0and y = 0.5 as a function of  $1 - \rho$  with base parameter triple  $(\beta, \gamma, b) = (1, 2.5, 4)$  in (3.39) based on 40,000 replications.

$1-\rho$	0.16	0.08	0.04	0.02	0.01
$ heta^*( ho)$	0.101	0.0519	0.0263	0.0132	0.00664
$\hat{p}$ for $y = 0.0$	0.050594	0.052946	0.054024	0.054544	0.054904
$e^{-\theta^* b}$	0.0807	0.0747	0.0720	0.0707	0.0701
$A_y$	0.627	0.708	0.750	0.771	0.783
$A_y^-$ LB in (3.20)	0.477	0.532	0.560	0.573	0.580
$A_y^+$ UB in (3.20)	0.789	0.894	0.947	0.974	0.987
s.e.	7.49E-05	5.64 E-05	5.13E-05	5.03E-05	5.01E-05
95% CI (lb)	0.05045	0.05284	0.05392	0.05445	0.05481
(ub)	0.05074	0.05306	0.05412	0.05464	0.05500
r.e.	0.001480	0.001065	0.000950	0.000923	0.000913
$P(W_y > b) / P(W > b)$	0.79534	0.79246	0.79200	0.79200	0.79377
diff w.r.t. last column	-0.00158	0.00131	0.00177	0.00177	0.00000
abs diff	0.00158	0.00131	0.00177	0.00177	0.00000
$A_y/ ho$	0.74662	0.76999	0.78125	0.78680	0.79107
diff w.r.t. last column	0.04445	0.02108	0.00982	0.00427	0.00000
abs diff	0.04445	0.02108	0.00982	0.00427	0.00000
$\hat{p}$ for $y = 0.5$	0.086646	0.092721	0.095707	0.096711	0.097186
$e^{-\theta^* b}$	0.0807	0.0747	0.0720	0.0707	0.0701
$A_y$	1.074	1.241	1.329	1.367	1.386
$A_y^-$ LB in (3.20)	0.789	0.894	0.947	0.974	0.987
$A_y^+$ UB in (3.20)	1.305	1.502	1.603	1.654	1.679
s.e.	1.25E-04	9.42E-05	8.49E-05	8.28E-05	8.28E-05
95% CI (lb)	0.08640	0.09254	0.09554	0.09655	0.09702
(ub)	0.08689	0.09291	0.09587	0.09687	0.09735
r.e.	0.001442	0.001016	0.000887	0.000856	0.000852
$P(W_y > b) / P(W > b)$	1.36208	1.38777	1.40307	1.40428	1.40505
diff w.r.t. last column	0.04297	0.01728	0.00198	0.00077	0.00000
abs diff	0.04297	0.01728	0.00198	0.00077	0.00000
$A_y/ ho$	1.27865	1.34842	1.38403	1.39507	1.40028
diff w.r.t. last column	0.12163	0.05186	0.01625	0.00521	0.00000
abs diff	0.12163	0.05186	0.01625	0.00521	0.00000

difference and absolute difference of the value with value in the final column of the table.

A close examination of Tables 3.5 and 3.6 show that there is a consistent sign in the differences in the second-to-last row, being positive for the  $M_t/H_2/1$  in Table 3.5 and negative for the  $(H_2)_t/M/1$ 

**Table 3.6:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $(H_2)_t / M / 1$  model for y = 0.0and y = 0.5 as a function of  $1 - \rho$  with base parameter triple  $(\beta, \gamma, b) = (1, 2.5, 4)$  in (3.39) based on 40,000 replications.

$1-\rho$	0.16	0.08	0.04	0.02	0.01
$ heta^*( ho)$	0.113	0.0548	0.0270	0.0134	0.00669
$\hat{p}$ for $y = 0$	0.038876	0.046701	0.050799	0.053020	0.053985
$e^{-\theta^*b}$	0.0593	0.0645	0.0670	0.0682	0.0689
$A_y$	0.655	0.724	0.758	0.777	0.784
$A_y$ LB	0.477	0.532	0.559	0.573	0.580
$A_y$ UB	0.840	0.920	0.960	0.980	0.990
s.e.	4.36E-05	4.56E-05	4.73E-05	4.88E-05	4.95E-05
95% CI (lb)	0.03879	0.04661	0.05071	0.05292	0.05389
(ub)	0.03896	0.04679	0.05089	0.05312	0.05408
r.e.	0.001123	0.000976	0.000932	0.000920	0.000917
$P(A_y > b) / P(A > b)$	0.78051	0.78763	0.78988	0.79280	0.79187
diff	0.01136	0.00424	0.00199	-0.00093	0.00000
abs diff	0.01136	0.00424	0.00199	0.00093	0.00000
$A_y/ ho$	0.78015	0.78747	0.78988	0.79279	0.79186
diff	0.01171	0.00439	0.00198	-0.00094	0.00000
abs diff	0.01171	0.00439	0.00198	0.00094	0.00000
$\hat{p}$ for $y = 0.5$	0.071241	0.084111	0.090923	0.094201	0.096045
$e^{-\theta^* b}$	0.0593	0.0645	0.0670	0.0682	0.0689
$A_y$	1.201	1.305	1.357	1.380	1.395
$A_y$ LB	0.840	0.920	0.960	0.980	0.990
$A_y$ UB	1.477	1.592	1.648	1.677	1.691
s.e.	7.61E-05	7.71E-05	7.93E-05	8.13E-05	8.21E-05
95% CI (lb)	0.07109	0.08396	0.09077	0.09404	0.09588
(ub)	0.07139	0.08426	0.09108	0.09436	0.09621
r.e.	0.001068	0.000917	0.000873	0.000863	0.000855
$P(W_y > b) / P(W > b)$	1.43030	1.41856	1.41378	1.40857	1.40881
diff	-0.02149	-0.00975	-0.00497	0.00024	0.00000
abs diff	0.02149	0.00975	0.00497	0.00024	0.00000
$A_y/ ho$	1.42963	1.41826	1.41378	1.40856	1.40878
diff	-0.02085	-0.00948	-0.00500	0.00023	0.00000
abs diff	0.02085	0.00948	0.00500	0.00023	0.00000

model Table 3.6. These consistent signs in Tables 3.5 and 3.6 suggest that the two cases  $M_t/H_2/1$ and  $(H_2)_t/M/1$  serve as one-sided bounds on RPBM. We provide strong theoretical support for this idea in Theorem A.2.1 and Corollary A.2.1 of Appendix A. Those results show that the one-sided bounds apply exactly to the asymptotic decay rates  $\theta^*$ , which is the dominant part of the actual tail probability. For the cases considered in Table 3.6, it is natural to wonder if the refinement of the rare-event algorithm for the first non-exponential interarrival time makes much difference. We show that it does not for these cases with higher  $\rho$  in §A.3.6 of Appendix A.

Tables 3.4, 3.5 and 3.6 show that the bounds  $A_y^-$  and  $A_y^+$  in (3.20) are not too close, and thus not good approximations for the actual  $A_y$ . Experiments show that the average of the two bounds is not a consistently good approximation for  $A_y$  either.

Simulation results over a wide range of y show that  $P(W_y > b)$  consistently increases from a minimum at y = 0 to a maximum at y = 0.5 and then decreases to back to the minimum at y = 1, with the values for y = 1/4 and y = 3/4 being approximately equal to P(W > b). It remains to establish theoretical supporting results.

### 3.5.4 Estimating the Moments of $W_y$

We now apply the extension of the algorithm in §3.4.5 to estimate the first two moments of  $W_y$ , reporting the estimated mean and standard deviation. In Table 3.7 we first show preliminary results for the stationary M/M/1 model, so that we can judge the algorithm against known exact

**Table 3.7:** Estimated mean E[W] and standard deviation SD(W) as a function of  $1 - \rho$  for five cases of the stationary M/M/1 queue:  $\mu = 1, \bar{\lambda} = \rho$ 

$1 - \rho$	0.16	0.08	0.04	0.02	0.01
$n_s$ in (3.36)	40,000	40,000	40,000	40,000	40,000
$\delta$ in (3.36)	0.001	0.001	0.001	0.001	0.001
largest $b$	41	86	173	345	691
P(W > 0)	0.8396	0.9201	0.9601	0.9799	0.9900
exact	0.8400	0.9200	0.9600	0.9800	0.9900
s.e. of $P(W > 0)$	6.86E-04	3.71E-04	1.93E-04	9.73E-05	4.98E-05
%95 CI of $P(W > 0)$	[0.8383, 0.8410]	[0.919,  0.921]	[0.9598,  0.9605]	[0.9797,  0.9801]	[0.9899,  0.9901]
E[W]	5.249	11.499	23.999	49.000	99.000
exact	5.250	11.500	24.000	49.000	99.000
s.e. of $E[W]$	1.59E-03	1.27E-03	9.51E-04	6.93E-04	4.94E-04
%95 CI of $E[W]$	[5.246, 5.252]	[11.497,  11.502]	[23.997, 24.001]	[48.999,  49.001]	[98.999, 99.001]
E[W W>0]	6.251	12.497	24.995	50.003	100.005
%95 CI of $E[W W>0]$	[6.238, 6.265]	[12.485, 12.510]	[24.983, 25.007]	[49.992,  50.014]	[99.994,  100.015]
$E[W^2]$	65.624	287.494	1199.982	4899.957	19,800.03
exact	65.625	287.500	1200.000	4900.000	19,800.00
s.e. of $E[W^2]$	1.50E-02	2.33E-02	3.40E-02	4.92E-02	7.04E-02
%95 CI of $E[W^2]$	[65.60,  65.65]	[287.45, 287.54]	[1199.9, 1200.1]	[4899.9, 4900.1]	[19,799.9,19,800.2]
SD[W]	6.170	12.460	24.981	49.990	99.995
exact	6.1695	12.450	24.980	49.990	99.995
$P(W > 0)/\rho$	0.9995	1.0002	1.0001	0.9999	1.0000
exact	1.0000	1.0000	1.0000	1.0000	1.0000
$(1-\rho)E[W]$	0.8398	0.9200	0.9600	0.9800	0.9900
$(1-\rho)SD[W]$	0.9873	0.9968	0.9992	0.9998	0.9999
$(1-\rho)E[W]/\rho$	0.9998	0.9999	0.9999	1.0000	1.0000
$(1-\rho)SD[W]/\rho$	0.8293	0.9171	0.9593	0.9798	0.9899
$(1-\rho)E[W W>0]$	1.0002	0.9998	0.9998	1.0001	1.0000
$(1-\rho)SD[W W>0]$	1.0002	1.0000	1.0000	1.0000	1.0000

results. For ease of comparison, we show the corresponding known exact values for P(W > 0), E[W],  $E[W^2]$  and SD(W). The first section of Table 3.7 with three rows shows the algorithm parameters. The final seven rows of Table 3.7 are included to show alternatives ways of scaling

aimed at achieving stable values across all values of  $1 - \rho$ . In this case, knowing that W has an exponential distribution except for an atom of mass  $1 - \rho$  at the origin, we are not surprised to see that the final two rows provide the best scaling. We will use those rows in the following tables for time-varying arrival-rate functions.

Tables 3.8 and 3.9 show corresponding estimates of the time varying mean  $E[W_y]$  and standard deviation  $SD(W_y)$  for the special case of y = 0.5 for associated  $M_t/M/1$  model with arrival-rate function in (2.4) for base parameter pairs  $(\beta, \gamma) = (1, 2.5)$  and  $(\beta, \gamma) = (4, 2.5)$  using the scaling convention in (3.39). Both have cycle length  $2\pi/\gamma$ , which equals 6.28/0.1 = 62.8 for  $\rho = 0.8$ . The higher relative amplitude in Table 3.9 leads to much larger mean values at y = 0.5, which tends to produce the largest values in the cycle. As can be seen from the online supplement, much lower values occur for y = 0, which tends to produce the least values.

**Table 3.8:** Estimated mean  $E[W_y]$  and standard deviation  $SD(W_y)$  as a function of  $1 - \rho$  for five cases of the  $M_t/M/1$  queue at y = 0.5:  $\mu = 1, \bar{\lambda} = \rho$  and base parameter pair  $(\beta, \gamma) = (1, 2.5)$ 

$1-\rho$	0.16	0.08	0.04	0.02	0.01
$n_s$ in (3.36)	40,000	40,000	40,000	40,000	40,000
$\delta$ in (3.36)	0.001	0.001	0.001	0.001	0.001
largest $b$	41	86	173	345	691
$P(W_y > 0)$	0.8801	0.9411	0.9714	0.9851	0.9930
s.e. of $P(W_y > 0)$	9.85E-04	6.54E-04	4.51E-04	2.92 E- 04	2.19E-04
%95 CI of $P(W_y > 0)$	[0.8782, 0.8820]	[0.9399,  0.9424]	[0.9705,  0.9723]	[0.9845,  0.9856]	[0.9926,  0.9934]
$E[W_y]$	6.839	14.927	31.194	63.667	128.411
std of $E[W_y]$	6.42E-03	1.20E-02	2.36E-02	4.69E-02	9.30E-02
%95 CI of $E[W_y]$	[6.827,  6.852]	[14.903, 14.950]	[31.147, 31.240]	[63.575,  63.759]	[128.228, 128.593]
$E[W_y W_y>0]$	7.771	15.860	32.113	64.632	129.315
%95 CI of $E[W_y W_y>0]$	[7.740, 7.803]	[15.814,  15.907]	[32.036, 32.189]	[64.501,  64.763]	[129.075, 129.554]
$E[W_y^2]$	97.057	427.685	1795.344	7344.665	29,673.77
std of $E[W_y^2]$	7.81E-02	0.302	1.207	4.829	19.314
%95 CI of $E[W_y^2]$	[96.90, 97.21]	[427.09, 428.28]	[1793.0, 1797.7]	[7335.2, 7354.13]	[29, 636, 29, 712]
$SD[W_y]$	7.091	14.314	28.676	57.369	114.824
$P(W_y > 0)/\rho$	1.0478	1.0230	1.0119	1.0052	1.0030
$(1-\rho)E[W_y W_y>0]$	1.2434	1.2688	1.2845	1.2926	1.2931
$(1-\rho)SD[W_y W_y>0]$	1.1301	1.1395	1.1433	1.1452	1.1472

Finally, Table 3.10 shows estimates of the time varying mean  $E[W_y]$  and standard deviation  $SD(W_y)$  for the special case of y = 0.5 for associated  $(H_2)_t/M/1$  model with arrival-rate function

**Table 3.9:** Estimated mean  $E[W_y]$  and standard deviation  $SD(W_y)$  as a function of  $1 - \rho$  for five cases of the  $M_t/M/1$  queue at y = 0.5:  $\mu = 1, \overline{\lambda} = \rho$  and base parameter pair  $(\beta, \gamma) = (4, 2.5)$  having larger relative amplitude

$1-\rho$	0.16	0.08	0.04	0.02	0.01
$n_s$ in (3.36)	40,000	40,000	40,000	40,000	40,000
$\delta$ in (3.36)	0.001	0.001	0.001	0.001	0.001
largest $b$	41	86	173	345	691
$P(W_y > 0)$	0.9728	0.9883	0.9967	0.9965	0.9993
s.e. of $P(W_y > 0)$	3.61E-03	2.69E-03	2.05E-03	1.16E-03	8.52E-04
%95 CI of $P(W_y > 0)$	[0.9657,  0.9799]	[0.9831,  0.9936]	[0.9927,  1.0000]	[0.9943,  0.9988]	[0.9976,  1.0000]
$E[W_y]$	15.148	33.583	70.677	145.183	294.222
std of $E[W_y]$	5.58E-02	1.13E-01	2.27E-01	4.59E-01	9.15 E-01
%95 CI $E[W_y]$	[15.04,  15.26]	[33.36,  33.81]	[70.23, 71.12]	[144.3, 146.1]	[292.4, 296.0]
$E[W_y W_y > 0]$	15.572	33.980	70.909	145.690	294.437
%95 CI of $E[W_y W_y>0]$	[15.35, 15.80]	[33.58, 34.39]	[70.2, 71.6]	[144.5, 147.0]	[292.4, 296.7]
$E[W_y^2]$	331.868	1528.127	6547.951	27,092.17	110,239.9
std of $E[W_y^2]$	1.023	4.263	17.227	69.632	0.785
%95 CI of $E[W_y^2]$	[329.9, 333.9]	[1519.8,  1536.5]	[6514,  6582]	[26,955, 27,228]	[109, 691,  110, 787]
$SD[W_y]$	10.119	20.007	39.405	77.551	153.861
$P(W_y > 0)/\rho$	1.1581	1.0743	1.0383	1.0169	1.0094
$(1-\rho)E[W_y W_y>0]$	2.4915	2.7184	2.8364	2.9138	2.9444
$(1-\rho)SD[W_y W_y>0]$	1.5892	1.5830	1.5704	1.5442	1.5371

in (2.4) for base parameter pairs  $(\beta, \gamma) = (1, 2.5)$ , but here we let  $\delta$  increase as  $1 - \rho$  decreases. Table 3.10 shows that the precision remains good for all  $\rho$ . (For the cases considered in Table 3.10, the refinement of the rare-event algorithm for the first non-exponential interarrival time does not make too much difference, but it matters more than for Table 3.6, as we show in §A.3.6 of Appendix A.)

### 3.6 The Supporting Heavy-Traffic FCLT

To explain the unified numerical results in §3.5, we now review and extend the heavy-traffic (HT) functional central limit theorem (FCLT) for periodic  $G_t/G/1$  queues in Theorem 3.2 of Whitt (2014). An extension of the HT FCLT in Whitt (2014) is needed because that HT FCLT is stated for the scaled arrival process and the scaled queue-length process, but not the scaled workload process that we consider here. A similar argument applies to the workload process, jointly with

$1 - \rho$	0.16	0.08	0.04	0.02	0.01
$ heta^*( ho)$	0.113	0.0548	0.0270	0.0134	0.00669
$n_s$	40,000	40,000	40,000	40,000	40,000
δ	0.001	0.002	0.004	0.008	0.016
largest $b$	41	86	173	345	691
$P(W_y > 0)$	0.8721	0.9382	0.9691	0.9853	0.9923
s.e. of $P(W_y > 0)$	7.36E-04	4.81E-04	3.18E-04	2.34E-04	1.51E-04
%95 CI of $P(W_y > 0)$	[0.8707, 0.8736]	[0.9373,  0.9391]	[0.9685,  0.9697]	[0.9848,  0.9857]	[0.9920,  0.9926]
$E[W_y]$	9.125	20.501	43.720	88.613	179.456
std of $E[W_y]$	5.56E-03	1.05E-02	2.07E-02	4.07E-02	8.18E-02
%95 CI of $E[W_y]$	[9.114, 9.135]	[20.480, 20.521]	[43.162,  43.243]	[88.533, 88.693]	[179.296, 179.616]
$E[W_y W_y > 0]$	10.462	21.851	45.114	89.937	180.845
%95 CI of $E[W_y W_y > 0]$	[10.432, 10.492]	[21.807, 21.895]	[44.510,  44.651]	[89.814, 90.060]	[180.630, 181.061]
$E[W_y^2]$	175.380	814.768	3489.720	$14,\!425.330$	58,633.918
std of $E[W_y^2]$	8.65E-02	0.350	1.424	5.703	23.026
%95 CI of $E[W_y^2]$	[175.21, 175.55]	[814.08, 815.46]	[3,486.93,  3,492.51]	[14, 414, 14, 436]	[58, 588, 58, 679]
$SD[W_y]$	9.598	19.862	40.289	81.074	162.571
$P(W_y > 0)/\rho$	1.0383	1.0198	1.0095	1.0054	1.0023
$(1-\rho)E[W_y]$	1.4599	1.6401	1.7488	1.7723	1.7946
$(1-\rho)SD[W_y]$	1.5357	1.5889	1.6116	1.6215	1.6257
$(1-\rho)E[W_y]/\rho$	1.7380	1.7827	1.8216	1.8084	1.8127
$(1-\rho)SD[W_y]/\rho$	1.2900	1.4618	1.5471	1.5891	1.6095
$(1-\rho)E[W_y W_y>0]$	1.6739	1.7481	1.8045	1.7987	1.8085
$(1-\rho)SD[W_y W_y>0]$	1.5316	1.5818	1.5828	1.6189	1.6243

**Table 3.10:** Estimated mean  $E[W_y]$  and standard deviation  $SD(W_y)$  as a function of  $1 - \rho$  for five cases of the  $(H_2)_t/M/1$  queue at y = 0.5:  $\mu = 1, \bar{\lambda} = \rho$  and base parameter pair  $(\beta, \gamma) = (1, 2.5)$ .

the other processes, but it is more natural to apply Theorem 9.3.4 of Whitt (2002) than Iglehart and Whitt (1970b), because the workload process is defined there in §9.2 essentially the same way as the workload is defined in §3.2.

The innovative part of Whitt (2014) is the new HT scaling in (3.38) to capture the impact of the periodicity in an interesting and revealing way, as demonstrated by the tables in §3.5. As shown in Whitt (2014), the periodicity has no impact on the heavy-traffic limit if this additional scaling is not included. (That elementary observation was made earlier by Falin (1989); the main contribution of Whitt (2014) is the new scaling.)

### 3.6.1 The Heavy-Traffic FCLT

We assume that the rate-1 arrival and service processes N and V specified in §3.2 are independent and each satisfies a FCLT. To state the result, let  $\hat{N}_n$  and  $\hat{S}_n^v$  be the scaled processes defined by

$$\hat{N}_n(t) \equiv n^{-1/2} [N(nt) - nt] \quad \text{and} \quad \hat{S}_n^v(t) \equiv n^{-1/2} [\sum_{i=1}^{\lfloor nt \rfloor} V_k - nt], \quad t \ge 0,$$
(3.41)

with  $\equiv$  denoting equality in distribution and  $\lfloor x \rfloor$  denoting the greatest integer less than or equal to x. We assume that

$$\hat{N}_n \Rightarrow c_a B_a \quad \text{and} \quad \hat{S}_n^v \Rightarrow c_s B_s \quad \text{in} \quad \mathcal{D} \quad \text{as} \quad n \to \infty,$$
(3.42)

where  $\mathcal{D}$  is the usual function space of right-continuous real-valued functions on  $[0, \infty)$  with left limits and  $\Rightarrow$  denotes convergence in distribution, as in Whitt (2002), while  $B_a$  and  $B_s$  are independent standard (mean 0, variance 1) Brownian motion processes (BM's). The assumed independence implies joint convergence in (3.42) by Theorem 11.4.4 of Whitt (2002).

We emphasize that GI assumptions are not needed, but that is an important special case. If the service times  $V_k$  are i.i.d. mean-1 random variables with variance = scv  $c_s^2$ , then the limit in (3.42) holds with service variability parameter  $c_s$ . Similarly, if the base arrival process is a renewal process or an equilibrium renewal process with times between renewals having mean 1 and variance = scv  $c_a^2$ , then the limit in (3.42) holds with arrival variability parameter  $c_a$ . (See Nieuwenhuis (1989) for theoretical support in the case of an equilibrium renewal process.)

Theorem 9.3.4 of Whitt (2002) refers to the conditions of Theorem 9.3.3, which requires a joint FCLT for the partial sums of the arrival and service processes, notably (3.9) on p. 295. That convergence follows from the FCLT's we assumed for N and V in (3.42) above. In particular, the assumed FCLT for N implies the associated FCLT for the partial sums of the interarrival times by Theorem 7.3.2 and Corollary 7.3.1 of Whitt (2002).

We create a model for each  $\rho$ ,  $0 < \rho < 1$ , by defining the arrival-rate function

$$\lambda_{\rho}(t) \equiv \rho + (1-\rho)\lambda_d((1-\rho)^2 t), \quad t \ge 0,$$
(3.43)

where  $\lambda_d$  is a periodic function with period  $C^*$  satisfying

$$\bar{\lambda}_d \equiv \frac{1}{C^*} \int_0^{C^*} \lambda_d(s) \, ds \equiv 0. \tag{3.44}$$

As a regularity condition, we also require that the function  $\lambda_d$  be an element of  $\mathcal{D}$ . As a consequence of (3.43) and (3.44), the average arrival rate is  $\bar{\lambda}_{\rho} = \rho$ ,  $0 < \rho < 1$ . Hence, (3.38) is a special case of (3.43); see §3.6.3 below.

We can also work with cumulative functions and let the cumulative arrival-rate function in model  $\rho$  be

$$\Lambda_{\rho}(t) \equiv \rho t + (1-\rho)^{-1} \Lambda_d((1-\rho)^2 t), \quad t \ge 0,$$
(3.45)

where

$$\Lambda_d(t) \equiv \int_0^t \lambda_d(s) \, ds, \qquad (3.46)$$

for  $\lambda_d$  again being the periodic function in (3.44). From (3.45)-(3.46), we see that the associated arrival-rate function obtained by differentiation in (3.45) is (3.43).

The time scaling in (3.43) and (3.45) implies that the period in model  $\rho$  with arrival-rate function  $\lambda_{\rho}(t)$  in (3.43) is  $C_{\rho} = C^*(1-\rho)^{-2}$ , where  $C^*$  is the period of  $\lambda_d(t)$  in (3.44). Thus the period  $c_{\rho}$  in model  $\rho$  is growing with  $\rho$ .

Now let  $A_{\rho}(t) \equiv N(\Lambda_{\rho}(t))$  be the arrival process, using the cumulative arrival-rate function  $\Lambda_{\rho}$ in (3.45) in place of  $\Lambda$  in (2.2). Let  $Q_{\rho}(t)$  and  $W_{\rho}(t)$  be the associated queue length process and workload process in the  $G_t/G/1$  model with arrival process  $A_{\rho}(t)$  in (3.43) and service times from the fixed service process V, constructed as in §9.2 of Whitt (2002). Then let associated scaled arrival, queue length and workload processes be defined by

$$\hat{A}_{\rho}(t) \equiv (1-\rho)[A_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}t],$$

$$\hat{Q}_{\rho}(t) \equiv (1-\rho)Q_{\rho}((1-\rho)^{-2}t) \text{ and } \hat{W}_{\rho}(t) \equiv (1-\rho)W_{\rho}((1-\rho)^{-2}t), \quad t \ge 0.$$
(3.47)

The scaled processes in (5.40) and the HT limit all have cycle length  $C^*$ .

The following heavy-traffic FCLT states that  $\hat{A}_{\rho}$  converges to periodic Brownian motion (PBM), while  $\hat{Q}_{\rho}$  and  $\hat{W}_{\rho}$  converge to a common reflected periodic Brownian motion (RPBM). To explain, let e be the identity function with  $e(t) = t, t \ge 0$ . By a PBM, we mean a process  $cB + \Lambda - e \equiv$  $\{cB(t) + \Lambda_d(t) - t : t \ge 0\}$ , where B is a BM and  $\Lambda_d$  is of the form (3.46), so that the process has periodic deterministic drift  $\lambda_d(t) - 1$ . Let  $\psi$  be the usual one-dimensional reflection map as on pp. 87, 290 and 439 of Whitt (2002). Given that  $cB + \Lambda - e$  is a PBM,  $\psi(cB + \Lambda - e)$  is a RPBM. To state the HT FCLT, let  $\mathcal{D}^k$  be the k-fold product space of  $\mathcal{D}$  with itself and let  $\stackrel{d}{=}$  denote equality in distribution. **Theorem 3.6.1.** (heavy-traffic limit extending Whitt (2014)) If, in addition to the definitions and assumptions in (5.26)-(5.40) above, the system starts empty at time 0, then

$$(\hat{A}_{\rho}, \hat{Q}_{\rho}, \hat{W}_{\rho}) \Rightarrow (X_a, Z, Z) \quad in \quad \mathcal{D}^3 \quad as \quad \rho \uparrow 1,$$
(3.48)

where

$$X_a \equiv c_a B_a + \Lambda_d - e, \quad X \equiv X_a - c_s B_s \quad and \quad Z \equiv \psi(X), \tag{3.49}$$

with  $B_a$  and  $B_s$  being independent BM's,  $\Lambda_d$  in (3.46) and  $c_a$  and  $c_s$  being the variability parameters in (3.42), so that  $X \stackrel{d}{=} c_x B$ , where  $c_x \equiv \sqrt{c_a^2 + c_s^2}$  and B is a BM.

The joint limit for  $(\hat{A}_{\rho}, \hat{Q}_{\rho})$  is established in Theorem 3.2 of Whitt (2014), which in turn follows quite directly from Iglehart and Whitt (1970b). (We remark that there is a typographical error in the translation term on the first line of (13) in the proof of Theorem 3.2 of Whitt (2014); it should be  $-(1-\rho)^{-2}t$  as in equation (11) there instead of  $-(1-\rho)^{-2}\rho t$ .) To treat the workload, we apply Theorem 9.3.4 of Whitt (2002), which implies that the limit for  $\hat{W}_{\rho}$  is the same as for the limit for  $\hat{Q}_{\rho}$ .

Unfortunately, the periodic feature makes the RPBM complicated, so that it remains to derive explicit expressions for its transient and periodic steady-state distributions. The present chapter contributes by developing an effective algorithm to calculate the periodic steady-state distribution.

### 3.6.2 Approximations for the Steady-State Workload

Our algorithm for the periodic steady-state distribution of RPBM calculates the periodic steadystate distribution of the scaled workload process in a  $GI_t/GI/1$  queue for suitably large  $\rho$  and uses Theorem 3.6.1 for justification. While that approach is intuitively reasonable, there are steps that remain to be justified. Proper justification requires an additional limit interchange argument, which has been done in some contexts, e.g., see Budhiraja and Lee (2009), but here is left for a topic of future research.

Hence, we assume that those steps are justified. In particular, we assume that the workload process and the limiting RPBM have proper periodic steady-state distributions for each  $\rho$  and that there is convergence in distribution of the scaled periodic steady state workload to the periodic steady state of RPBM as  $\rho \uparrow 1$ . In particular, in addition to the limit  $\hat{W}_{\rho} \Rightarrow Z$  in  $\mathcal{D}$  as  $\rho \uparrow 1$  established in Theorem 3.6.1, we assume that

$$W_{\rho}((k+y)c_{\rho}) \Rightarrow W_{\rho,y}(\infty) \quad \text{in } \mathbb{R} \quad \text{as} \quad k \to \infty,$$
(3.50)

where  $P(W_{\rho,y}(\infty) < \infty) = 1$  for all  $\rho$  and  $y, 0 < \rho < 1$  and  $0 \le y < 1$ , or, equivalently,

$$\hat{W}_{\rho}((k+y)C^*) \Rightarrow \hat{W}_{\rho,y}(\infty) \quad \text{in } \mathbb{R} \quad \text{as} \quad k \to \infty,$$
(3.51)

where  $P(\hat{W}_{\rho,y}(\infty) < \infty) = 1$  for all  $\rho$  and  $y, 0 < \rho < 1$  and  $0 \le y < 1$ , and

$$Z((k+y)C^*) \Rightarrow Z_y(\infty) \quad \text{in } \mathbb{R} \quad \text{as} \quad k \to \infty,$$
 (3.52)

where  $P(Z_y(\infty) < \infty) = 1$  for all  $y, 0 \le y < 1$ . With these assumptions, our algorithm applies to RPBM using the approximation

$$P(Z_y(\infty) > x) \approx P(\hat{W}_{\rho,y}(\infty) > x) \tag{3.53}$$

where  $\rho$  is chosen to be suitably large.

### 3.6.3 Application to the Sinusoidal Arrival-Rate Function

For the sinusoidal example in (2.4), we let

$$\lambda_d(t) \equiv \lambda\beta \sin\left(\gamma t\right), \quad t \ge 0, \tag{3.54}$$

for  $\lambda_d(t)$  in (3.44), so that the cycle length is  $C^* = 2\pi/\gamma$ . With (3.54) and  $\bar{\lambda} \equiv \rho$ , (3.43) becomes (3.38), so that the cycle length in model  $\rho$  is  $c_{\rho} = C^*(1-\rho)^{-2} = 2\pi/(\gamma(1-\rho)^2)$ . When we consider the periodic steady-state workload, the time scaling is gone but we still have the spatial scaling. When the traffic intensity is  $\rho$ , we multiply by  $1 - \rho$ ; i.e., we have

$$\hat{W}_{\rho,y}(\infty) = (1-\rho)W_{\rho,y}(\infty).$$
 (3.55)

Hence, to have asymptotically convergent models, we should choose parameter four-tuples  $(\bar{\lambda}_{\rho}, \beta_{\rho}, \gamma_{\rho}, b_{\rho})$ indexed by  $\rho$  as indicated in (3.39).

### **3.6.4** Approximations for the $G_t/G/1$ Model

To apply the heavy-traffic FCLT to generate approximations for the performance of the periodic steady-state workload in a general periodic  $G_t/G/1$  model (without i.i.d. assumptions), we assume that the assumptions in §3.6.1 are satisfied so that Theorem 3.6.1 is valid. We then approximate the model by a  $GI_t/GI/1$  model which has the same HT FCLT limit process. In other words, we approximate the underlying rate-1 arrival counting process N by a renewal process with i.i.d. mean-1 times between renewals having scv  $c_a^2$ , where  $c_a$  is the arrival process variability parameter in the assumed FCLT (3.42). Similarly, we approximate the sequence of mean-1 service times  $\{V_k\}$ by a sequence of mean-1 i.i.d. random variables with a scv equal to  $c_s^2$ , where  $c_s$  is the service variability parameter in the assumed FCLT (3.42). Both approximations are exact for GI.

To construct the specific GI arrival and service processes, we follow the approximation scheme in §3 of Whitt (1982). We apply the same method for the interarrival times  $U_k$  of N as we do to the service times  $V_k$ , so we only discuss the service times. If  $c_s^2 \approx 1$ , then we use a mean-1 exponential (M) distribution; if  $c_s^2 > 1$ , then we use a mean-1 hyperexponential  $(H_2)$  distribution with pdf  $f_V(x) = p_1\mu_1e^{-\mu_1x} + p_2\mu_2e^{-\mu_2x}$ , with  $p_1 + p_2 = 1$ , having parameter triple  $(p_1, \mu_1, \mu_2)$ . To reduce the parameters to two (the mean and scv), we assume balanced means, i.e.,  $p_1/\mu_1 = p_2/\mu_2$ , as in (3.7) of Whitt (1982). If  $c_s^2 < 1$  and if  $c_s^2 \approx 1/k$  for some integer k, then we use a mean-1 Erlang  $(E_k)$  distribution (sum of k i.i.d. exponential variables), otherwise if  $c_s^2 < 1$ , then we use the D+Mdistribution, i.e., a sum of a deterministic constant (D) and an exponential (M) distribution with rate  $\mu$ , which has pdf  $f_V(x) = \mu e^{-\mu(x-d)}$ ,  $x \ge d$ , as in (3.11) and (3.12) of Whitt (1982).

### 3.7 Conclusions

We have developed a new algorithm to calculate the distribution of the periodic steady-state remaining workload  $W_y$ , at time yC within a periodic cycle of length C,  $0 \le y < 1$ , in a general  $GI_t/GI/1$  single-server queue with periodic arrival-rate function. The key model assumption is the representation in (2.2) of the arrival process as a time-transformation of a rate-1 process. The algorithm is based on the new representation of  $W_y$  in (3.1) derived in §3.1.1 and §3.2. In §3.4 we developed an algorithm for computing the exact tail probabilities  $P(W_y > b)$  in the  $GI_t/GI/1$  model based on the established rare-event simulation algorithm for the associated stationary GI/GI/1 model. That connection is supported by the close relation between the two models, established in  $\S3.3$ .

We also have shown that the algorithm can be applied together with the heavy-traffic FCLT in Whitt (2014) reviewed in §3.6 to also calculate the periodic steady-state distribution and moments of reflected periodic Brownian motion (RPBM). In addition, the algorithm can be applied to approximate the tail probabilities in the more general  $G_t/G/1$  model by choosing special parameters (the squared coefficients of variation (scv) of interrenewal times) in the  $GI_t/GI/1$  model to insure that the two systems obey the same heavy-traffic FCLT.

We have verified the effectiveness of the algorithm for  $GI_t/GI/1$  queues and RPBM by conducted extensive simulation experiments for the  $GI_t/GI/1$  model with sinusoidal arrival rate in §3.1.3 and a range of traffic intensities. Some of these are reported in §3.5 and in the supplement Appendix A. It remains to investigate the algorithm for  $G_t/G/1$  queues more general than  $GI_t/GI/1$ .

## Part II

# Performance Stabilization in Time-Varying Single-Server Queues

### Chapter 4

# The Service-Rate Control Problem

Simulation is used to evaluate the performance of alternative service-rate controls designed to stabilize performance in a queue with time-varying arrival rates, service in order of arrival and unlimited waiting space. Both Markovian and non-Markovian models are considered. The simulation experiments show that a rate-matching service-rate control successfully stabilizes the expected queue length, but not the expected waiting time, while a new square-root service-rate control, based on assuming that a pointwise-stationary approximation is appropriate, approximately stabilizes the expected waiting time when the arrival rate changes slowly compared to the expected service time. We also show results for two square-root service-rate controls related to the many-server squareroot staffing formula, which are not effective for the single-server setting. This chapter is based on Ma and Whitt (2015b) with reference to Whitt (2015).

### 4.1 Introduction

In this chapter we study alternative service-rate controls to stabilize performance in a single-server queue with time-varying arrival rate and independent and identically distributed (i.i.d.) service requirements specified separately from the service rate actually provided. Our study parallels Liu and Whitt (2012b), He et al. (2016) and earlier papers cited there that develop time-varying staffing levels (number of servers) to stabilize performance in multi-server queues with flexible staffing.

The present problem of service-rate control is important for systems with only a few servers or with inflexible staffing. In many applications, even though the available service resources are fixed, it is possible to change the processing rate. Two important examples are airplane landings and airport security lines. For the airplane landing example, the number of runways is fixed, but it may be possible to change the rate of airplane landings by controlling the required separation distance between airplanes. Similarly, for the TSA inspection example, the number of security lines is fixed, but it may be possible to increase the inspection rate by relaxing inspection requirement. In these settings the possible service rates that can be achieved may be limited, but to gain insight into the potential benefits of controlling the service rate, we study the idealized case of a single server where the service rate is fully subject to control.

We conduct simulation experiments to study the performance of the service-rate controls to stabilize performance in these systems. We consider four different service-rate controls: a ratematching control that makes the service rate proportional to the arrival rate and three square-root service-rate controls. The first square-root service-rate control is a natural analog of the offered-load square-root-staffing formula used for many-server queues, where the offered load is the expected number of busy servers in an associated infinite-server system with the same arrival rate and a service-time distribution. Since the service-time distribution is unavailable in advance, we use the service-requirement distribution. The second square-root service-rate control is a variant of the first, in which the arrival rate is used in place of the offered load. The third square-root service-rate control is obtained by solving a quadratic equation, based on a steady-state heavytraffic approximation assuming that a pointwise-stationary approximation (PSA) is appropriate; see Green et al. (2007).

We show that the rate-matching control stabilizes the expected queue length, but not the expected waiting time, consistent with theoretical results established in Whitt (2015). We show that the expected waiting time tends to be inversely proportional to the arrival rate. We show that the first two square-root service-rate controls that are analogs of the square-root staffing formula for multiple server queues stabilize the mean waiting times to some extent, but not fully. We show that the final square-root control based on the PSA is effective for long cycles, where the PSA is effective, but not more generally.

The remainder of this chapter is organized as follows. In  $\S4.2$  we define the different service-rate controls with reference to theorems in Whitt (2015); in  $\S4.3$  we describe the simulation experiments we conduct using the algorithms in Chapter 2; in  $\S4.4$  we show some of the simulation results

verifying the performance of various service-rate controls; and in §4.5 we draw conclusions.

### 4.2 The Service-Rate Controls

In this section, we specify the different service-rate controls that we consider and show theoretical results developed in Whitt (2015). We use the same model for  $G_t/G_t/1$  queue as in §2.2 and the same service-time model as in §2.4.

### 4.2.1 The Rate-Matching Control

The first service-rate control is the rate-matching control

$$\mu(t) \equiv \lambda(t)/\rho, \quad t \ge 0, \tag{4.1}$$

where  $\rho$  is the desired traffic intensity. Clearly, the instantaneous traffic intensity is  $\rho(t) \equiv \lambda(t)/\mu(t)$  is constant.

Theorem 4.2.1 and Theorem 4.2.2 from Whitt (2015) show that with the rate-matching servicerate control, the queue-length process Q(t) is a time transformation of the stationary model and thus the steady-state queue-length distribution is stabilized. We let  $Q_1(t)$  be the queue-length process with constant arrival rate 1 and constant service rate  $\frac{1}{\rho}$ , which together with the stationary departure process  $D_1(t)$  is defined as below:

$$Q_1(t) \equiv A_1(t) - D_1(t), \quad t \ge 0, \tag{4.2}$$

$$D_1(t) \equiv N_s(\int_0^t \mu_1(s) \mathbf{1}_{\{Q_1(s)>0\}} \, ds) = N_s(\int_0^t \rho^{-1} \mathbf{1}_{\{Q_1(s)>0\}} \, ds), \quad t \ge 0, \tag{4.3}$$

where  $A_1 \equiv N_a$  and  $D_1(t)$  is the total number of departures in the interval [0, t].

The resulting Theorem 4.2.1 follows from the time-transformation construction of the arrival process and service process in (2.2) and (2.3), which leads to the stabilization of steady-state queue-length distribution result in Theorem 4.2.2.

**Theorem 4.2.1.** (Theorem 3.1 from Whitt (2015): time transformation of a stationary model) For (A, D, Q) with the rate-matching service-rate control and the stationary single-server model  $(A_1, D_1, Q_1)$  defined above,

$$(A(t), D(t), Q(t)) = (A_1(\Lambda(t)), D_1(\Lambda(t)), Q_1(\Lambda(t))), \quad t \ge 0.$$
(4.4)

**Theorem 4.2.2.** (Theorem 4.2 from Whitt (2015): stabilizing the queue-length distribution and the steady-state delay probability) Let  $Q_1(t)$  be the queue length process when  $\lambda(t) = 1$ ,  $t \ge 0$ . If  $Q_1(t) \Rightarrow Q_1(\infty)$  as  $t \to \infty$ , where  $P(Q_1(\infty) < \infty) = 1$ , then also

$$Q(t) \Rightarrow Q_1(\infty) \quad in \quad \mathbb{R} \quad as \quad t \to \infty,$$

$$(4.5)$$

and

$$P(W(t) > 0) = P(Q(t) \ge 1) \to \rho \quad as \quad t \to \infty.$$

$$(4.6)$$

As for the virtual waiting time process, Theorem 4.2.3 gives an expression for W(t) involving the stationary waiting time process  $W_1(t)$  and thus Theorem 4.2.4 presents a heavy-traffic limit for the steady-state virtual waiting time process, which is not stable, being asymptotically inversely proportional to the arrival-rate function.

**Theorem 4.2.3.** (Theorem 5.1 from Whitt (2015): constructing the virtual waiting time) The virtual waiting time W(t) can be represented as

$$W(t) = \Lambda_t^{-1}(W_1(\Lambda(t)), \quad t \ge 0,$$
(4.7)

where  $W_1(t)$  is the waiting time at time t in the associated stationary G/G/1 model and  $\Lambda_t^{-1}$  is the inverse of

$$\Lambda_t(v) = \Lambda(t+v) - \Lambda(t), \quad v \ge 0 \quad and \quad t \ge 0, \tag{4.8}$$

which is strictly increasing and continuous. If  $W_1(t)$  has its stationary distribution  $W_1^*$ , then  $W(t) \stackrel{d}{=} \Lambda_t^{-1}(W_1^*).$ 

We now refer to the scaled arrival rate function  $\bar{\lambda}_n(t)$  and scaled cumulative arrival rate function  $\bar{\Lambda}_n(t)$  in (5.17) of Whitt (2015) using the usual heavy-traffic scaling. The scaled queueing processes are defined as follows. Let

$$\hat{Q}_{1,n}(t) \equiv n^{-1/2} Q_{1,n}(nt), \quad t \ge 0,$$
(4.9)

so that  $\hat{Q}_n(t) = \hat{Q}_{1,n}(\bar{\Lambda}_n(nt)), t \ge 0$  by Theorem 4.2.1. Let  $W_n(t)$  be the virtual waiting time at time t in model n and define the associated scaled processes

$$\hat{W}_n(t) \equiv n^{-1/2} W_n(nt), \quad t \ge 0.$$
 (4.10)

Let  $\mathcal{D}^k$  be the k-fold product space of  $\mathcal{D}$  with itself with the usual product topology. Let R(t; a, b) be reflected Brownian motion (RBM) with drift -a and diffusion coefficient b.

**Theorem 4.2.4.** (Theorem 5.2 from Whitt (2015): heavy-traffic limit for the time-varying waiting time) Let the system start empty. Under the scaling assumptions above,

$$(\hat{Q}_n, \hat{W}_n) \Rightarrow (\hat{Q}, \hat{W}) \quad in \quad \mathcal{D}^2 \quad as \quad n \to \infty,$$

$$(4.11)$$

where

$$\hat{W}(t) \equiv \hat{Q}(t)/\lambda_f(t)$$
 and  $\hat{Q}(t) \equiv R(\Lambda_f(t); -1, c_a^2 + c_s^2), \quad t \ge 0.$  (4.12)

with  $\lambda_f$  in (2.4). As a consequence, for each T > 0,

$$\sup_{0 \le t \le T} \left\{ \hat{W}_n(t) - (\hat{Q}_n(t)/\lambda_f(t)) \right\} \Rightarrow 0 \quad as \quad n \to \infty$$
(4.13)

and, for each  $x \ge 0$ ,

$$P(\hat{Q}_n(t) > x) \to e^{-2x/(c_a^2 + c_s^2)}$$
 and  $P(\lambda_f(t)\hat{W}_n(t) > x) \to e^{-2x/(c_a^2 + c_s^2)}$  (4.14)

as first  $n \to \infty$  and then  $t \to \infty$ .

The simulation experiments show that the rate-matching control indeed stabilizes the expected queue length (in fact, the entire queue-length distribution), but not the expected waiting time. Simulation results also show that the heavy-traffic approximation resulting from Theorem 4.2.4 works very well as long as the cycle length is not too short (If the cycle length is too short, the queueing process converges to the stationary queue, where the periodicity is lost). The waiting time heavy-traffic approximation for expected stationary waiting time divided by the arrival rate:

$$E[W(t)] \approx \frac{E[W(\infty)]}{\lambda(t)} \approx \frac{\rho^2 (c_a^2 + c_s^2)}{2(1-\rho)\lambda(t)},$$
(4.15)

### 4.2.2 Two Square-Root Controls

We consider two variants of the classical square-root staffing rule for multi-server queues, which lets the time-varying number of servers (staffing) be

$$s(t) \equiv m(t) + \nu_s \sqrt{m(t)}, \quad t \ge 0, \tag{4.16}$$

where  $\nu_s$  is a constant and the m(t) is the offered load, which is the expected number of busy servers in the infinite-server system with the same arrival process and service times. The first square-root control is the direct analog

$$\mu(t) \equiv m(t) + \nu_m \sqrt{m(t)}, \quad t \ge 0,$$
(4.17)

where both m(t) and  $\nu_m$  need to be modified. Since we have time-varying service rates, it is unclear what service times should be used in the infinite-server model. We use the service-requirement distribution directly. For the sinusoidal arrival rate function in (2.4), explicit formulas for m(t) is given in Eick et al. (1993); for exponential service times,  $m(t) = 1 + (\beta/(1 + \gamma^2) (sin\gamma t - \gamma cos\gamma t))$ .

The second variant of (4.16) is (4.17) with  $\lambda(t)$  instead of m(t), i.e.,

$$\mu(t) \equiv \lambda(t) + \nu_{\lambda} \sqrt{\lambda(t)}, t \ge 0, \tag{4.18}$$

Simulations experiments show that these service-rate controls adapted from the multi-server staffing formula stabilize the performance to some extent but are not truly effective for the single-server system.

### 4.2.3 The PSA-Based Square-Root Control

To obtain a service-rate control that is effective for stabilizing the mean waiting time, we start by assuming that the PSA approximation is appropriate. PSA assumes that performance of the queue at time t can be approximated by a stationary queue with parameters taking effect at time t. Specifically, the waiting time and queue length of the time-varying queue at time t can be approximated by a stationary queue with constant arrival rate  $\lambda(t)$  and constant service rate  $\mu(t)$ .

We can thus use a time-varying heavy-traffic approximation for the stationary queue at each time t

$$E[W(t)] \approx \rho(t)v/\mu(t)(1-\rho(t)) = \lambda(t)v/(\mu(t)^2 - \mu(t)\lambda(t)), \quad t \ge 0,$$
(4.19)

where v is a variability parameter, e.g.,  $v = c_a^2 + c_s^2$ ; see §5.1 of Whitt (1983). (For M/GI/1, this is the exact steady-state formula.) To stabilize, we set E[W(t)] = w and obtain a quadratic equation for  $\mu(t)$ , yielding

$$\mu(t) \equiv \lambda(t) + (\lambda(t)/2) \left( \sqrt{(\lambda(t) + \nu_{PSA})/\lambda(t)} - 1 \right), \quad t \ge 0.$$
(4.20)

Simulation experiments verify that this control stabilizes the expected waiting time when the periodic cycles are not too short (when PSA is appropriate), but not when the cycles are short.

We have seen that the rate-matching control stabilizes the queue-length process but not the waiting time process, while PSA-based square-root control approximately stabilizes the expected waiting time in heavy traffic with PSA assumption but not the queue-length process. Therefore two questions naturally occur. The first is whether it is possible to stabilize both the queue-length process and the waiting time process. Theorem 4.2.5 shows any control that stabilizes the waiting time distribution cannot also stabilize the mean number waiting in queue, which proof is based on the time-varying version of Little's Law.

**Theorem 4.2.5.** (Theorem 7.1 from Whitt (2015): impossibility of stabilizing both) Consider a  $G_t/G_t/1$  system starting empty in the distant past. Suppose that a service-rate control makes P(W(t) > x) independent of t for all  $x \ge 0$ . Then the only arrival rate functions for which the mean number waiting in queue  $E[(Q(t)-1)^+]$  is constant, independent of t, are the constant arrival rate functions.

The second question is whether we can develop other controls that can stabilize waiting time in common situations without PSA assumption or not in heavy traffic. We develop a damped time-lag service-rate control in Chapter 5 which performs fairly well in stabilizing virtual waiting time.

### 4.3 The Experiments

### 4.3.1 Estimating Performance Measures

We mainly evaluate two performance measures for each service-rate control, the expected number of customers in the system, E[Q(t)], and the expected virtual waiting time E[W(t)]. The virtual waiting time at time t is defined as the waiting time of a potential or hypothetical arrival (a virtual arrival) at time t, where the waiting time is the time from arrival until starting service.

For each simulation replication, we generate the arrival process and service process using the efficient simulation algorithm developed in Chapter 2. We consider a fixed time interval [0, T] and calculate the performance measures at time points  $k\Delta t, 1 \leq k \leq 1000$ , where  $\Delta t = T/1000$ . We use  $T = 2 \times 10^4$  for  $\gamma = 0.001$  and  $T = 2 \times 10^3$  for the other values of  $\gamma$ . We use the longer time interval for very small  $\gamma$  because we want to see the performance over at least several cycles (which each are of length  $2\pi/\gamma$ ). On the other hand, we cannot only fix the number of cycles, because we need enough absolute time to remove the impact of the initial transient.

To calculate these two performance measures, we first derive values of the cumulative arrival function A(t) and the cumulative departure function D(t) at time points  $j\Delta t, 1 \leq j \leq 1000$  from customers' arrival times  $A_k$  and departure times  $D_k$ . Then we compute Q(t) = A(t) - D(t) and  $W(t) = (W_{A(t)} + V_{A(t)} - (t - A_{A(t)}))^+$  at those time points, where the virtual waiting time W(t) actually depends on information after time t, because the service time  $V_{A(t)}$  may depend on future service-rate function. But this future effect has been properly accounted for, because the service times have already been generated, according to §2.4.

We generate 10,000 independent replications to estimate mean values of performance measures and to construct their confidence intervals at each of those time points. This sample size is large enough to produce reliable estimation. We estimate the mean values E[Q(t)] and E[W(t)] by taking the average over all replications and construct 95% confidence intervals for these mean values. Since we have a very large sample sizes, z-statistics are essentially the same as the natural t-statistics.

### 4.3.2 The Study Cases

### 4.3.2.1 The Rate Functions

We use the sinusoidal arrival rate function in (2.4) with parameters  $\beta = 0.2$  and  $\gamma = 10^{j}$  for  $-3 \leq j \leq 2$  to cover a range of different cycle lengths  $2\pi/\gamma$ . The service rate controls are then as specified in §4.2. For the infinite-server offered load m(t) with this sinusoidal arrival rate function, formulas are given in Eick et al. (1993).

### 4.3.2.2 Interval Distributions for the Base Renewal Processes

We use renewal processes with i.i.d. interval times having mean 1 for the base processes N and  $N_s$  used to construct the arrival and service process. We use the squared coefficient of variation (scv, variance divided by the square of the mean),  $c_a^2$  and  $c_s^2$ , to characterize the variability. We consider three different distributions: exponential ( $c^2 = 1$ ), hyperexponential (mixture of two exponentials,  $H_2$ ,  $c^2 > 1$ ) and Erlang (sums of two i.i.d. exponentials,  $E_2$ ,  $c^2 = 0.5$ ) to represent a range of variability. The  $H_2$  distribution has mean 1 and scv  $c^2 = 4$ , assuming balanced means  $p_1\lambda_1^{-1} = p_2\lambda_2^{-1}$  as in Whitt (1982); it has density  $h(x) = p_1\lambda_1e^{-\lambda_1x} + p_2\lambda_2e^{-\lambda_2x}$ , where  $p_1 = (5 + \sqrt{15})/10$ ,  $p_2 = 1 - p_1$  and  $\lambda_i = 2p_i, i = 1, 2$ . The simulation experiments consider various combinations of these distributions for the arrival and service processes. Some results are for the Markovian  $M_t/M_t/1$  model, while others are for the non-Markovian  $GI_t/GI_t/1$  systems.

### 4.4 Simulation Results

In this section, we display simulation results to show the performance of the different service-rate controls.

### 4.4.1 The Rate-Matching Control

Figures 4.1 and 4.2 show the performance of the rate-matching control for the Markovian  $M_t/M_t/1$ system. Figure 4.1 shows the time-varying means E[Q(t)] and E[W(t)] for three values of  $\gamma$ : 0.001,



Figure 4.1: Estimated E[Q(t)] for the rate-matching control in the  $M_t/M_t/1$  system with different  $\gamma$ : 0.001 (left), 0.1 (middle) and 10 (right).

0.1 and 10.0. In each case we show the performance over three cycles of length  $2\pi/\gamma$ , for which the total length is inversely proportional to  $\gamma$ . We show the 95% confidence interval for E[Q(t)] as well as the estimate itself. Figure 4.1 shows that E[Q(t)] is stabilized in all cases, but E[W(t)] is not. Both means are stabilized for  $\gamma = 10.0$  because the cycles are very short, making the arrival process nearly the same as a homogeneous Poisson process (implied by Theorem 1 of Whitt (1984)).

Figure 4.2 compares the estimated E[W(t)] to the heavy-traffic approximation in 4.15. Figure 4.2 shows that this heavy-traffic approximation works well provided that  $\gamma$  is not too large (the cycles are not too short). A small time-shift error appears at  $\gamma = 0.1$  and significant deviation appears for  $\gamma \geq 1$ . (Above we observed that the rapidly fluctuating arrival rate for the very short cycles makes the model nearly the same as if the arrival rate were constant, equal to its average.)

Figures 4.3 and 4.4 present performance results for E[Q(t)] and E[W(t)], respectively, using the rate-matching control applied to three non-Markovian  $GI_t/GI_t/1$  systems and three values of



Figure 4.2: Comparison of estimated E[W(t)] to its heavy traffic approximation in (4.15) under the rate-matching control in  $M_t/M_t/1$  system with different values of  $\gamma$ : 0.001 (top left), 0.01 (top right), 0.1 (bottom left) and 1 (bottom right).

 $\gamma$ . (We use  $(H_2/E_2)$  to specify that N is a  $H_2$  renewal process, while  $N_s$  is an  $E_2$  renewal process, and similarly for other cases.) As in stationary models, the performance tends to be proportional to the total variability  $c_a^2 + c_s^2$ . Otherwise, the story is essentially the same as for the  $M_t/M_t/1$ model.

### 4.4.2 Two Square-Root Controls

Figures 4.5 and 4.6 presents performance results of the square-root controls related to the manyserver staffing formula applied to the Markovian  $M_t/M_t/1$  model with different values of  $\gamma$ . The


Figure 4.3: Estimated E[Q(t)] for the rate-matching control in three different  $G_t/G_t/1$  models,  $(H_2/H_2)$ ,  $(H_2/E_2)$  and  $(E_2/E_2)$ , and three different values of  $\gamma$ : 0.001 (left), 0.1 (middle) and 10 (right).



Figure 4.4: Estimated E[W(t)] for the rate-matching control in three different  $G_t/G_t/1$  models,  $(H_2/H_2)$ ,  $(H_2/E_2)$  and  $(E_2/E_2)$ , and three different values of  $\gamma$ : 0.001 (left), 0.1 (middle) and 10 (right)

first variant in (4.17) is shown in Figure 4.5, while the second variant in (4.18) is shown in Figure 4.6. When  $\gamma$  gets larger, m(t) is more different from  $\lambda(t)$  and performance is more different for these two controls.

These figures show that neither of these controls is consistently effective. When  $\gamma$  is very small, the offered load m(t) is very close to the arrival rate  $\lambda(t)$ , explaining why the two left-most plots are very similar.



Figure 4.5: Estimated means E[Q(t)] and E[W(t)] for the square-root control in (4.17) for the  $M_t/M_t/1$  system with different values of  $\gamma$ : 0.001 (left), 0.1 (middle) and 1 (right).



Figure 4.6: Estimated means E[Q(t)] and E[W(t)] for the square-root control in (4.18) for the  $M_t/M_t/1$  system with different values of  $\gamma$ : 0.001 (left), 0.1 (middle) and 1 (right).

## 4.4.3 The PSA Square-Root Control

Figure 4.7 shows the results of the PSA square-root service-rate control in (4.20) applied to the  $M_t/M_t/1$  system, while Figure 4.8 shows its application to corresponding  $(H_2/H_2)$ ,  $(H_2/E_2)$ , and  $(E_2/E_2)$ ,  $GI_t/GI_t/1$  systems. When  $\gamma = 0.001$ , so that the cycles are long and arrival rates change very slowly compared to service times, we see that E(W(t)) is stabilized, as intended (while E(Q(t)) is not). When  $\gamma = 0.1$ , so that the cycles are much shorter and PSA is no longer appropriate, E(W(t)) becomes periodic.



Figure 4.7: Estimated E[Q(t)] and E[W(t)] for the PSA square-root control in (4.20) in the  $M_t/M_t/1$  model for two values of  $\gamma$ : 0.001 (left) and 0.1 (right).

## 4.5 Conclusions

In this chapter we conducted simulation experiments evaluating the performance of alternative service-rate controls for single-server queues with time-varying arrival rates. This service-rate control problem arises when there is few number of servers or inflexible staffing. Our time-varying single-server model is an idealized setting of real applications and our results provide guidance on changing staffing/ service-rate/ service on/off for real applications.

In this chapter we have used the efficient simulation algorithm for  $G_t/G_t/1$  queue developed in Chapter 2 to conduct simulation experiments to evaluate the performance of four candidate servicerate controls. The model is a single-server queue with service in order of arrival, unlimited waiting space and a time-varying arrival rate function. The service-rate controls apply to arbitrary arrival



Figure 4.8: Estimated E(W(t)) (left column) and E(Q(t)) (right column) for the PSA square-root control in (4.20) in three  $G_t/G_t/1$  models,  $(H_2/H_2)$ ,  $(H_2/E_2)$  and  $(E_2/E_2)$ , for two values of  $\gamma$ : 0.001 (top) and 0.1 (bottom).

rate functions, but for these experiments we used the sinusoidal periodic arrival rate function in (2.4) with average arrival rate 1, relative amplitude  $\beta = 0.2$  and various time-scaling factors  $\gamma$ . The service requirements were i.i.d. random variables specified separately from the service-rate control. The arrival processes were mostly non-homogeneous Poisson processes, but the method applies to very general arrival processes that can be represented as a deterministic time transformation of a stationary point process. Experiments were conducted for stationary processes constructed from renewal processes with non-exponential as well as exponential distributions. This allows representing different levels of stochastic variability.

The simulation experiments confirmed theoretical results in Theorem 4.2.2 and Theorem 4.2.4 showing that the rate-matching control in (4.1) stabilizes the expected queue length E[Q(t)] after an initial transient period, but not the expected waiting time, and that the heavy-traffic approximation for mean waiting time in (4.15) performs fairly well. Simulation results also showed that the PSAbased square-root service-rate control in (4.20) stabilizes the mean waiting time when the arrival rate function changes slowly (for long cycles relative to the mean service time) so that PSA is effective. The simulation experiments also presented that the other two service-rate controls in (4.17) and (4.18) that are modifications of the classical square-root staffing formula for manyserver queues in (4.16) are not so effective in the present context. From Theorem 4.2.5, we see that it is impossible for any service-rate control to stabilize both the waiting time distribution and the mean number waiting time in queue. Since most of the time we are more interested in stabilizing the mean waiting time, we will develop a damped time-lag service-rate control in Chapter 5 that performs fairly well in stabilizing waiting time when PSA is not appropriate,

## Chapter 5

# Damped Time-Lag Service-Rate Control

We consider a single-server queue with unlimited waiting space, the FCFS discipline, a periodic arrival-rate function and i.i.d. service requirements, where the service-rate function is subject to control. We previously showed in Chapter 4 that a rate-matching control, where the service rate is made proportional to the arrival rate, stabilizes the queue length process, but not the (virtual) waiting time process. In order to minimize the maximum expected waiting time (and stabilize the expected waiting time), we now consider a modification of the service-rate control involving two parameters: a time lag and a damping factors. We develop an efficient simulation search algorithm to find the best time lag and damping factor. That simulation algorithm is an extension of our rareevent simulation algorithm for the  $GI_t/GI/1$  queue (in Chapter 3) to the  $GI_t/GI_t/1$  queue, allowing the time-varying service rate. To gain insight into these controls, we establish a heavy-traffic limit with periodicity in the fluid scale. That produces a diffusion control problem for the stabilization, which we solve numerically by the simulation search in the scaled family of systems with  $\rho \uparrow 1$ . The state space collapse in that theorem shows that there is a time-varying Little's law in heavytraffic, implying that the queue length and waiting time cannot be simultaneously stabilized in this limit. We conduct simulation experiments showing that the new control is effective for stabilizing the expected waiting time for a wide range of model parameters, but we also show that it cannot stabilize the expected waiting time perfectly. This chapter is an edited version of Ma and Whitt (2018a).

## 5.1 Introduction

#### 5.1.1 A Nonstationary Stochastic Design Problem

In this chapter, we address an open problem in Chapter 4, which considered the problem of stabilizing performance over time, i.e., making a specified time-dependent performance measure a target constant function, in a single-server queue with unlimited waiting space, the first-come first-served (FCFS) discipline and a time-varying arrival-rate function. The stabilization is to be achieved with a deterministic service-rate function, under the assumption that the customer service requirements are specified independently of the service-rate control. This is a stochastic design problem instead of a real-time stochastic control problem; i.e., the service-rate control is to be determined in advance, assuming full knowledge of the model, but without knowledge of the system state (e.g., the value of the stochastic queue length process) that will actually prevail at any time.

As explained in Chapter 4, variants of this service rate control are performed in response to timevarying demand, in many service operations, such as hospital surgery rooms and airport inspection lines, but little is known about the ideal timing and extent of service rate changes. Service-rate controls for single-server queues are also of current interest within more complex systems, such as in energy-efficient data centers in cloud computing Kwon and Gautam (2016) and in business process management Suriadi et al. (2017).

In Chapter 4 it was shown that a rate-matching control, where the service rate is made proportional to the arrival rate, stabilizes the queue length process, but not the (virtual) waiting time process. In this chapter we develop an algorithm to approximately stabilize the expected waiting time at a target level. It uses a modification of the service-rate control involving two parameters: a time lag and a damping factor.

## 5.1.2 Related Literature

There is a large literature on similar stochastic design problems involving setting staffing levels (the number of servers) in a multi-server queue to stabilize performance in face of time-varying demand, e.g., Defraeye and van Nieuwenhuyse (2013), Feldman et al. (2008), He et al. (2016), Jennings

et al. (1996), Liu (2018), Liu and Whitt (2012b), Pender and Massey (2017), Stolletz (2008), Whitt (2018). For a single-server queue, the direct analog would be turning on and off the server, which is a restrictive extreme version of the service-rate control we consider.

The dynamic control problem of turning on and off the server in specified system states has received considerable attention in the stationary setting, starting with Yadin and Naor (1963), Heyman (1968). Similar dynamic control problems for single-server queues including service-rate controls have been analyzed as Markov decision processes in Adusumilli and Hasenbein (2010), George and Harrison (2001) and references therein. We emphasize that our design problem is different; our service-rate control is for a nonstationary model and must be set in advance, without knowledge of the system state. In many cases, our new problem is more realistic, because arrival rates are often strongly time-varying and can be reasonably well estimated in advance, while changes to the service rate may be difficult to implement without advance planning. Of course, in general both problems are important.

Given the extensive research on the staffing design problem for many-server queues, it is natural to consider variants of the successful staffing algorithms, but it is now well known that the behavior of many-server queues tends to be dramatically different from single-server queues. That difference can be seen by comparing the many-server fluid models in Liu and Whitt (2012a) to the singleserver fluid models in Chen and Mandelbaum (1991), as discussed on p. 836 of Liu and Whitt (2011). A simple fluid model supporting the rate-matching control in Chapter 4 is supported by our heavy-traffic weak law of large numbers in Theorem 5.6.1, see Corollary 5.6.1, but we are working to go beyond that.

Hence, it should not be surprising that service-rate controls using variants of the established many-server staffing algorithms are no longer effective for single-server queues. For example, a natural analog of the square-root staffing function from Jennings et al. (1996) was considered as a candidate service-rate control in (4.18), but was found to be ineffective, as illustrated by Figure 4.6. Also variants of the iterated staffing algorithm (ISA) in Feldman et al. (2008) and Defraeye and van Nieuwenhuyse (2013) were found to be ineffective, evidently because the controls have impact over greater time intervals (are less "local") with single-server systems.

As indicated in Whitt (2015), controlling the service rate to meet time-varying demand is analogous to Kleinrock's classic service-capacity-allocation problem in a stationary Markovian Jackson network Kleinrock (1964); we allocate service capacity over time instead over space (different queues within the network).

#### 5.1.3 The Rate-Matching Control

Given that the service requirements are specified independently, the actual service times resulting from a time-varying control are relatively complicated, but a construction is given in §2.4. In Chapter 4, several controls were considered, but most attention was given to the *rate-matching control*, which chooses the service rate to be proportional to the arrival rate; i.e., for a given target traffic intensity  $\rho$ , the service-rate function is (4.1). In Chapter 4, Theorem 4.2.2 shows that the rate-matching control stabilizes the queue-length process; Theorem 4.2.3 gives an expression for the waiting-time with the rate-matching control, while Theorems 4.2.4 establishes heavy-traffic limits showing that the queue-length is asymptotically stable, but the waiting time is not, being asymptotically inversely proportional to the arrival-rate function.

## 5.1.4 The Open Problem

The open problem from Chapter 4 is developing a service-rate control that can stabilize the expected waiting time. (We only discuss the continuous-time virtual waiting time process in this chapter, which is the waiting time of a potential or hypothetical customer if it were to arrive at that time, and so omit "virtual.") Toward that end, we now study a modification of the rate-matching control. Without loss of generality, we write the periodic arrival-rate function as

$$\lambda(t) \equiv \rho(1+s(t)), \quad t \ge 0, \tag{5.1}$$

where  $0 < \rho < 1$  and s is a periodic function with period C satisfying

$$\bar{s} \equiv \frac{1}{C} \int_0^C s(u) \, du \equiv 0. \tag{5.2}$$

As a regularity condition, we require that

$$s_L \le s(t) \le s_U$$
 for all  $t$  with  $-1 \le s_L \le 0 \le s_U < \infty$ . (5.3)

Most of our numerical examples will be for a sinusoidal function, where  $s(t) = \beta \sin(\gamma t)$  for s(t)in (5.1), so that we have (2.4), where  $\beta$  is the relative amplitude, with  $0 \le \beta \le 1$  and the period is  $C = 2\pi/\gamma$ . But the damped time-lag control developed in this chapter is not limited to the sinusoidal case. The simulation optimization algorithm illustrated later in section 5.5.4 applies to more general arrival rate functions and section 5.9.1.2 shows the performance of the damped time-lag control for a piecewise-linear single-peak arrival rate function.

In the periodic setting of (5.1)-(5.3), we consider the rate-matching control in (4.1) modified by a *time lag*  $\eta$  and *damping factor*  $\xi$ ; in particular,

$$\mu(t) \equiv 1 + \xi s(t - \eta), \quad t \ge 0, \tag{5.4}$$

for  $0 < \xi \leq 1$  and  $\eta > 0$ . Thus, the average arrival rate and service rate are  $\bar{\lambda} = \rho$  and  $\bar{\mu} = 1$ , so that the long-run traffic intensity is  $\bar{\rho} \equiv \bar{\lambda}/\bar{\mu} = \rho$ . However, the instantaneous traffic intensity  $\rho(t) \equiv \lambda(t)/\mu(t)$  can satisfy  $\rho(t) > 1$  for some t in each periodic cycle, e.g., if  $\rho(1+\beta) > 1-\beta$  or, equivalently, if  $\beta > (1-\rho)/(1+\rho)$  in the setting of (2.4) and (5.4).

## 5.1.5 Formulation of Optimal Control Problems

Because it is directly of interest, and because we want to allow for imperfect stabilization, we formulate our control problem as minimizing the maximum expected waiting time over a periodic cycle [0, C]. We formulate the main optimization problem as a min-max problem, i.e.,

$$w^* \equiv \min_{\mu(t) \in \mathcal{M}(1)} \max_{0 \le y \le 1} \{ E[W_y] \},$$
(5.5)

where  $E[W_y]$  is the expected (periodic) steady-state (virtual) waiting time starting at time yCwithin a cycle of length C,  $0 \le y < C$ , and  $\mathcal{M}(m)$  is the set of all periodic service-rate functions with average rate m, which we take to be  $m \equiv 1$ .

Given that the average arrival rate is  $\rho < 1$ , the obvious reference case is the mean waiting time E[W] in the associated stationary model, which for the M/GI/1 model is

$$E[W] = \frac{\rho(1+c_s^2)}{2(1-\rho)}$$
(5.6)

and thus  $E[W] = \rho/(1-\rho)$  in the M/M/1 model. However, in general E[W] is not a lower bound for the average of the periodic steady-state mean  $E[W_y]$  over a cycle; see Remark 5.2.1 and Example 5.2.1.

We have not yet solved this general optimization problem in (5.5). Here are open problems, applying to the Markovian  $M_t/M_t/1$  model and generalizations:

- 1. For the general periodic problem, what is the solution (value of  $w^*$  and set of optimal servicerate functions  $\mu^*(t)$  as a function of the model)?
- 2. For the sinusoidal special case in (2.4), what is the solution?
- 3. To what extent do the optimal solutions stabilize the expected waiting time  $E[W_y]$  over time? In particular, is it possible to stabilize  $E[W_y]$  perfectly?

**Remark 5.1.1.** (stabilizing the full waiting time distribution) Theorems 4.2.2 shows that the ratematching control stabilizes the delay probability  $P(W_y > 0)$ , while Corollary 5.1 of Whitt (2015) shows that the rate-matching control cannot stabilize the mean waiting time. Theorem 4.2.4 establishes a heavy-traffic limit (with periodicity in the stronger fluid scale in §5.6 here) that shows that it is not possible to stabilize the queue length and waiting time processes simultaneously. Thus, we conclude that it is not possible to stabilize the full waiting time distribution. Hence, the open problems above are only for the mean. In this chapter we primarily focus on the mean, but we also show that it stabilizes the entire distribution to some extent in §5.9.1.

In this chapter, we only consider the restricted set of controls in (5.4). Now our goal is

$$w^{*}(\eta,\xi) \equiv \min_{\eta,\xi} \max_{0 \le y \le 1} \{ E[W_{y}] \}.$$
(5.7)

For practical purposes, this two-parameter control is appealing for its simplicity. We also find that it is quite effective, even though it cannot stabilize  $E[W_u]$  perfectly.

We also consider the associated stabilization control, where (5.7) is replaced by

$$w_{stab}^*(\eta,\xi) \equiv \min_{\eta,\xi} \{ \max_{0 \le y \le 1} \{ E[W_y] \} - \min_{0 \le y \le 1} \{ E[W_y] \} \}.$$
(5.8)

In our sinusoidal examples, where there is strong symmetry, we find that the solutions to (5.7) and (5.8) are the same (but we have no proof), but neither stabilizes perfectly. For more general periodic arrival rate functions, we detect differences.

## 5.1.6 Organization of the Chapter

This chapter involves some challenging technical methods. Hence, we present the more accessible results first. We start in  $\S5.2$  by presenting two simulation examples to illustrate the effectiveness

of our new algorithm. Then in §5.3 we introduce the two technical tools we will apply: (i) an extension of the rare-event simulation algorithm for the  $GI_t/GI/1$  model from Chapter 3 to the  $GI_t/GI_t/1$  model with a general service-rate control and (ii) heavy-traffic limits involving scaling of the underlying deterministic periodic arrival-rate function.

We start in earnest in §5.4. We elaborate on the model and key processes representing the workload and the waiting time in §5.4. Theorem 5.4.1 shows that the rate-matching control stabilizes the workload process as well as the queue length process. We discuss the extension of the rare-event simulation algorithm from Chapter 3to our setting and its application to perform simulation search in §5.5. In both §5.4 and §5.5 we will be brief because we can draw upon Chapter 4 and Chapter 3.

We establish our main heavy-traffic limits with the periodicity in the stronger fluid scaling (see (5.11)) in §5.6. We present the proof of the main heavy-traffic limit, Theorem 5.6.2, in §5.7. We establish heavy-traffic limits with the periodicity in the weaker diffusion scaling (see (5.12)) in §5.8.

We give simulation examples in  $\S5.9$ . In  $\S5.9.1$  we present simulation results using the fluid scaling in  $\S5.6$ ; in  $\S5.9.2$  we present simulation results using the diffusion scaling in  $\S5.8$ . We draw conclusions in  $\S5.10$ .

## 5.2 Simulation Examples

To illustrate the effectiveness of our new algorithm, we show results for two simulation examples. We consider the Markovian  $M_t/M_t/1$  model with the sinusoidal arrival rate function in (5.1)-(5.3) and (2.4). The first example has model parameters  $(\rho, \beta, \gamma) = (0.8, 0.2, 0.1)$ , so that the average arrival rate is  $\bar{\rho} = 0.8$ , the average service time is 1 and the cycle length is  $C = 2\pi/\gamma = 62.8$ . Figure 5.1 (left) shows the expected steady-state waiting time  $E[W_y]$  together with the corresponding expected workload  $E[L_y]$  and the product  $\lambda(y)E[W_y]$ , all for  $0 \leq y < 1$ . The second example on the right differs only by increasing  $\rho$  from 0.8 to 0.95. Figure 5.1 also shows the upper and lower 95% confidence-interval bounds for  $E[L_y]$  and  $E[W_y]$  with black dashed lines, but these can only be seen by zooming in.

Figure 5.1 shows that the expected waiting time  $E[W_y]$  is well stabilized at a value somewhat higher than the expected steady-state waiting time for the stationary M/M/1 model, which is  $\rho/(1-\rho)$  (4 on the left and 19 on the right). The maximum deviation (maximum - minimum)



Figure 5.1: Estimates of the periodic steady-state values of  $E[W_y]$  (blue solid line),  $E[L_y]$  (red dashed line) and  $\lambda(y)E[W_y]$  (green dotted line) for the optimal control  $(\eta^*, \xi^*)$  for the sinusoidal example with parameter triples  $(\rho, \beta, \gamma) = (0.8, 0.2, 0.1)$  (left) and (0.95, 0.2, 0.1) (right), so that the cycle length is  $C = 2\pi/\gamma = 62.8$ . The optimal controls are (5.84, 0.84) for  $\rho = 0.8$  and (15.1, 2.13) for  $\rho = 0.95$ .

over a cycle is 0.0335 is for  $\rho = 0.8$  and 0.4653 for  $\rho = 0.95$ . Thus the maximum relative errors are about 0.8% for  $\rho = 0.8$  and 2.2% for  $\rho = 0.95$ , clearly adequate for practical applications. Nevertheless, careful simulations and statistical analysis allow us to conclude that it is impossible to stabilize the expected waiting time perfectly with this control. To see the contrasting view with the rate-matching control for this same model, see 4.1. (See Chapter 4 for more examples.)

It is natural to wonder if there is any order in the optimal controls found for  $\rho = 0.8$  and  $\rho = 0.95$  in Figure 5.1. The dependence on  $\rho$  is revealed by the main heavy-traffic limit theorem, Theorem 5.6.2.

Remark 5.2.1. (the cost of periodicity) The difference between the stable average waiting time in Figure 5.1 and the value  $\rho/(1-\rho)$  for the stationary model (4 on the left and 19 on the right) might be called "the average cost of periodicity," but we point out that the overall average waiting time with a service-rate control could be much less than in the stationary model. The classical results for the periodic  $M_t/GI/1$  queue in Rolski (1981, 1989a) so not apply because, in general, the service times are neither independent of the arrival process nor i.i.d.; See Example 5.2.1. **Example 5.2.1.** (small expected waiting times with periodicity) To illustrate a nonstationary model with a low average expected waiting time, consider the  $M_t/M_t/1$  model with the two-level arrival-rate function with period C:

$$\lambda(t) \equiv \rho b \mathbb{1}_{[(C/2) - \delta, (C/2) + \delta)}(t), \quad 0 \le t < C \quad \text{and} \quad b\delta = C,$$
(5.9)

where  $\delta < C/2$  and  $1_A$  is the indicator function of the set A, i.e.,  $1_A(t) = 1$  if  $t \in A$  and 0 otherwise. Let the service-rate function be as in (5.4) with  $\eta = 2\delta$  and  $\xi = 1$ . Then the number of arrivals in the interval  $[(C/2) - \delta, (C/2) + \delta)$  has a Poisson distribution with mean  $\rho C$ , while the number of potential departures in the interval  $[(C/2) + \delta, (C/2) + 3\delta)$  has a Poisson distribution with mean C. Thus, for  $\rho < 1$  and  $C = b\delta$  suitably large, the net input over the interval  $[(C/2) - \delta, (C/2) + \delta)$  is approximately Gaussian with mean  $-(1 - \rho)b\delta$  and variance  $(1 + \rho)b\delta$ , which is unlikely to be positive. By choosing  $\delta$  suitably small and  $b\delta$  suitably large, subject to specified  $\rho$ , we can make the maximum steady-state expected waiting time, and thus the average, approach 0. One way to explain this phenomenon is to observe that the interarrival times and service times will be highly correlated.

Remark 5.2.2. (the single-parameter alternative) It is natural to wonder if we could use only the single control parameter  $\eta$ , fixing  $\xi = 1$ . If we let  $\xi = 1$  and optimize over  $\eta$  in the setting of Figure 5.1, then for  $\rho = 0.8$  ( $\rho = 0.95$ ) we get  $\eta^* = 5.93$  ( $\eta^* = 28.3$ )and a maximum deviation of 0.4109 (3.034), which yields about 10% (14%) relative error instead of 0.8% (2.2%). Hence, we use the two control parameters.

## 5.3 The Key Technical Tools

In this section we discuss the two technical tools that we use.

#### 5.3.1 A Simulation Search Algorithm

Our primary tool for finding good  $(\eta, \xi)$  controls is a simulation search algorithm. For that purpose, we extend the rare-event simulation algorithm for the time-varying workload process in the periodic  $GI_t/GI/1$  model in Chapter 3 to the  $GI_t/GI_t/1$  model, where the service rate is time-varying as well. (The notation  $GI_t$  means that the process is a deterministic time transformation of a renewal process; see §5.4.) The workload L(t) represents the amount of work in service time in the system at time t, while the waiting time can be represented as the first-passage time

$$W(t) = \inf \{ u \ge 0 : \int_{t}^{t+u} \mu(s) \, ds = L(t) \}.$$
(5.10)

The waiting time W(t) coincides with the workload L(t) when  $\mu(t) = 1$  for all t, but not otherwise.

As in Chapter 3, the rare-event simulation algorithm calculates the periodic steady-state workload  $L_y$  and waiting time  $W_y$ , starting at time yC within a cycle of length C,  $0 \le y < 1$ . We employ a search over the parameters  $(\eta, \xi)$ , as discussed in §5.5, in order to solve the optimization problems (5.7) and (5.8). The search part is relatively elementary because we have only two control parameters. For background on simulation optimization, see Fu (2015), Jian and Henderson (2015) and the references there.

The computational complexity for one control vector  $(\eta, \xi)$  is essentially the same as in Chapter 3. In particular, the program running time tends to be proportional to the number of replications and number of y values, which for the case  $\rho = 0.8$  in Figure 5.1 were taken to be 40,000 and 40, respectively. That required about 100 minutes on a desktop computer. As indicated in §3.4.7, the running time tends to be of order  $(1 - \rho)^{-1}$ , so that the cases with high traffic intensity are more challenging. The simulation search is performed in stages, with fewer y values and replications in the early stages, but the full long run at the end to confirm performance.

#### 5.3.2 Heavy-Traffic Limits

To better understand how the control parameters and performance depend on the model parameters, we establish heavy-traffic (HT) limits, which involve considering a family of models indexed by  $\rho$  and letting  $\rho \uparrow 1$ , drawing on our previous work in Whitt (2014, 2015) and in Chapter 3. That previous work shows that the scaling is very important, because there are several possibilities. We use the conventional HT scaling of time by  $(1 - \rho)^{-2}$  (usually denoted by n) and space by  $1 - \rho$ (usually denoted by  $1/\sqrt{n}$ ), as in Chapters 5 and 9 of Whitt (2002), but if we do so without also scaling the arrival-rate function, then the HT limit is easily seen to be the same as if the periodicity were replaced by the constant long-run average, as shown by Falin Falin (1989).

To obtain insight into the periodic dynamics, it is thus important to also scale the arrival-rate function, which is initially specified in (5.1) with (5.2) and (5.3). However, the papers Whitt (2014)

and Whitt (2015) actually use two different HT scalings of the arrival-rate function. Our main HT scaling in  $\S5.6$  follows Whitt (2015) and has periodicity in the fluid scale, i.e.,

$$\lambda_{\rho}(t) \equiv \rho(1 + s((1 - \rho)^2 t)), \quad t \ge 0,$$
(5.11)

but in  $\S5.8$  we also consider the scaling from Whitt (2014) and  $\S3.6$ , which has the periodicity in diffusion scale, i.e.,

$$\lambda_{\rho}(t) \equiv \rho(1 + (1 - \rho)s((1 - \rho)^2 t)), \quad t \ge 0.$$
(5.12)

The extent of the periodicity is stronger in (5.11) than in (5.12), because of the extra factor  $(1 - \rho)$  before s in (5.12). The workload and the waiting time have the same HT limit with the diffusion-scale scaling in (5.12), but different limits with the fluid-scale scaling in (5.11). To capture the clear differences shown in Figure 5.1, we obviously want the stronger fluid scaling in (5.11). The HT functional central limit theorem (FCLT) in Theorem 5.6.2 for the scaling in (5.11) in §5.6 helps interpret Figure 5.1.

It is important to note that if we have constant service rate with this scaling, then the waiting times explode as  $\rho \uparrow 1$ , because the instantaneous traffic intensity  $\rho(t) \equiv \lambda(t) > 1$  over intervals growing as  $\rho \uparrow 1$ ; this case is analyzed in Choudhury et al. (1997).

We also establish a HT functional weak law of large numbers (FWLLN) in Theorem 5.6.1, which yields a deterministic fluid approximation. However, it is not very useful, because it shows that our proposed control with  $\xi = 1$  stabilizes the waiting time perfectly for all  $\eta$  as  $\rho \uparrow 1$  (But it helps to see that nothing bad happens.)

## 5.4 The Model

In this section we specify the general model, defining the arrival process in §5.4.1 and the basic queueing stochastic processes in §5.4.2. We specialize to the periodic  $G_t/G_t/1$  model in §5.4.3. We show that the workload is stabilized by the rate-matching control in (4.1), extending the results for the queue-length process in Whitt (2015).

#### 5.4.1 The Arrival Process

We represent the periodic arrival counting process A as a deterministic time transformation of an underlying rate-1 counting process N with associated sequence of interarrival times  $\{U_k : k \ge 1\}$  by (2.1) and (2.2). This is a common representation when N is a rate-1 Poisson process; then A is a nonhomogeneous Poisson process (NHPP). For the  $G_t/G_t/1$  model, N is understood to be a rate-1 stationary point process. Hence, for the  $GI_t/GI_t/1$  model, N is an equilibrium renewal process with time between renewals having mean 1, for which the first inter-renewal time  $U_1$  has the equilibrium distribution. The representation in (2.2) has been used frequently for processes N more general than NHPP's, an early source being by Massey and Whitt (1994).

For the sinusoidal arrival-rate function in (2.4), the associated cumulative arrival-rate function is (2.6).

We only consider the case  $\rho < 1$ , under which a proper steady-state exists under regularity conditions (which we do not discuss here). Behavior differs for short cycles and long cycles. For the case of a constant service rate, there are two important cases for the relative amplitude: (i)  $0 < \beta < \rho^{-1} - 1$  and (ii)  $\rho^{-1} - 1 \leq \beta \leq 1$ . In the first case, we have  $\rho(t) < 1$  for all t, where  $\rho(t) \equiv \lambda(t)$  is the instantaneous traffic intensity, but in the second case we have intervals with  $\rho(t) \geq 1$ , where significant congestion can build up. If there is a long cycle as well, the system may be better understood from fluid and diffusion limits, as in Choudhury et al. (1997). However, that difficulty can be avoided by a service-rate control.

## **5.4.2** The General $G_t/G_t/1$ Model

We consider a modification of the standard single-server queue with unlimited waiting space where customers are served in order of arrival. Let  $\{V_k\}$  be the sequence of service requirements. As in §2.2, we separately define the rate at which service is performed from the service requirement. Given the arrival counting process A(t) defined in §5.4.1, let the total input of work over the interval [0, t] be the random sum as in (3.4).

Let service be performed at time t at rate  $\mu(t)$  whenever there is work to perform. Paralleling the cumulative arrival rate  $\Lambda(t)$  defined in (2.1), let the cumulative available service rate be

$$M(t) \equiv \int_0^t \mu(s) \, ds, \quad t \ge 0.$$
 (5.13)

Let the net-input process of work be  $X(t) \equiv Y(t) - M(t), t \ge 0$ . Then we can apply the reflection map to the net input process X(t) to represent the workload (the remaining work in service time) at time t, starting empty at time 0, as

$$L(t) = X(t) - \inf \{X(s) : 0 \le s \le t\} = \sup \{X(t) - X(s) : 0 \le s \le t\}, \quad t \ge 0.$$

In this setting it is elementary that the continuous-time (virtual) waiting time (before starting service) at time t, which we denote by W(t), can be related to L(t).

**Lemma 5.4.1.** (waiting time representation) The waiting time at time t can be represented as

$$W(t) = M_t^{-1}(L(t)), \quad t \ge 0,$$
(5.14)

where  $M_t^{-1}$  is the inverse of  $M_t(u) \equiv M(t+u) - M(t)$  for M(t) in (5.13).

**Proof** By definition,

$$W(t) = \inf \{ u \ge 0 : \int_{t}^{t+u} \mu(s) \, ds = L(t) \}$$
  
=  $\inf \{ u \ge 0 : M(t+u) - M(t) = L(t) \} = M_{t}^{-1}(L(t)),$  (5.15)

for  $M_t(u)$  above, as claimed in (5.14).

## 5.4.3 The Periodic $G_t/G_t/1$ Model

As in Chapter 3, we consider the periodic steady state of the periodic  $G_t/G_t/1$  model with arrivalrate function in (5.1). For that purpose, we exploit the arrival process construction in (2.2) in terms of the stationary processes  $N \equiv \{N(t) : t \ge 0\}$  and  $V \equiv \{V_k : k \ge 1\}$ . Let the associated service-rate function  $\mu(t)$  also be periodic with cycle length C, with average service rate be  $\bar{\mu} = 1$ , and bounds  $0 < \mu_L \le \mu(t) \le \mu_U < \infty$ , for  $0 \le t \le C$ .

As in Chapter 3 and earlier in Loynes (1962) and Chapter 6 in Sigman (1995), we now convert the standard representation of the workload process in §5.4 to a simple supremum by using a reverse-time construction. To do so, we extend the stationary processes  $\{N(t)\}$  and  $\{V_k\}$  to the entire real line. We regard the periodic arrival-rate and service-rate as defined on the entire real line as well, with the functions fixed by their position within the periodic cycle at time 0. With those conditions, the reverse-time construction is achieved by letting the interarrival times and service times be ordered in reverse time going backwards from time 0. Then  $\tilde{A}(t)$  counts the number of arrivals in [-t, 0],  $\tilde{Y}(t)$  is the total input in [-t, 0] and  $\tilde{X}(t)$  is the net input in [-t, 0], for  $t \ge 0$ . To exploit the reverse-time representation, let  $\tilde{\Lambda}_y(t)$  as in (3.2) be the reverse-time cumulative arrival-rate function starting at time yC within the periodic cycle [0, C],  $0 \le y < 1$ , and  $\tilde{\Lambda}_y^{-1}$  is its inverse function, which is well defined because  $\tilde{\Lambda}_y(t)$  is continuous and strictly increasing.

As an analog of (3.2) for the cumulative service rate, let

$$\tilde{M}_y(t) \equiv M(yC) - M(yC - t), \quad t \ge 0,$$
(5.16)

We let the service requirements  $V_k$  come from a general stationary sequence with  $E[V_k] = 1$ .

With this reverse-time representation, the workload at time yC in the system starting empty at time yC - t can be represented as

$$L_{y}(t) = \sup_{0 \le s \le t} \{\tilde{X}_{y}(s)\}$$
  
$$\stackrel{d}{=} \sup_{0 \le s \le t} \left\{ \sum_{k=1}^{N(\tilde{\Lambda}_{y}(s))} V_{k} - \tilde{M}_{y}(s) \right\} = \sup_{0 \le s \le \tilde{\Lambda}_{y}(t)} \left\{ \sum_{k=1}^{N(s)} V_{k} - \tilde{M}_{y}(\tilde{\Lambda}_{y}^{-1}(s)) \right\}, \quad (5.17)$$

where  $\tilde{X}_y$  is the reverse-time net input of work starting at time yC within the cycle of length C. The other quantities in (5.17) are the reverse-time cumulative arrival-rate function  $\tilde{\Lambda}_y(t)$  in (3.2) with inverse  $\tilde{\Lambda}_y^{-1}(t)$  and the reverse-time cumulative service-rate function  $\tilde{M}_y$  in (5.16) with inverse  $\tilde{M}_y^{-1}$ .

The equality in distribution in (5.17) holds because N is a stationary point process, which is a point process with stationary increments and a constant rate.

As  $t \to \infty$ ,  $L_y(t) \uparrow L_y(\infty) \equiv L_y$  w.p.1 as  $t \to \infty$ , for

$$L_y \stackrel{d}{=} \sup_{s \ge 0} \Big\{ \sum_{k=1}^{N(s)} V_k - \tilde{M}_y(\tilde{\Lambda}_y^{-1}(s)) \Big\}, \qquad 0 \le y < 1.$$
(5.18)

Even though (5.17) is valid for all t, we think of the system starting empty at times -kC, for  $k \ge 1$ , so that we let yC - t = -kC or, equivalently, we stipulate that t = C(k + y),  $0 \le y < C$ , and consider successive values of k and let  $k \to \infty$  to get (5.18). That makes (5.17) valid to describe the distribution of L(C(k + y)) for all  $k \ge 1$ .

We now observe that the time transformation in (5.17) shows that the periodic  $G_t/G_t/1$  model is actually equivalent to a  $G/G_t/1$  model with a stationary arrival process and a new cumulative service rate function  $\tilde{M}_y(\tilde{\Lambda}_y^{-1}(t))$ . **Corollary 5.4.1.** (conversion of  $G_t/G_t/1$  to an equivalent  $G_t/G/1$ ) In addition to representing the periodic steady-state workload  $L_y$  in a periodic  $G_t/G_t/1$  model as a periodic steady-state workload in a periodic  $G/G_t/1$  model, which has a stationary stochastic input and a deterministic service rate, as shown in (5.18) above, we can represent it as a periodic steady-state workload in a periodic  $G_t/G/1$  model, which has a periodic stochastic input and a constant service rate, via

$$L_y = \sup \left\{ \sum_{k=1}^{N(\tilde{\Lambda}_y(\tilde{M}_y^{-1}(s)))} V_k - s : s \ge 0 \right\}.$$
 (5.19)

**Corollary 5.4.2.** (the associated periodic steady-state waiting time) The periodic steady-state waiting time associated with the periodic steady-state workload in (5.18) is

$$W_y = \tilde{M}_y^{-1}(L_y), \quad 0 \le y < 1.$$
 (5.20)

**Proof** Apply the reasoning of Lemma 5.4.1.

In Whitt (2015) we showed that the rate-matching service-rate control in (4.1) stabilizes the queue-length process. Now we establish the corresponding result for the workload.

**Theorem 5.4.1.** (stabilizing the periodic workload) If the rate-matching control in (4.1) is used, then  $L_y \stackrel{d}{=} L$  for  $L_y$  in (5.18), where L is the steady-state workload in the associated (stable) stationary G/G/1 model, i.e.,

$$L \stackrel{\mathrm{d}}{=} \sup_{s \ge 0} \Big\{ \sum_{k=1}^{N(s)} V_k - \rho^{-1} s \Big\},$$
(5.21)

which is independent of y.

**Proof** With the rate matching control, we have  $M(t) = C\Lambda(t)$  and  $\tilde{M}_y(t) = C\tilde{\Lambda}_y(t), t \ge 0$ . As a consequence,  $\tilde{M}_y(\tilde{\Lambda}_y^{-1}(t)) = Ct, t \ge 0$ , so that

$$L_{y} \stackrel{d}{=} \sup_{s \ge 0} \left\{ \sum_{k=1}^{N(s)} V_{k} - \tilde{M}_{y}(\tilde{\Lambda}_{y}^{-1}(s)) \right\}$$
$$\stackrel{d}{=} \sup_{s \ge 0} \left\{ \sum_{k=1}^{N(s)} V_{k} - Cs \right\} \stackrel{d}{=} L. \quad \bullet \tag{5.22}$$

## 5.5 The Simulation Search Algorithm

The rare-event simulation algorithm from Chapter 3 exploits the classic rare-event simulation algorithm for the GI/GI/1 queue, exploiting importance sampling using an exponential change of measure, as in Ch. XIII of Asmussen (2003) and Ch. VI of Asmussen and Glynn (2007). Hence our simulation algorithm applies to the  $GI_t/GI_t/1$  queue. It was shown in Chapter 3 that the algorithm is effective for estimating the mean as well as small tail probabilities. (Also see Chapter 2.)

## 5.5.1 The $GI_t/GI_t/1$ Model

In the  $GI_t/GI_t/1$  setting, the underlying rate-1 process N is an equilibrium renewal process, which means that  $U_1$  has the stationary-excess or equilibrium distribution  $U_e$ , which may be different from the i.i.d. distributions of  $U_k$ ,  $k \ge 2$ . Also in the  $GI_t/GI_t/1$  setting, the service times  $V_k$ 's are i.i.d. with distribution V, and are independent of the arrival process.

The simulation algorithm exploits the discrete-time representation of the workload  $L_y$  in (5.18) and the waiting time  $W_y$ , i.e.,

$$L_{y} \stackrel{d}{=} \sup_{s \ge 0} \left\{ \sum_{k=1}^{N(s)} V_{k} - \tilde{M}_{y}(\tilde{\Lambda}_{y}^{-1}(s)) \right\}$$
$$\stackrel{d}{=} \sup_{n \ge 0} \left\{ \sum_{k=1}^{n} V_{k} - \tilde{M}_{y}(\tilde{\Lambda}_{y}^{-1}(\sum_{k=1}^{n} U_{k})) \right\},$$
$$W_{y} \stackrel{d}{=} M_{y}^{-1}(L_{y}), \qquad 0 \le y < 1.$$
(5.23)

where  $M_y$  is the same as  $M_t$ , which is the forward integral of the service rate starting from position y within a cycle.

We exploit the rare-event simulation algorithm in Chapter 3, which is based on an exponential change of measure. In that setting, we use the underlying measure  $P_{\theta^*}$  determined for GI/GI/1queue. We again use the same notations  $X_k(\rho) = V_k - \rho^{-1}U_k$  and partial sum process  $S_n \equiv \sum_{k=1}^n X_k$ for GI/GI/1 and define the new associated process

$$Q_n \equiv \sum_{k=1}^n V_k - \tilde{M}_y(\tilde{\Lambda}_y^{-1}(\sum_{k=1}^n U_k)),$$

which is the process inside the supremum function. To avoid duplication of notation, we let the likelihood function here be denoted by  $\Psi$  instead of L. Then the estimator of the rare-event

probability for  $W_y$  can be derived as below:

$$P(W_{y} > b) = P(M_{y}^{-1}(L_{y}) > b) = P(L_{y} > M_{y}(b))$$

$$= P(\tau_{M_{y}(b)}^{Q} < \infty) = E_{\theta^{*}}[\Psi_{\tau_{M_{y}(b)}^{Q}}(\theta^{*})]$$

$$= E_{\theta^{*}}[m_{X_{1}}(\theta^{*})m_{X}(\theta^{*})^{(\tau_{M_{y}(b)}^{Q}-1)}\exp(-\theta^{*}S_{\tau_{M_{y}(b)}^{Q}})]$$

$$= m_{X_{1}}(\theta^{*})E_{\theta^{*}}[\exp(\theta^{*}S_{\tau_{M_{y}(b)}^{Q}})], \qquad (5.24)$$

where  $\tau_{M_y(b)}^Q$  is the stopping time of process  $Q_n$  at level  $M_y(b)$ ,  $\Psi_{\tau_{M_y(b)}^Q}(\theta^*)$  is the exponentially tilted likelihood ratio for process  $Q_n$  at  $n = \tau_{M_y(b)}^Q$ ,  $m_X(\theta^*)$  is the moment generating function of X at  $\theta^*$ . The first  $X_1(\rho)$  in the partial sum  $S_{\tau_{M_y(b)}^Q}$  has a different distribution from  $\{X_k, k \ge 2\}$ .

## 5.5.2 The Extended Algorithm

Here is a summary of the extended algorithm to estimate the tail probabilities in the  $GI_t/GI_t/1$ queue with average service rate 1 and average arrival rate  $\rho$ :

- 1. Construct a table of the inverse cumulative arrival-rate function  $\rho \tilde{\Lambda}_{y}^{-1}$  (same as for  $GI_{t}/GI/1$ ).
- 2. Determine the required length of partial sums  $(n_s)$  needed in each application (same as for  $GI_t/GI/1$ ).
- 3. For each replication, we generate the required vectors of exponentially tilted interarrival times  $\rho^{-1}\tilde{U}$  and service times  $\tilde{V}$  from  $F_{\rho^{-1}U}^{-\theta^*}$  and  $F_V^{\theta^*}$  respectively (same as for  $GI_t/GI/1$ ).
- 4. Calculate the associated vectors of  $S_n$  and  $Q_n$  and find out the stopping time  $\tau_{M_y(b)}^Q$ , which is the hitting time of  $Q_n$  at level  $M_y(b)$ . This step is different from for  $GI_t/GI/1$  in that first we need to calculate  $M_y(b)$  as the hitting level instead of b and second we calculate vector  $Q_n$  different from  $R_n$  in an additional function  $\tilde{M}_y$  in the second term.
- 5. Use the above estimator to calculate the tail probability  $P(W_y > b)$  for each replication (same as for  $GI_t/GI/1$ ).
- 6. Run N i.i.d. replications and calculate the mean of the estimated values of  $P(W_y > b)$  (same as for  $GI_t/GI/1$ ).

#### 5.5.3 Explicit Representations for the Sinusoidal Case

Here we summarize the expressions for all the basic deterministic rate functions in our sinusoidal examples, extending (2.4), (5.4) and (2.6):

$$\tilde{\Lambda}_{y}(t) = \rho(t + \frac{\beta}{\gamma}(\cos(\gamma(t - yC)) - \cos(\gamma yC)))$$

$$M(t) = t - \xi \frac{\beta}{\gamma}(\cos(\gamma(t - \eta)) - \cos(\gamma \eta))$$

$$M_{y}(t) = t - \xi \frac{\beta}{\gamma}(\cos(\gamma(t + yC - \eta)) - \cos(\gamma(yC - \eta)))$$

$$\tilde{M}_{y}(t) = t + \xi \frac{\beta}{\gamma}(\cos(\gamma(t + \eta - yC)) - \cos(\gamma(\eta - yC))).$$
(5.25)

## 5.5.4 The Search Algorithm

We use an elementary iterative search algorithm, fixing an initial value of  $\eta$  at the mean for the steady-state model,  $\rho/(1-\rho)$ , and searching first over  $\xi$  and then over each variable iteratively until we get negligible improvement. That simple approach is substantiated by estimating the structure of the objective function. Figure 5.2 illustrates by showing the maximum waiting time  $\max_{0 \le y \le C} \{E[W_y]\}$  in the setting of Figure 5.1 (left). Figure 5.2 shows estimates of the maximum waiting time  $\max_{0 \le y \le C} \{E[W_y]\}$  as a function of  $(\eta, \xi)$  in  $[0, 20] \times [0, 5]$  (left)  $[3, 9] \times [0.6, 1.0]$  (right) in that setting. Figure 5.2 shows that the function is not convex as a function of  $\eta$ , but suggests that it is unimodal with a unique global minimum, supporting our simple procedure. The plots for the maximum deviation  $\max_{0 \le y \le C} \{E[W_y]\} - \max_{0 \le y \le C} \{E[W_y]\}$  are similar.

We perform the search with fewer points y and replications in the initial stages, and then confirm with more points, 40 values of y and 40,000 replications, which yields excellent statistical precision, as can be seen from the narrow confidence interval bands in Figure 5.1.

## 5.6 Supporting Heavy-Traffic Limits

In this section we obtain a heavy-traffic (HT) functional weak law of large numbers (FWLLN) and a HT functional central limit theorem (FCLT) for the periodic  $G_t/G_t/1$  model with a general servicerate control of the form in (5.4). The HT FCLT produces a limit depending on an asymptotic time lag  $\hat{\eta}$  and damping factor  $\hat{\xi}$ , which arise from HT limits; see condition (5.52) in Theorem 5.6.2



Figure 5.2: Three-dimensional plots of estimates of the maximum waiting time  $\max_{0 \le y \le C} \{E[W_y]\}$  for  $(\eta, \xi)$  in  $[0, 20] \times [0, 5]$  (left)  $[3, 9] \times [0.6, 1.0]$  (right).

and the conclusion in (5.44). Thus we reduce the optimization problems over the parameter pairs  $(\eta_{\rho}, \xi_{\rho})$  in (5.7) and (5.8), asymptotically as  $\rho \uparrow 1$ , to diffusion control problems with the parameter pairs  $(\hat{\eta}, \hat{\xi})$ .

## 5.6.1 The Underlying Rate-One Processes

As in much of the HT literature, we start by introducing basic rate-1 stochastic processes, but here we consider service requirements instead of service times. We assume that the rate-1 arrival and service-requirements processes N and V specified in §5.4 are independent and each satisfies a FCLT. To state the result, let  $\hat{N}_n^a$  and  $\hat{S}_n^v$  be the scaled processes defined by

$$\hat{N}_{n}^{a}(t) \equiv n^{-1/2}[N(nt) - nt] \quad \text{and} \quad \hat{S}_{n}^{v}(t) \equiv n^{-1/2}[\sum_{i=1}^{\lfloor nt \rfloor} V_{k} - nt], \quad t \ge 0,$$
(5.26)

with  $\equiv$  denoting equality in distribution and  $\lfloor x \rfloor$  denoting the greatest integer less than or equal to x. We assume that

$$\hat{N}_n^a \Rightarrow c_a B_a \quad \text{and} \quad \hat{S}_n^v \Rightarrow c_s B_s \quad \text{in} \quad \mathcal{D} \quad \text{as} \quad n \to \infty,$$
(5.27)

where  $\mathcal{D}$  is the usual function space of right-continuous real-valued functions on  $[0, \infty)$  with left limits and  $\Rightarrow$  denotes convergence in distribution, as in Whitt (2002), while  $B_a$  and  $B_s$  are independent standard (mean 0, variance 1) Brownian motion processes (BM's). The assumed independence implies joint convergence in (5.27) by Theorem 11.4.4 of Whitt (2002).

We emphasize that GI assumptions are not needed, but that is an important special case. If the service times  $V_k$  are i.i.d. mean-1 random variables with variance, also the squared coefficient of variation (scv),  $c_s^2$ , then the limit in (5.27) holds with service variability parameter  $c_s$ . Similarly, if the base arrival process is a renewal process or an equilibrium renewal process with times between renewals having mean 1 and variance (and scv)  $c_a^2$ , then the limit in (5.27) holds with arrival variability parameter  $c_a$ . (See Nieuwenhuis (1989) for theoretical support in the case of an equilibrium renewal process.)

For the queueing HT FCLT, we will apply Theorem 9.3.4 of Whitt (2002), which refers to the conditions of Theorem 9.3.3. Those conditions require a joint FCLT for the partial sums of the arrival and service processes, notably (3.9) on p. 295. That convergence follows from the FCLT's we assumed for  $\hat{N}_n^a$  and  $\hat{S}_n^v$  in (5.27) above. In particular, the assumed FCLT for  $N_n^a$  implies the associated FCLT for the partial sums of the interarrival times by Theorem 7.3.2 and Corollary 7.3.1 of Whitt (2002).

#### 5.6.2 A Family of Models

As a basis for the HT FCLT, we create a model for each  $\rho$ ,  $0 < \rho < 1$ . We do so by defining the arrival-rate and service-rate functions.

#### 5.6.2.1 The Arrival-Rate and Service-Rate Functions.

Let the arrival-rate function in model  $\rho$  be as in (5.11) in the setting of (5.1)-(5.3). As a further regularity condition, we also require that the function s be an element of the function space  $\mathcal{D}$ , as in Whitt (2002). Then the associated cumulative arrival-rate function in model  $\rho$  be

$$\Lambda_{\rho}(t) \equiv \rho(t + (1 - \rho)^{-2}S((1 - \rho)^{2}t), \quad t \ge 0,$$
(5.28)

where

$$S(t) \equiv \int_0^t s(u) \, du, \tag{5.29}$$

for s again being the periodic function in (5.1)-(5.3). From (5.28)-(5.29), we see that the associated arrival-rate function obtained by differentiation in (5.28) is indeed  $\lambda_{\rho}(t)$  in (5.11).

The time scaling in (5.11) and (5.28) implies that the period in model  $\rho$  with arrival-rate function  $\lambda_{\rho}(t)$  in (5.11) is  $C_{\rho} = C(1-\rho)^{-2}$ , where C is the period of s(t) in (5.1)-(5.3). Thus the period  $C_{\rho}$  in model  $\rho$  is growing with  $\rho$ . This scaling follows Lemma 5.1 and Theorem 5.2 of Whitt (2015), with n there replaced by  $(1-\rho)^{-2}$ . In particular, the scaling here is in fluid or FWLLN scale, and thus is different from the diffusion or FCLT scaling in Theorem 3.2 of Whitt (2014) and Theorem 3.6.1.

Let  $A_{\rho}(t) \equiv N(\Lambda_{\rho}(t))$  be the arrival process in model  $\rho$ , which is obtained by using the cumulative arrival-rate function  $\Lambda_{\rho}$  in (5.28) in place of  $\Lambda$  in (2.2). Given that definition, we see that the cumulative arrival rate is indeed

$$E[A_{\rho}(t)] = E[N(\Lambda_{\rho}(t))] = \Lambda_{\rho}(t), \quad t \ge 0.$$
(5.30)

We now define associated scaled time-varying service-rate functions. These are the rate-matching service-rate functions in Chapter 4 modified by a time lag and a damping factor. In particular,

$$\mu_{\rho}(t) \equiv 1 + \xi_{\rho} s((1-\rho)^{2}(t-\eta_{\rho})) \text{ and} M_{\rho}(t) \equiv \int_{0}^{t} \mu_{\rho}(u) \, du = t + (1-\rho)^{-2} \xi_{\rho} S((1-\rho)^{2}(t-\eta_{\rho})), \quad t \ge 0,$$
(5.31)

where s is the periodic function with period C in (5.2), while  $\eta_{\rho}$  is the  $\rho$ -dependent time lag and  $\xi_{\rho}$  is the  $\rho$ -dependent damping factor. From (5.31) and (5.2), we see that the average service rate is  $\bar{\mu}_{\rho} = 1$  for all  $\rho$ . As a consequence, the average traffic intensity is  $\bar{\lambda}_{\rho}/\bar{\mu}_{\rho} = \rho$  for all  $\rho$ , while the instantaneous traffic intensity at time t is  $\lambda_{\rho}(t)/\mu_{\rho}(t)$ ,  $t \geq 0$ , which is a more complicated periodic function, again with period C.

#### 5.6.2.2 The Associated Queueing Processes

Having defined the family of arrival processes  $A_{\rho}(t)$  and deterministic service-rate functions  $M_{\rho}(t)$ above, we define the other queueing processes  $Y_{\rho}(t)$ ,  $X_{\rho}(t)$ ,  $L_{\rho}(t)$  and  $W_{\rho}(t)$  as in §5.4.2. Let the completed-work process be defined by

$$C'_{\rho}(t) \equiv Y_{\rho}(t) - L_{\rho}(t), \quad t \ge 0.$$
 (5.32)

We now can apply Lemma 5.4.1 in §5.4 to express the waiting time process as

$$W_{\rho}(t) \equiv \inf \{ u \ge 0 : M_{\rho}(t+u) - M_{\rho}(t) \ge L_{\rho}(t) \}, \quad t \ge 0.$$
(5.33)

The (virtual) waiting time  $W_{\rho}(t)$  represents the time that a hypothetical arrival at time t would have to wait before starting service.

As in (2.3), we can define the queue-length process (number in system) and the departure process in model  $\rho$  jointly. We can also express the departure process in terms of the workload process instead of the queue-length process by

$$D_{\rho}(t) \equiv N_s \left( \int_0^t \mu_{\rho}(s) \mathbf{1}_{\{L_{\rho}(s) > 0\}} \, ds \right), \quad t \ge 0,$$
(5.34)

but we do not focus on the departure and queue-length processes here.

## 5.6.3 The Scaled Queueing Processes

We start with the FWLLN-scaled processes. First let the scaled deterministic rate functions be

$$\bar{\Lambda}_{\rho}(t) \equiv (1-\rho)^2 \Lambda_{\rho}((1-\rho)^{-2}t) \text{ and } \bar{M}_{\rho}(t) \equiv (1-\rho)^2 M_{\rho}((1-\rho)^{-2}t), \quad t \ge 0,$$
 (5.35)

for  $\Lambda_{\rho}(t)$  in (5.28) and  $M_{\rho}(t)$  in (5.31). We immediately see that

$$\bar{\Lambda}_{\rho} \to \Lambda_{f} \quad \text{in} \quad \mathcal{D} \quad \text{as} \quad \rho \uparrow 1,$$

$$(5.36)$$

where

$$\Lambda_f(t) \equiv t + S(t), \quad t \ge 0, \tag{5.37}$$

for S(t) in (5.29).

Let the FWLLN-scaled arrival arrival stochastic process be defined by

$$\bar{A}_{\rho}(t) \equiv (1-\rho)^2 A_{\rho}((1-\rho)^{-2}t),$$
 (5.38)

Let the input, net-input, workload, completed-work and waiting-time components of the FWLLNscaled the vector  $(\bar{A}_{\rho}, \bar{Y}_{\rho}, \bar{X}_{\rho}, \bar{L}_{\rho}, \bar{C}'_{\rho}, \bar{W}_{\rho})$  be defined in the same way.

Then let the associated FCLT-scaled deterministic rate functions be defined by

$$\hat{\Lambda}_{\rho}(t) \equiv (1-\rho)[\Lambda_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}\Lambda_{f}(t)],$$
  
$$\hat{M}_{\rho}(t) \equiv (1-\rho)[M_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}\Lambda_{f}(t)]$$
(5.39)

for  $\Lambda_f$  in (5.37). Let the associated FCLT-scaled stochastic processes be defined by

$$\hat{A}_{\rho}(t) \equiv (1-\rho)[A_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}\Lambda_{f}(t)],$$

$$\hat{Y}_{\rho}(t) \equiv (1-\rho)[Y_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}\Lambda_{f}(t)],$$

$$\hat{X}_{\rho}(t) \equiv (1-\rho)X_{\rho}((1-\rho)^{-2}t),$$

$$\hat{L}_{\rho}(t) \equiv (1-\rho)L_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}\Lambda_{f}(t)],$$

$$\hat{W}_{\rho}(t) \equiv (1-\rho)W_{\rho}((1-\rho)^{-2}t), \quad t \ge 0.$$
(5.40)

## 5.6.4 The HT FWLLN

We start with the HT FWLLN. The limit provides a deterministic fluid approximation. However, simple fluid approximations evidently are too crude to provide much help. Corollary 5.6.1 below shows that the rate-matching control stabilizes both the workload and the waiting time for the fluid approximation.

Let  $\mathcal{D}^k$  be the k-fold product space of  $\mathcal{D}$  with itself, let  $\Rightarrow$  denote convergence in distribution and let  $x \circ y$  be the composition function defined by  $(x \circ y)(t) \equiv x(y(t))$ . Let  $a \wedge b \equiv \min \{a, b\}$  and let  $\psi : D \to D$  be the standard one-dimensional reflection map as in §13.5 of Whitt (2002), i.e.,

$$\psi(x)(t) \equiv x(t) - (\inf \{x(s) : 0 \le s \le t\} \land 0), \quad t \ge 0.$$
(5.41)

**Theorem 5.6.1.** (*HT FWLLN*) Under the definitions and assumptions in §5.6 above, if  $\xi_{\rho} \rightarrow \xi$ and  $\eta_{\rho} \rightarrow \eta$  as  $\rho \uparrow 1$ , and the system starts empty at time 0, then

$$\bar{M}_{\rho} \to M_f \quad in \quad \mathcal{D}, \quad where \quad M_f(t) \equiv t + \xi S(t - \eta)$$

$$(5.42)$$

and

$$(\bar{A}_{\rho}, \bar{Y}_{\rho}, \bar{X}_{\rho}, \bar{L}_{\rho}, \bar{C}'_{\rho}, \bar{W}_{\rho}) \Rightarrow (\bar{A}, \bar{Y}, \bar{X}, \bar{L}, \bar{C}', \bar{W}) \quad in \quad \mathcal{D}^{6} \quad as \quad \rho \uparrow 1$$
(5.43)

for  $(\bar{A}_{\rho}, \bar{Y}_{\rho}, \bar{X}_{\rho}, \bar{L}_{\rho}, \bar{C}'_{\rho}, \bar{W}_{\rho})$  defined in (5.38), where

$$\bar{A}(t) \equiv \bar{Y}(t) \equiv \Lambda_f(t), \quad \bar{X}(t) \equiv S(t) - \xi S(t - \eta), \quad t \ge \eta,$$

$$\bar{L}(t) \equiv \sup_{0 \le s \le C} \{X(t) - X(t - s)\}, \quad t \ge C + \eta,$$

$$\bar{C}'(t) \equiv \bar{Y}(t) - \bar{L}(t), \quad and$$

$$\bar{W}(t) \equiv \inf \{u \ge 0 : M_f(t + u) - M_f(t) \ge \bar{L}(t)\}, \quad t \ge 0.$$
(5.44)

for  $\Lambda_f(t)$  in (5.37) with S(t) in (5.29),  $M_f(t)$  in (5.42) and  $\psi$  being the reflection map in (5.41).

**Proof** We successively apply the continuous mapping theorem (CMT) using the functions in §12.7 and §13.2-13.6 of Whitt (2002). First, observe that (5.42) is a minor modification of (5.36). Let  $\bar{N}^a_{\rho}$  and  $\bar{S}_{\rho}$  denote  $\bar{N}^a_n$  and  $\bar{S}^v_n$ , respectively, where, paralleling (5.26), we let  $\bar{N}^a_n(t) \equiv n^{-1}N(nt)$ and  $\bar{S}^v_n \equiv n^{-1}S^v_{\lfloor nt \rfloor}$ ,  $t \ge 0$ , and then let  $n = (1 - \rho)^{-2}$ . Then observe that  $\bar{A}_{\rho} = \bar{N}^a_{\rho} \circ \bar{\Lambda}_{\rho}$  and  $\bar{Y}_{\rho} = \bar{S}_{\rho} \circ \bar{A}_{\rho}$ , so that we can apply the CMT with the composition map. The limit for  $\bar{X}_{\rho}$  follows from the CMT with addition and then the limit for  $\bar{L}_{\rho}$  follows from the CMT with the reflection map in (5.41). To establish the limit for the scaled waiting time  $\bar{W}_{\rho}(t)$  in  $\mathcal{D}$  we apply the CMT with the inverse function. Finally, the limit for  $\bar{C}_{\rho}$  again follows from the CMT with addition.

We obtain stronger results in special cases:

**Corollary 5.6.1.** (FWLLN for the rate-matching service rate control) In addition to the conditions of Theorem 5.6.1, if  $\eta = 0$  and  $\xi = 1$ , then  $M_f(t) = \Lambda_f(t)$ ,  $t \ge 0$ , and then  $\bar{X}(t) = \bar{L}(t) = \bar{W}(t) = 0$ for all  $t \ge 0$ , while  $\bar{C} = \bar{Y} = \bar{A} = \Lambda_f$ .

**Remark 5.6.1.** (stabilization achieved by many fluid models) It is evident that the conclusion of Corollary 5.6.1 holds for any single-server fluid model with arrival rate  $\lambda(t)$  and service rate  $\mu(t)$ provided that  $\mu(t) \geq \lambda(t)$  for all t. The  $(\eta, \xi)$  controls are intended to address the time-varying arrival rate in the more general stochastic setting.

As a modification of Corollary 5.6.1, we can have all customers wait exactly  $\eta$  if we provide no service until time  $\eta$ .

**Corollary 5.6.2.** (stabilizing the waiting time at any positive value) In addition to the conditions of Theorem 5.6.1, if  $\xi = 1$  and  $M_f(t) = 0$ ,  $0 \le t < \eta$ , then  $M_f(t) = \Lambda_f(t - \eta)$ ,  $t \ge \eta$ , for a fixed time lag  $\eta > 0$ , so that

$$\bar{L}(t) = \bar{X}(t) \equiv \bar{X}_{\eta}(t) = \Lambda_f(t) - \Lambda_f(t-\eta) = \int_{t-\eta}^t \lambda_f(s) \, ds > 0 \tag{5.45}$$

and

$$W(t) = \eta \quad \text{for all} \quad t \ge \eta. \tag{5.46}$$

**Corollary 5.6.3.** (sinusoidal with damped time lag) In addition to the conditions of Theorem 5.6.1, suppose that

$$s(t) \equiv \beta \sin\left(\gamma t\right), \quad t \ge 0,\tag{5.47}$$

for positive constants  $\beta$  and  $\gamma$  with  $\beta < 1$ , so that s(t) is periodic with period  $C \equiv C_{\gamma} = 2\pi/\gamma$ . Then

$$S(t) = (\beta/\gamma)(1 - \cos(\gamma t)), \quad t \ge \eta, \tag{5.48}$$

so that

$$\bar{L}(t) = (\beta/\gamma)([\xi\cos(\gamma(t-\eta)) - \cos(\gamma t)] 
+ \sup_{0 \le s \le C} \{\cos(\gamma(t-s)) - \xi\cos(\gamma(t-\eta-s))\} 
= (\beta/\gamma)([\xi\cos(\gamma(t-\eta)) - \cos(\gamma t)] 
+ \sup_{0 \le s \le C} \{\cos(\gamma s) - \xi\cos(\gamma(s-\eta))\}), \quad t \ge c + \eta.$$
(5.49)

For the special case  $\xi = 1$ ,  $\overline{W}(t) = \eta$ . If in addition, and  $\eta < \pi/\gamma$ , the supremum in (5.49) is attained at  $s^* = (\pi/2\gamma) - (\eta/2)$ , so that

$$\bar{L}(t) = \left(\frac{\beta}{\gamma}\right) \left(\left[\cos\left(\gamma(t-\eta)\right) - \cos\left(\gamma t\right)\right] + \left[\cos\left((\pi/2) - (\gamma\eta/2)\right) - \cos\left((\pi/2) + (\gamma\eta/2)\right)\right]\right)$$
(5.50)

for  $t \ge C + \eta$ . As  $\eta \downarrow 0$ ,

$$\bar{L}(t)/\eta \to 1 + \beta \sin\left(\gamma t\right) = 1 + s(t). \tag{5.51}$$

**Remark 5.6.2.** (the impact of high or low frequency) Corollary 5.6.3 shows the impact of high or low frequency. First, it is well known that high frequency has negligible impact, because performance tends to be determined by the behavior of the cumulative arrival rate function  $\Lambda(t)$  in (2.2) rather than the rate function  $\lambda(t)$ . From (5.48) and (5.49), we see that  $S(t) \to 0$  and  $\bar{L}(t) \to 0$  as  $\gamma \to \infty$ . On the other hand, for any fixed  $t, s(t) \to 0$  as  $\gamma \to 0$ .

## 5.6.5 The HT FCLT

We now state our main HT result: the HT FCLT with periodicity in fluid scale, as in (5.11). We present the proof in §5.7 after discussing consequences here.

**Theorem 5.6.2.** (HT FCLT) In addition to the definitions and assumptions in §5.6 above, including the scaled arrival-rate function in (5.11), assume that the periodic function s(t) in (5.2) is continuous and

$$(1-\rho)\eta_{\rho} \to \hat{\eta} \quad and \quad \frac{\xi_{\rho}-1}{1-\rho} \to \hat{\xi} \quad as \quad \rho \uparrow 1,$$
 (5.52)

where  $0 \leq \hat{\eta} < \infty$  and  $0 \leq \hat{\xi} < \infty$ . Then there is a limit for the scaled cumulative service-rate functions  $\hat{M}_{\rho}$  in (5.31) and (5.39); i.e.,

$$\hat{M}_{\rho}(t) \equiv (1-\rho)[M_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}(t+S(t))] \rightarrow \hat{M}(t) \equiv -s(t)\hat{\eta} + S(t)\hat{\xi} \quad in \quad \mathcal{D} \quad as \quad \rho \uparrow 1$$
(5.53)

for s(t) in (5.2) and S(t) in (5.29). If, in addition, the system starts empty at time 0, then

$$(\hat{A}_{\rho}, \hat{Y}_{\rho}, \hat{X}_{\rho}, \hat{L}_{\rho}, \hat{W}_{\rho}, \hat{C}'_{\rho}) \Rightarrow (\hat{A}, \hat{Y}, \hat{X}, \hat{L}, \hat{W}, \hat{C}') \quad in \quad \mathcal{D}^5 \quad as \quad \rho \uparrow 1$$
(5.54)

for  $(\hat{A}_{\rho}, \hat{Y}_{\rho}, \hat{X}_{\rho}, \hat{L}_{\rho}, \hat{W}_{\rho}, \hat{C}'_{\rho})$  defined in (5.40), where

$$\hat{A}(t) \equiv (c_a B_a - e) \circ \Lambda_f(t), \quad \hat{Y}(t) \equiv (c_x B - e) \circ \Lambda_f(t), \quad \hat{C}'(t) \equiv \hat{Y}(t) - \hat{L}(t),$$

$$\hat{X}(t) \equiv \hat{Y}(t) - \hat{M}(t) = \hat{Y}(t) + s(t)\hat{\eta} - S(t)\hat{\xi}$$

$$= (c_x B \circ \Lambda_f)(t) - \Lambda_f(t) + s(t)\hat{\eta} - S(t)\hat{\xi},$$

$$\hat{L}(t) \equiv \psi(\hat{X})(t) \quad and \quad \hat{W}(t) \equiv \hat{L}(t)/\mu_f(t), \quad t \ge 0.$$
(5.55)

with  $c_x \equiv \sqrt{c_a^2 + c_s^2}$ , B a BM,  $\psi$  the reflection map in (5.41) and  $\mu_f(t) \equiv \lambda_f(t) \equiv 1 + s(t)$ ,  $t \ge 0$ , the limiting arrival-rate function, the dervative of  $\Lambda_f$  in (5.37).

We now draw attention to some important consequences. First, Theorem 5.6.2 establishes a HT time-varying (TV) Little's law (LL), paralleling the many-server heavy-traffic (MSHT) TV LL in Sun and Whitt (2018) and exposed for the rate-matching control in Theorem 4.2.4. This is a time-varying version of the familiar state-space collapse, which goes back to the early HT papers, e.g., Whitt (1971). We remark that the relation is different from the time-varying LL discussed in Bertsimas and Mourtzinou (1997), Fralix and Riano (2010) and Sigman and Whitt (2018), Whitt and Zhang (2018).

**Corollary 5.6.4.** (*HT time-varying Little's law*) Under the conditions of Theorem 5.6.2, the limit processes are related by

$$\hat{L}(t) = \lambda_f(t)\hat{W}(t), \quad t \ge 0, \quad w.p.1.$$
 (5.56)

We now consider an alternative deterministic limit to the HT FWLLN in Theorem 5.6.1. Now we assume that the FCLT holds with the variability parameter set equal to 0. For this purpose, we assume that s(t) is differentiable and let  $\dot{s}(t)$  be its derivative. **Corollary 5.6.5.** (the case of no variability) If  $c_x = 0$  and s(t) is differentiable in addition to the conditions of Theorem 5.6.2, then

$$\hat{X}(t) = -t + s(t)\hat{\eta} - S(t)(\hat{\xi} + 1), \quad t \ge 0,$$
(5.57)

so that  $\hat{L}(t) = \hat{W}(t) = 0$  for all  $t \ge 0$  if and only if

$$\frac{d\hat{X}(t)}{dt} = -1 + \hat{\eta}\dot{s}(t) - (\hat{\xi} + 1)s(t) \le 0, \quad t \ge 0.$$
(5.58)

In the sinusoidal case with  $s(t) \equiv \beta \sin \gamma t$  in (2.4),

$$\frac{d\hat{X}(t)}{dt} = -1 + \hat{\eta}\beta\gamma\cos\gamma t - (\hat{\xi} + 1)\beta\sin\gamma t, \quad t \ge 0.$$
(5.59)

For  $\beta = 1$  and  $\gamma \rightarrow 0$ ,

$$\frac{dX(t)}{dt} \to -1 - (\hat{\xi} + 1)\beta \sin \gamma t, \quad t \ge 0,$$
(5.60)

which is strictly positive over subintervals if  $\hat{\xi} > 0$ .

For the nondegenerate sinusoidal arrival rate function, the derivative in (5.58) of Corollary 5.6.5 implies it is not always possible to stabilize the limiting time-varying diffusion process  $\hat{W}$  with  $\hat{\xi} > 0$ in Theorem 5.6.2. We conjecture that it is never possible to stabilize it perfectly.

We now establish conditions for the optimality of an  $(\hat{\eta}^*, \hat{\xi}^*)$  control for the limiting diffusion control problem for either formulation (5.7) or (5.8). Our proof will exploit uniform integrability (UI); see p. 31 of Billingsley (1999).

**Corollary 5.6.6.** (optimality for the limiting diffusion process) Consider the special case of the  $GI_t/GI_t/1$  model with  $E[U_k^{2+\epsilon}] < \infty$  and  $E[V_k^{2+\epsilon}] < \infty$  for some  $\epsilon > 0$ . If  $(\eta_{\rho}^*, \xi_{\rho}^*) \to (\hat{\eta}^*, \hat{\xi}^*)$  as  $\rho \to 1$ , where  $(\eta_{\rho}^*, \eta_{\rho}^*)$  is the optimal control for problem (5.7) or (5.8), then the limiting control  $(\hat{\eta}^*, \hat{\xi}^*)$  is optimal for the corresponding diffusion control problem.

**Proof** We let  $(\tilde{\eta}, \tilde{\xi})$  be any alternative control for the limiting diffusioon process. Then let  $(\tilde{\eta}_{\rho}, \tilde{\xi}_{\rho})$  be an associated control for model  $\rho$ ,  $0 < \rho < 1$ , where  $\tilde{\eta}_{\rho} \equiv \tilde{\eta}/(1-\rho)$  and  $\tilde{\eta}_{\rho} \equiv 1 + (1-\rho)\tilde{\xi}$ . Then, by this construction, condition (5.52) holds for the family  $(\tilde{\eta}_{\rho}, \tilde{\xi}_{\rho})$ . We next want to show that the convergence in distribution can be extended to convergence of the means for all t, which requires uniform integrability uniformly in t; see p. 31 of Billingsley (1999). We use the bounds on the second moments to show that it holds.

Toward that end, we exploit the upper bounds for the workload process in the  $G_t/G_t/1$  model in terms of the associated workload process in the stationary G/G/1 model from §3.3. These bounds extend directly to the  $G_t/G_t/1$  model by virtue of Corollary 5.4.1. These bounds show that the mean workload is bounded above uniformly in y over the interval [0, C]. These bounds also apply to the waiting time process because  $W(t) \leq L(t)/\mu_L$ , where  $\mu_L > 0$  is a lower bound on the service rate, which follows from (5.3) and (5.4). For the stationary GI/GI/1 model, finite second moments imply the existence of the first moments of the waiting time and uniform integrability needed for convergence; see p. 31 of Billingsley (1999) and §X.2 and X.7 of Asmussen (2003).

Finally, we observe that our optimal policy  $(\eta_{\rho}^*, \xi_{\rho}^*)$  has expected value greater than or equal to the alternative policy  $(\tilde{\eta}_{\rho}, \tilde{\xi}_{\rho})$  for all  $\rho$ , while both converge as  $\rho \to 1$ . Hence, the limit of the optimal policies,  $(\hat{\eta}^*, \hat{\xi}^*)$  must be at least as good as  $(\tilde{\eta}, \tilde{\xi})$ .

We apply Corollary 5.6.6 to support our numerical calculations by observing that  $(\eta_{\rho}^*, \xi_{\rho}^*)$  when scaled as in (5.52) converges to a limit. We thus deduce that the limit must be the optimal policy for the diffusion. However, this numerical evidence is not a mathematical proof. Moreover, while the numerical evidence is good, it is not exceptionally good, especially for  $\xi_{\rho}^*$  as can be seen from Table 5.1 in §5.9.1 below.

## 5.7 Proof of Theorem 5.6.2

To establish (5.53), apply (5.31) and (5.39) to obtain

$$\hat{M}_{\rho}(t) \equiv (1-\rho)[M_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}(t+S(t))] 
= (1-\rho)^{-1}[\xi_{\rho}S(t-(1-\rho)^{2}\eta_{\rho}) - S(t)] 
= (1-\rho)^{-1}[\xi_{\rho}S(t-(1-\rho)^{2}\eta_{\rho}) - \xi_{\rho}S(t)] + (1-\rho)^{-1}[\xi_{\rho}S(t) - S(t)] 
\rightarrow -\hat{\eta}s(t) + \hat{\xi}S(t) \text{ in } \mathcal{D} \text{ as } \rho \uparrow 1,$$
(5.61)

where on the third line we have subtracted and added the term  $\xi_{\rho}S(t)$  and on the last line we have differentiated using

$$(1-\rho)^2 \eta_{\rho}/(1-\rho) = (1-\rho)\eta_{\rho} \rightarrow \hat{\eta} \text{ as } \rho \uparrow 1$$

by assumption (5.52). We used the assumed continuity of s to have S be continuously differentiable, so that the derivative of S(t) holds uniformly in t over bounded intervals. We next establish (5.54). First, the limit for  $\hat{A}_{\rho}$  is given in Lemma 5.1 of Whitt (2015), but we need to make an adjustment because the arrival rate in model  $\rho$  is chosen to be  $\rho$  here as opposed to 1 before. From (5.28), (5.36) and (5.39), we see that

$$\bar{\Lambda}_{\rho}(t) = \rho \Lambda_{f}(t) \to \Lambda_{f}(t) \quad \text{in } \mathcal{D} \quad \text{as } \rho \to 1$$
$$\hat{\Lambda}_{\rho}(t) = (1-\rho)^{-1} \rho \Lambda_{f}(t) - (1-\rho)^{-1} \Lambda_{f}(t) = -\Lambda_{f}(t)$$
(5.62)

for all  $\rho$ , where  $\Lambda_f(t)$  is defined in (5.37). Then the limit for  $\hat{A}_{\rho}$  follows from the standard argument for random sums. The key is to observe that

$$\hat{A}_{\rho} = \hat{N}_{\rho} \circ \bar{\Lambda}_{\rho} + \hat{\Lambda}_{\rho}, \qquad (5.63)$$

where  $\hat{N}_{\rho}$  is defined to be  $\hat{N}_n$  in (5.26) for  $n = (1-\rho)^{-2}$ . So we can start with the joint convergence

$$\left(\hat{N}_{\rho}, \bar{\Lambda}_{\rho}, \hat{\Lambda}_{\rho}\right) \Rightarrow (c_a B_a, \Lambda_f, -\Lambda_f) \quad \text{in} \quad \mathcal{D}^3 \quad \text{as} \quad \rho \to 1,$$
(5.64)

We then apply convergence preservation with the map  $g(x, y, z) = x \circ y + z$  (composition plus addition) as in §13.3 of Whitt (2002) to get  $\hat{A}_{\rho} \Rightarrow c_a B_a \circ \Lambda_f - \Lambda_f = (c_a B_a - e) \circ \Lambda_f$  in  $\mathcal{D}$ .

Similarly, given that  $\bar{N}_{\rho} \equiv (1-\rho)^2 N((1-\rho)^{-2}t) \Rightarrow e$  and  $\bar{A}_{\rho} \equiv (1-\rho)^2 A((1-\rho)^{-2}t)$ ,

$$\bar{A}_{\rho} = (1-\rho)^2 N(\Lambda_{\rho}((1-\rho)^{-2}t)) = (1-\rho)^2 N((1-\rho)^{-2}\rho\Lambda_f(t))$$
$$= \bar{N}_{\rho}(\rho\Lambda_f(t)) \Rightarrow \Lambda_f \quad \text{in} \quad \mathcal{D} \quad \text{as} \quad \rho \to 1.$$
(5.65)

A variant of the random-sum argument holds for  $\hat{Y}_{\rho}$  too. In particular, we start with the joint convergence

$$\left(\hat{S}_{\rho}, \bar{A}_{\rho}, \hat{A}_{\rho}\right) \Rightarrow (c_s B_s, \Lambda_f, c_a B_a \circ \Lambda_f - \Lambda_f) \quad \text{in} \quad \mathcal{D}^3 \quad \text{as} \quad \rho \to 1,$$
 (5.66)

The joint convergence holds by virtue of Theorems 11.4.4 and 11.4.5 of Whitt (2002). We then apply convergence preservation with the map  $g(x, y, z) = x \circ y + z$  (composition plus addition) as in §13.3 of Whitt (2002) to get

$$\hat{Y}_{\rho} = \hat{S}_{\rho} \circ \bar{A}_{\rho} + \hat{A}_{\rho} \Rightarrow c_s B_s \circ \Lambda_f + c_a B_a \circ \Lambda_f - \Lambda_f 
\stackrel{d}{=} c_x B \circ \Lambda_f - \Lambda_f \quad \text{in} \quad \mathcal{D} \quad \text{as} \quad \rho \to 1.$$
(5.67)

Then the limits for  $\hat{X}_{\rho}$  and  $\hat{L}_{\rho}$  follow from the continuous mapping theorem with the standard reflection map reasoning, e.g., as in Chapter 9 of Whitt (2002), even though the service rate function is now more general.

However, the waiting time requires a new treatment. The limit follows from the definition of the scaled service-rate control in (5.31) and the first-passage-time representation of the waiting time in (5.33). The structure and result are similar to the Puhalskii Puhalskii (1994) theorem and related results in §13.7 of Whitt (2002), but they evidently do not apply directly.

We will apply Taylor's theorem to a perturbation of S in (5.29). The essential idea is that

$$(1-\rho)^{-1}[S(t+(1-\rho)u) - S(t)] \to s(t)u \text{ as } \rho \to 1$$
 (5.68)

uniformly in t and u over bounded intervals. Just as in (5.61), we use the assumed continuity of s to have S be continuously differentiable, so that the derivative of S(t) holds uniformly in t and u over bounded intervals.

For the specific application, let

$$\tilde{S}_{\rho}(t,u) \equiv (1-\rho)^{-1} \xi_{\rho} [S(t+(1-\rho)u - (1-\rho)^2 \eta_{\rho}) - S(t-(1-\rho)^2 \eta_{\rho})]$$
(5.69)

and

$$\zeta_{\rho}(t,u) \equiv \tilde{S}_{\rho}(t,u) - s(t)u.$$
(5.70)

By combining (5.68) and the two limits in condition (5.52), we see that  $\zeta_{\rho}(t, u)$  is asymptotically negligible as  $\rho \to 1$  uniformly in t and u over bounded intervals. We will use this at the critical final step in the following representation.

To start, let  $\tilde{M}_{\rho}(t, u) \equiv M_{\rho}((1-\rho)^{-2}t+u)$ . Then, from (5.40) and (5.33),

$$\hat{W}_{\rho}(t) \equiv (1-\rho)W_{\rho}((1-\rho)^{-2}t) 
= (1-\rho)\inf\{u \ge 0: \tilde{M}_{\rho}(t, u) - \tilde{M}_{\rho}(t, 0)) \ge L_{\rho}((1-\rho)^{-2}t)\} 
= \inf\{u \ge 0: \tilde{M}_{\rho}(t, (1-\rho)^{-1}u) - \tilde{M}_{\rho}(t, 0) \ge L_{\rho}((1-\rho)^{-2}t)\} 
= \inf\{u \ge 0: (1-\rho)[\tilde{M}_{\rho}(t, (1-\rho)^{-1}u) - \tilde{M}_{\rho}(t, 0)] \ge \hat{L}_{\rho}(t)\} 
= \inf\{u \ge 0: u + \tilde{S}_{\rho}(t, u) \ge \hat{L}_{\rho}(t)\} 
= \inf\{u \ge 0: u + s(t)u + \zeta_{\rho}(t, u) \ge \hat{L}_{\rho}(t)\} 
= \inf\{u \ge 0: u\lambda_{f}(t) + \zeta_{\rho}(t, u) \ge \hat{L}_{\rho}(t)\}, \quad t \ge 0,$$
(5.71)

where  $\lambda_f(t) = t + s(t)$  by (5.37) and we apply Taylor's theorem with (5.69) and (5.70) in line 6 to obtain that  $\zeta_{\rho}(t, u)$  is asymptotically negligible as  $\rho \to 1$  uniformly over both t and u over bounded subintervals. For the final step, to simplify, we make the entire argument deterministic by using the Skorohod representation theorem, as in Theorem 3.2.2 of Whitt (2002), to replace the stochastic convergence  $\hat{L}_{\rho} \Rightarrow \hat{L}$  in  $\mathcal{D}$  by associated convergence w.p.1. Then we see from line 6 of (5.71) that in the infinimum it suffices to consider u only just beyond  $\hat{L}(t)/\lambda_f(t)$ , which for t in a bounded interval is bounded for each sample path, because  $\lambda_f(t)$  has been assumed to be bounded below, while  $\hat{L}(t)$ is bounded above, for t in a bounded interval. Thus, we can write

$$\frac{\hat{L}_{\rho}(t) - K\zeta_{\rho}^{\uparrow}(t)}{\lambda_{f}(t)} \le \hat{W}_{\rho}(t) \le \frac{\hat{L}_{\rho}(t) + K\zeta_{\rho}^{\uparrow}(t)}{\lambda_{f}(t)}$$
(5.72)

for t and u over specified bounded intervals, K an appropriate positive constant and

$$\zeta_{\rho}^{\uparrow}(t) \equiv \sup_{0 \le u \le \bar{u}} |\zeta_{\rho}(t, u)|$$

for an appropriate  $\bar{u}$ . Given that  $\hat{L}_{\rho} \Rightarrow \hat{L}$  and  $\zeta_{\rho}^{\uparrow} \to 0$  in  $\mathcal{D}$ , we can use the standard sandwiching argument (uniformly over bounded time intervals) to obtain convergence  $\hat{W}_{\rho}(t) \Rightarrow \hat{L}(t)/\lambda_f(t) \equiv \hat{W}(t)$  in  $\mathcal{D}$ , which completes the proof.  $\blacksquare$ 

## 5.8 A HT FCLT with Periodicity

In this section we establish a HT FCLT with periodicity holding in the weaker diffusion scale instead of in the fluid scale, as was done in §5.6. The scaling here follows Whitt (2014) and §3.6 instead of Whitt (2015). In this scaling the HT limits of the waiting time coincides with the HT limit for the workload process, and so does not capture the differences we see in the simulations in previous sections.

## 5.8.1 An Alternative Family of Models

We start with the same basic rate-1 processes in §5.6.1. We then create a model for each  $\rho$ ,  $0 < \rho < 1$ , now using (5.12) instead of (5.11). That yields the family of cumulative arrival rate functions

$$\Lambda_{\rho}(t) \equiv \rho(t + (1 - \rho)^{-1} S((1 - \rho)^2 t)), \quad t \ge 0,$$
(5.73)

for S in (5.29). Differentiating in (5.73) yields the arrival-rate function in (5.12). Just as before, the time scaling in (5.12) and (5.73) implies that the period in model  $\rho$  with arrival-rate function
$\lambda_{\rho}(t)$  in (5.12) is  $C_{\rho} = C(1-\rho)^{-2}$ , where C is the period of s in (5.1)-(5.3). Thus the period  $C_{\rho}$  in model  $\rho$  is growing with  $\rho$ .

# 5.8.2 An Associated Family of Service-Rate Controls

Just as in §5.6.2.1, we define associated service-rate controls. Closely paralleling (5.12) and (5.73), we define associated scaled time-varying service-rate functions using the control parameters  $\eta_{\rho}$  and  $\xi_{\rho}$ , i.e., for all  $t \ge 0$ ,

$$\mu_{\rho}(t) \equiv 1 + (1-\rho)\xi_{\rho}s(t-\eta_{\rho}) \text{ and}$$

$$M_{\rho}(t) \equiv \int_{0}^{t} \mu_{\rho}(s) \, ds = t + (1-\rho)^{-1}\xi_{\rho}S((1-\rho)^{2}(t-\eta_{\rho})). \tag{5.74}$$

Just as in (5.73), differentiation of  $M_{\rho}(t)$  in (5.74) shows that it is consistent with  $\mu_{\rho}(t)$ . As a consequence of (5.74), the average service rate is  $\bar{\mu}_{\rho} = 1$ ,  $0 < \rho < 1$ .

### 5.8.3 The Scaled Queueing Processes

We use the same processes introduced in §5.4, but new scaling. Let the scaled arrival-rate and service-rate functions be defined for  $t \ge 0$  by

$$\hat{\Lambda}_{\rho}(t) \equiv (1-\rho)[\Lambda_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}t] = \rho S(t) - t \hat{M}_{\rho}(t) \equiv (1-\rho)[M_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}t] = \xi_{\rho}S(t-(1-\rho)^{2}\eta_{\rho}).$$
(5.75)

Clearly,  $\hat{\Lambda}_{\rho}(t) \to S(t) - t$  as  $\rho \to 1$  uniformly over bounded intervals of t. The key is what happens to  $\hat{M}_{\rho}(t)$ . From (5.75), we get

**Lemma 5.8.1.** (*HT limit of*  $\hat{M}_{\rho}(t)$ ) If  $\xi_{\rho} \to 1$  and  $(1-\rho)^2 \eta_{\rho} \to 0$ , then  $\hat{M}_{\rho}(t) \to S(t)$  uniformly over bounded intervals of t.

Then let associated scaled stochastic processes be defined by

$$\hat{A}_{\rho}(t) \equiv (1-\rho)[A_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}t],$$

$$\hat{Y}_{\rho}(t) \equiv (1-\rho)[Y_{\rho}((1-\rho)^{-2}t) - (1-\rho)^{-2}t],$$

$$\hat{X}_{\rho}(t) \equiv (1-\rho)X_{\rho}((1-\rho)^{-2}t),$$

$$\hat{L}_{\rho}(t) \equiv (1-\rho)L_{\rho}((1-\rho)^{-2}t),$$

$$\hat{C}_{\rho}' \equiv \hat{Y}_{\rho} - \hat{L}_{\rho},$$

$$\hat{W}_{\rho}(t) \equiv (1-\rho)W_{\rho}((1-\rho)^{-2}t),$$

$$t \ge 0.$$
(5.76)

Note that the translation terms in  $\hat{\Lambda}_{\rho}$  and  $\hat{M}_{\rho}$  in (5.75) are different from the translation terms in (5.35), while the translation terms in  $\hat{A}_{\rho}$  and  $\hat{Y}_{\rho}$  in (5.76) are different from the translation terms in (5.40). Thus, the statement of the heavy-traffic limit below is different (and weaker).

### 5.8.4 The HT FCLT with Periodicity

Just as in §5.6, the following heavy-traffic FCLT states that  $\hat{A}_{\rho}$  and  $\hat{Y}_{\rho}$  converge to periodic Brownian motions (PBM's). However, unlike §5.6,  $\hat{X}_{\rho}$  converges to an *ordinary* Brownian motion (BM),  $\hat{L}_{\rho}$  and  $\hat{W}_{\rho}$  converge to the *same ordinary* reflected Brownian motion (RBM), while  $\hat{C}'_{\rho}$  has a complicated limit. We thus show that  $\hat{L}_{\rho}$  and  $\hat{W}_{\rho}$  are asymptotically stable and Markov. Note that the scaling condition on  $(\eta_{\rho}, \xi_{\rho})$  here are implied by condition (5.52) in Theorem 5.6.2, but as noted above the conclusion is different and weaker, because of the different translation terms.

**Theorem 5.8.1.** (heavy-traffic limit extending Theorem 3.2 of Whitt (2014) and 3.6.1) If, in addition to the definitions and assumptions in (5.73)-(5.76) above,  $(1-\rho)^2\eta_{\rho} \to 0$  and  $\xi_{\rho} \to 1$  as  $\rho \to 1$  and the system starts empty at time 0, then

$$(\hat{\Lambda}_{\rho}, \hat{M}_{\rho}, \hat{A}_{\rho}, \hat{Y}_{\rho}, \hat{X}_{\rho}, \hat{L}_{\rho}, \hat{W}_{\rho}, \hat{C}'_{\rho}) \Rightarrow (\hat{\Lambda}, \hat{M}, \hat{A}, \hat{Y}, \hat{X}, \hat{L}, \hat{W}, \hat{C}')$$
(5.77)

in  $\mathcal{D}^8$  as  $\rho \to 1$  for  $(\hat{\Lambda}_{\rho}, \hat{M}_{\rho})$  defined in (5.75) and  $(\hat{A}_{\rho}, \hat{Y}_{\rho}, \hat{X}_{\rho}, \hat{L}_{\rho}, \hat{W}_{\rho}, \hat{C}'_{\rho})$  defined in (5.76), where

$$\hat{\Lambda} \equiv S - e, \quad \hat{A} \equiv c_a B_a + S - e, \quad \hat{M} \equiv S,$$

$$\hat{Y} \equiv \hat{A} + c_s B_s, \quad \hat{X} \equiv \hat{Y} - S \stackrel{d}{=} c_x B - e,$$

$$\hat{L} \equiv \psi(\hat{X}), \quad \hat{W} \equiv \psi(\hat{X}) \quad and \quad \hat{C}' \equiv \hat{Y} - \hat{L},$$
(5.78)

with  $B_a$  and  $B_s$  being independent BM's, S in (5.29),  $c_a$  and  $c_s$  being the variability parameters in (5.27),  $c_x \equiv \sqrt{c_a^2 + c_s^2}$  and B is a BM.

**Proof** We will be brief because most of the argument is essentially the same as in Whitt (2014) and Chapter 3. First, the limit for  $\hat{A}_{\rho}$  is given in Theorem 3.2 of Whitt (2014). Then the limit for  $\hat{Y}_{\rho}$  follows from Theorem 9.3.4 of Whitt (2002), as noted in the proof of 3.6.1. (See C(t) in (9.2.4) and  $C_n$  in (9.3.4) and Theorem 9.3.4 of Whitt (2002).) Then the limits for  $\hat{X}_{\rho}$  and  $\hat{L}_{\rho}$  follow from the standard reflection mapping argument as in even though the service rate function is now more general. Again, the waiting time requires a new treatment. The limit follows from the first-passagetime representation in (5.33). In particular, paralleling (5.71), letting  $\tilde{M}_{\rho}(t, u) \equiv M_{\rho}((1-\rho)^{-2}t+u)$ , we have

$$\hat{W}_{\rho}(t) \equiv (1-\rho)W_{\rho}((1-\rho)^{-2}t) 
= (1-\rho)\inf\{u \ge 0: \tilde{M}_{\rho}(t,u) - \tilde{M}_{\rho}(t,0) \ge L_{\rho}((1-\rho)^{-2}t)\} 
= \inf\{u \ge 0: \tilde{M}_{\rho}(t,(1-\rho)^{-1}u) - \tilde{M}_{\rho}(t,0) \ge L_{\rho}((1-\rho)^{-2}t)\} 
= \inf\{u \ge 0: (1-\rho)[\tilde{M}_{\rho}(t,(1-\rho)^{-1}u) - \tilde{M}_{\rho}(t,0)] \ge \hat{L}_{\rho}(t)\} 
= \inf\{u \ge 0: u + \zeta_{\rho}(t,u) \ge \hat{L}_{\rho}(t)\},$$
(5.79)

for  $t \geq 0$ , where

$$\zeta(t,u) \equiv \xi_{\rho} \left[ S(t+(1-\rho)u - (1-\rho)^2 \eta_{\rho}) - S(t-(1-\rho)^2 \eta_{\rho}) \right],$$
(5.80)

which is asymptoically negligible as  $\rho \to 1$  uniformly in compact intervals, given the conditions on  $\eta_{\rho}$  and  $\xi_{\rho}$ . As technical support for the last step, note that

$$S(t+\epsilon) - S(t) \le s_U \epsilon \quad \text{for all} \quad \epsilon > 0, \tag{5.81}$$

for  $s_U$  in (5.3). Also add and subtract  $\xi_{\rho}S(t)$  and treat the two terms separately, i.e.,

$$\xi_{\rho}S(t + (1 - \rho)u - (1 - \rho)^{2}\eta_{\rho}) - S(t) = \xi_{\rho}S(t) - S(t) + \xi_{\rho}S(t + (1 - \rho)u - (1 - \rho)^{2}\eta_{\rho}) - \xi_{\rho}S(t).$$

Hence, we can apply the continuous mapping theorem for the inverse in §13.6 of Whitt (2002) to get  $\hat{W}_{\rho} \Rightarrow \hat{L}$  in  $\mathcal{D}$  as  $\rho \to 1$ , jointly with the other limits.

# 5.9 Simulation Examples

### 5.9.1 In the Setting of §5.6

In this section we report results of simulation experiments to evaluate the new optimal  $(\eta_{\rho}^*, \xi_{\rho}^*)$ controls as a function of  $\rho$  for models scaled according to Theorem 5.6.2, specifically by (5.11), (5.28) and (5.31), so that we can see the systematic behavior.

#### 5.9.1.1 Sinusoidal Examples

We start with sinusoidal examples and then consider non-sinusoidal examples. Table 5.1 shows results for four values of the traffic intensity  $\rho$  with  $\rho \uparrow 1$  for the sinusoidal model in (5.1)-(5.4) with HT scaling in (5.11) with parameters  $(\rho, \beta_{\rho}, \gamma_{\rho}) = (\rho, 0.2, 2.5(1-\rho)^2)$ . For this case, we found that the solutions to optimization problems (5.7) and (5.8) are identical, to within our statistical precision. Hence, our solutions are for both problems.

Table 5.1 shows the estimated optimal controls  $\eta_{\rho}^*$  and  $\xi_{\rho}^*$  in each case, plus scaled versions consistent with condition (5.52). Table 5.1 shows that the relative error is roughly independent of  $\rho$ , being less than 1% in each case. Table 5.1 also shows that the limit  $\hat{\eta}^* \approx 1.45$  is rapidly approached by  $(1-\rho)\eta_{\rho}^*/\rho$ , while the limit  $\hat{\xi}^* \approx 1.8$  is roughly approached by  $(\xi_{\rho}^* - 1)/(1-\rho)$ , both of which are consistent with condition (5.52). The results support Theorem 5.6.2, but unfortunately the rate of convergence in the control parameters is not fast. Evidently the optimal damping control  $\xi_{\rho}^*$  is more problematic.

For the model in Table 5.1, Figure 5.3 shows the expected periodic steady-state virtual waiting time (solid blue line), the expected steady-state workload (the dashed red line) and arrival rate multiplied by the mean waiting time (the dotted green line) for  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right). As in Figure 5.1, the 95% confidence interval bands are included, but they can only be seen by zooming in.

We also considered alternative values of the relative amplitude  $\beta$ . Table 5.2 shows the solutions to the minimum-deviation optimization problem in (5.8) for the sinusoidal model in Table 5.1 except  $\beta$  has been increased to  $\beta = 0.8$  from 0.2. Table 5.2 shows that the relative error is roughly independent of  $\rho$ , but the relative error has increased to about 10% from about 1% in Table 5.1. Unlike in Figure 5.3, it is evident that the  $(\eta_{\rho}^*, \xi_{\rho}^*)$  control does not stabilize the expected waiting

**Table 5.1:** The (identical) solutions to the minimax and minimum-deviation optimization problems in (5.7) and (5.8) for the sinusoidal model in (5.1)-(5.4) with HT scaling in (5.11) with parameters  $(\rho, \beta_{\rho}, \gamma_{\rho}) = (\rho, 0.2, 2.5(1 - \rho)^2)$ . The mean waiting times are reported with and without space scaling.

ρ	0.8	0.9	0.95	0.975
$\beta_{ ho} \equiv \beta$	0.2	0.2	0.2	0.2
$\gamma_ ho$	0.1	0.025	0.00625	0.0015625
$\eta^*_ ho$	5.80	12.94	27.7	56.6
$\hat{\eta}_{\rho}^* \equiv (1-\rho)\eta_{\rho}^*/\rho$	1.45	1.44	1.46	1.45
$\xi^*_ ho$	0.842	0.889	0.931	0.960
$\hat{\xi}_{\rho}^* \equiv (1-\xi_{\rho}^*)/(1-\rho)$	0.79	1.11	1.38	1.60
$\max E[W_y]$	4.03	9.10	19.29	39.61
$(1-\rho)\max E[W_y]/ ho$	1.008	1.011	1.015	1.016
$\max E[W_y] - \min E[W_y]$	0.032	0.091	0.143	0.364
avg $E[W_y]$	4.02	9.07	19.21	39.47
$(1-\rho)$ avg $E[W_y]/\rho$	1.005	1.007	1.011	1.012
relative error	0.8%	1.0%	0.7%	0.9%

time perfectly, either for fixed  $\rho$  or asymptotically as  $\rho \to 1$ .

From cases with  $0.2 \leq \beta \leq 0.9$  and  $0.8 \leq \rho \leq 0.975$ , we conclude that  $\hat{\eta}_{\rho}^* \equiv (1-\rho)\eta_{\rho}/\rho$  and  $\hat{\xi}_{\rho}^* \equiv (1-\xi_{\rho})/(1-\rho)$  are nondecreasing in  $\rho$ , while  $\hat{\eta}_{\rho}^*$  ( $\hat{\xi}_{\rho}^*$ ) is nondecreasing (nonincreasing) in  $\beta$ . The relative error tends to be independent of  $\rho$  but is increasing in  $\beta$ . The relative error for  $\beta = 0.5$  was about 4%, while the relative error for  $\beta = 0.9$  was about 22%. The difficulty as  $\beta \uparrow 1$  can be partially understood by the rate-matching control, where  $E[W_y] \approx c/\lambda_f(t)$  by Theorem 4.2.4, where c is the stable value, which has minimum and maximum values  $c/(1+\beta)$  and  $c/(1-\beta)$ , which deviate greatly as  $\beta \uparrow 1$ . (The constant c is the stable value of the expected queue length.) Tables 5.1 and 5.2 also show that the limiting optimal controls ( $\hat{\eta}^*, \hat{\xi}^*$ ) as well as the relative error depend on  $\beta$ .

Unlike the rate-matching control in Chapter 4, which stabilizes the entire queue-length distri-

Figure 5.3: Plots over one cycle for the example in Table 5.1. The first row shows the expected periodic steady-state virtual waiting time (solid blue line), the expected steady-state workload (the dashed red line) and arrival rate multiplied by the mean waiting time (the dotted green line) for  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right) in the base case ( $\beta, \gamma$ ) = (0.2, 2.5). The second row shows arrival rate divided by  $\rho$  and service rate for  $\rho = 0.8$  (left) and  $\rho = 0.95$  in the same base case. The optimal control parameters are ( $\eta_{\rho}^*, \xi_{\rho}^*$ ) = (5.80, 0.84) for  $\rho = 0.8$  and (27.7, 0.93) for  $\rho = 0.95$ . The maximum minus minimum of  $E[W_y]$  over a cycle equals 0.0321 for  $\rho = 0.8$  and 0.1425 for  $\rho = 0.95$ .



bution, the optimal modified  $(\eta, \xi)$  control neither stabilizes the mean perfectly nor does it stabilize the entire waiting-time distribution. However, it appears to do a reasonable job of both. Figures 5.4, 5.5 and 5.6 illustrate by showing plots of the time-varying (i) standard deviation  $SD[W_y]$ , (ii)

**Table 5.2:** The solutions to the minimum-deviation optimization problem in (5.8) for the sinusoidal model in Table 5.1 except  $\beta$  has been increased to  $\beta = 0.8$  from 0.2. The reported average mean waiting times are reported with and without space scaling.

ρ	0.8	0.9	0.95	0.975
$\beta_{ ho} \equiv \beta$	0.8	0.8	0.8	0.8
$\gamma_ ho$	0.1	0.025	0.00625	0.0015625
$\eta^*_ ho$	6.08	15.4	33.6	70.3
$\hat{\eta}_{\rho}^* \equiv (1-\rho)/\rho\eta_{\rho}^*$	1.52	1.71	1.77	1.80
$\xi^*_ ho$	0.874	0.893	0.929	0.960
$\hat{\xi}_{\rho}^* \equiv (1-\xi_{\rho}^*)/(1-\rho)$	0.63	1.07	1.42	1.60
$\max E[W_y] - \min E[W_y]$	0.54	1.32	2.28	4.55
avg $E[W_y]$	4.33	10.68	23.97	51.76
$(1-\rho)$ avg $E[W_y]/\rho$	1.08	1.19	1.14	1.26
relative error	12.5%	12.4%	9.5%	8.8%

the delay probability  $P(W_y > 0)$  and (iii) the full complementary cdf (ccdf)  $\{P(W_y > x) : x \ge 0\}$ for the two cases in Figure 5.3, i.e., for  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right).

Very roughly, Figures 5.3, 5.4, 5.5 and 5.6 are consistent with the time-varying waiting-time distribution being exponential as in the M/M/1 stationary model. We should not be surprised that the results look similar for  $\rho = 0.8$  and  $\rho = 0.95$  because they are scaled to be part of the family of systems satisfying the heavy-traffic limit. To consider a very different case, Figure 5.7 shows plots of the estimated waiting-time ccdf (left) and pdf (right) for  $\rho = 0.95$  in Figure 5.1, which is not scaled. Without the scaling, the cycles are relatively short. Figure 5.7, especially the pdf, shows much more varied behavior in this difficult short-cycle setting.

We now show two candidate modifications of the control used in Figure 5.1. First, Figure 5.8 shows the analog of Figure 5.1, where we fix  $\xi = 1$  and only use the single control parameter  $\eta$ . As we remarked in Remark 5.2.2 in §5.2, if we let  $\xi = 1$  and optimize over  $\eta$ , then for  $\rho = 0.8$  we get  $\eta_{\rho}^* = 5.93$  and a maximum deviation of 0.4109, which yields about 10% relative error instead of less than 1%. For  $\rho = 0.95$ ,  $\eta_{\rho}^* = 28.3$ , the maximum deviation is 3.034 and the relative error is



Figure 5.4: Plots over one cycle for the example in Table 5.1. Estimates of the periodic steadystate standard deviation  $SD[W_y]$  for  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right), shown in red solid line. Also displayed are the fluid arrival rate  $\lambda_f = 1 + s(t)$  (blue dashed line) and fluid service rate  $\mu_f = 1 + \xi_{\rho}^* s(t - \eta_{\rho}^*)$  (blue dotted line).



Figure 5.5: Plots over one cycle for the example in Table 5.1. Estimates of the periodic steadystate probability of delay  $P(W_y > 0)$  for  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right), shown in red solid line. Also displayed are the fluid arrival rate  $\lambda_f = 1 + s(t)$  (blue dashed line) and fluid service rate  $\mu_f = 1 + \xi_{\rho}^* s(t - \eta_{\rho}^*)$  (blue dotted line).



Figure 5.6: Plots over one cycle for the example in Table 5.1. Estimates of the periodic steadystate ccdf  $P(W_y > x)$  for four values of y: 0, 1/4, 1/2, 3/4 for  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right).



Figure 5.7: Plots over one cycle for the example in Figure 5.1 with  $\rho = 0.95$  but without the heavy-traffic scaling. Estimates of the ccdf (left) and pdf (right) for four values of y: 0, 1/4, 1/2, 3/4.

about 14%.

Second, Figure 5.9 shows the consequences of a direct HT approximation in the setting of Figure 5.1, obtained by letting  $\hat{\eta}^* \approx 1.45$ ,  $\eta_{\rho} \approx 1.45/(1-\rho)$ ,  $\hat{\xi}^* \approx 1.80$  and  $\xi_{\rho} \approx 1-1.8(1-\rho)$ , based on Table 5.1. For  $\rho = 0.8$ ,  $(\eta_{\rho}^*, \xi_{\rho}^*) = (7.25, 0.64)$  and the maximum deviation is 0.6005, yielding about 15% relative error. For  $\rho = 0.95$ ,  $(\eta_{\rho}^*, \xi_{\rho}^*) = (29.0, 0.91)$  and the maximum deviation is 0.9220



Figure 5.8: Estimates of the expected waiting time  $E[W_y]$  for the one-parameter  $\eta$  control with  $\xi \equiv 1$ , for the sinusoidal example in Figure 5.1 with parameter triple  $(\rho, \beta, \gamma) = (0.8, 0.2, 0.1)$  and  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right). For  $\rho = 0.8$ ,  $\eta_{\rho}^* = 5.93$  the maximum deviation is 0.4109 and the relative error is about 10%; for  $\rho = 0.95$ ,  $\eta_{\rho}^* = 28.3$ , the maximum deviation is 3.034 and the relative error is about 14%.

yielding about 5% relative error. Unlike in Figure 5.8, we see that the direct HT approximation improves as  $\rho$  increases, but the direct two-parameter optimal control is better.

Finally, Figure 5.10 plots two deterministic functions associated with the diffusion limit for the case  $\beta = 0.2$ ,  $\gamma = 2.5$ ,  $\hat{\eta} = 1.45$  and  $\hat{\xi} = -1.8$ . On the left appears  $\hat{M}(t) = -\hat{\eta}s(t) + \hat{\xi}S(t) = -1.5\beta\sin(\gamma t) - 1.5(\beta/\gamma)(1 - \cos(\gamma t))$  together with  $s(t) = \beta\sin(\gamma t)$  and  $S(t) = (\beta/\gamma)(1 - \cos(\gamma t))$ . On the right appears the diffusion limit for the net input  $\hat{X}(t) = -t - M(t)$  when  $c_x = 0$ . The plot on the right is consistent with condition (5.59) for no workload or waiting when  $c_x = 0$  in Corollary 5.6.5.



Figure 5.9: Estimates of the expected waiting time  $E[W_y]$  (solid red line) with the heavy-traffic control exploiting the estimated limiting controls  $\hat{\eta}^* \approx 1.45$  and  $\hat{\xi}^* = 1.8$ , so that  $\eta_{\rho}^* \approx 1.45/(1-\rho)$ and  $\xi_{\rho}^* \approx 1 - 1.8(1-\rho)$ . The plots are for the sinusoidal example in Figure 5.1 with parameter triple  $(\rho, \beta, \gamma) = (0.8, 0.2, 0.1)$  and  $\rho = 0.8$  (left) and  $\rho = 0.95$  (right). Also displayed are  $E[L_y]$ ,  $\lambda_f E[W_y]$  and 95% confidence interval bands, which require zooming in to see.

#### 5.9.1.2 Non-Sinusoidal Examples

We now turn to non-sinusoidal examples. We consider the piecewise-linear single-peak periodic arrival-rate function:

$$\lambda_{f}(y) = 1 - \beta + \frac{2\beta}{pC}y, \quad 0 \le y < pC \text{ and} \\ \lambda_{f}(y) = 1 + (\frac{1+p}{1-p})\beta - \frac{2\beta}{C(1-p)}y, \quad pC \le y < C,$$
(5.82)

where  $p \in [0,1)$ . This arrival-rate function increases linearly from  $1 - \beta$  to  $1 + \beta$  on [0, pC] and decreases linearly from  $1 + \beta$  to  $1 - \beta$  on [pC, C]. The periodic arrival rate function with traffic intensity  $\rho$  within one cycle is then  $\rho \lambda_f(y)$ .

Table 5.3 reports the optimal  $\eta$  and  $\xi$  for p = 1/2 (symmetric) and 1/3 (asymmetric) for  $\rho = 0.8$ ,  $\beta = 0.5$  and cycle length C = 60. We consider the two objective functions: the maximum expected waiting time and the maximum expected waiting time deviation. Then the following Figures 5.11 and 5.12 plot the expected waiting time under the optimal control for the cases p = 1/2 and p = 1/3. Table 5.3 shows a clear difference in the objective functions in the asymmetric case, but



Figure 5.10: Deterministic functions associated with the diffusion limit for the case  $\beta = 0.2$ ,  $\gamma = 2.5$ ,  $\hat{\eta} = 1.45$  and  $\hat{\xi} = -1.8$ . On the left appears  $\hat{M}(t) = -\hat{\eta}s(t) + \hat{\xi}S(t) = -1.5\beta\sin(\gamma t) - 1.5(\beta/\gamma)(1 - \cos(\gamma t))$  together with  $s(t) = \beta\sin(\gamma t)$  and  $S(t) = (\beta/\gamma)(1 - \cos(\gamma t))$ . On the right appears the diffusion limit for the net input  $\hat{X}(t) = -t - M(t)$  when  $c_x = 0$ , showing that condition condition (5.59) holds.

not in the symmetric case.

p	1/2	1/2	1/3	1/3
objective	(5.7)	(5.8)	(5.7)	(5.8)
ρ	0.8	0.8	0.8	0.8
C	60	60	60	60
$\eta_{ ho}$	5.9	5.9	5.8	5.8
$\frac{1- ho}{ ho}\eta_{ ho}$	1.48	1.48	1.45	1.45
$\xi_ ho$	0.86	0.86	0.84	0.87
$\frac{1-\xi_{ ho}}{(1- ho)}$	0.7	0.7	0.8	0.65
$\max E[W_y]$	4.1787	4.1787	4.2160	4.2175
$\max E[W_y] - \min E[W_y]$	0.3081	0.3081	0.4200	0.3468
avg $E[W_y]$	4.08	4.08	4.08	4.09
relative error	7.6%	7.6%	10.3%	8.5%

**Table 5.3:** The optimal  $\eta$  and  $\xi$  for  $\rho = 0.8$ ,  $\beta = 0.5$ , cycle length C = 60.

Figure 5.11: For  $\rho = 0.8$ , p = 1/2,  $\beta = 0.5$ , this figure plots the expected waiting time under optimal control that minimizes the maximum  $E[W_y]$ .



Figure 5.12: For  $\rho = 0.8$ , p = 1/3,  $\beta = 0.5$ , this figure plots the expected waiting time under optimal control that minimizes the maximum  $E[W_y]$ .



Just as for the sinusoidal examples (compare Tables 5.1 and 5.2), stabilizing the mean waiting time becomes more difficult as  $\beta$  increases toward the upper limit 1. The most difficult case is  $\beta = 1$ , where the arrival-rate function is 0 at the end points 0 and C. Table 5.4 shows the severe performance degradation in this case. Insight into the difficult cases with zero or near-zero  $\lambda_f(t)$  can be gained from the time-varying Little's law in Corollary 5.6.4 and steps (5.71) and (5.72) in the proof of Theorem 5.6.2.

p	1/2	1/2	1/3	1/3
objective	(5.7)	(5.8)	(5.7)	(5.8)
ρ	0.8	0.8	0.8	0.8
C	60	60	60	60
$\eta_{ ho}$	5.9	5.9	6.3	6.3
$\frac{1- ho}{ ho}\eta_{ ho}$	1.48	1.48	1.58	1.58
$\xi_ ho$	0.89	0.92	0.84	0.88
$\frac{1-\xi_{\rho}}{(1-\rho)}$	0.55	0.4	0.8	0.6
$\max E[W_y]$	4.8087	4.8812	4.7656	4.9072
$\max(E[W_y]) - \min(E[W_y])$	1.3715	1.2726	1.8288	1.7775
avg $E[W_y]$	4.32	4.38	4.35	4.40
relative error	31.7%	29.0%	42.0%	40.4%

**Table 5.4:** The optimal  $\eta$  and  $\xi$  for  $\rho = 0.8$ ,  $\beta = 1$ , cycle length C = 60.

# 5.9.2 In the Alternative Scaling of §5.8

We now consider four simulation examples in the alternative heavy-traffic scaling in §5.8. This is the same heavy-traffic scaling as in §3.6. We consider the base case of  $\beta = 1$ ,  $\gamma = 2.5$ , and use

$$(\bar{\lambda}_{\rho}, \beta_{\rho}, \gamma_{\rho}, b_{\rho}) = (\rho, (1-\rho)\beta, (1-\rho)^2\gamma, (1-\rho)^{-1}b).$$

Specifically, we consider cases with  $\rho = 0.84, 0.92, 0.96, 0.98$ . Here we use the lags  $\eta_{\rho} = 5.25, 11.5, 24, 49$  calculated by  $\rho/(1-\rho)$ , the scaler  $\xi_{\rho} = \rho$ . (These are consistent with Theorem 5.8.1.)

Figures 5.13-5.14 show the expected periodic steady-state waiting time (the solid blue line) and the expected steady-state workload (the dashed red line). Figures 5.13 and 5.14 show that the stabilization is not achieved well for the lower traffic intensities, but the stabilization improves for both curves as  $\rho$  increases. Both processes get quite well stabilized at  $\rho = 0.98$ , consistent with Theorem 5.8.1.

Figure 5.13: the expected periodic steady-state virtual waiting time (the blue line) and the expected steady-state workload (the red line) for  $\rho = 0.84$ ,  $\beta = 0.16$ ,  $\gamma = 0.064$ ,  $\eta_{\rho} = 5.25$ ,  $\xi_{\rho} = 0.84$ , yielding a maximum deviation 0.0699 (left) and  $\rho = 0.92$ ,  $\beta = 0.08$ ,  $\gamma = 0.016$ ,  $\eta_{\rho} = 11.5$ ,  $\xi_{\rho} = 0.92$ , yielding a maximum deviation 0.0408 (right).



Figure 5.14: the expected periodic steady-state virtual waiting time (the blue line) and the expected steady-state workload (the red line) for  $\rho = 0.96$ ,  $\beta = 0.04$ ,  $\gamma = 0.004$ ,  $\eta_{\rho} = 24$ ,  $\xi_{\rho} = 0.96$ ,, yielding a maximum deviation 0.0228 (left) and  $\rho = 0.98$ ,  $\beta = 0.02$ ,  $\gamma = 0.001$ ,  $\eta_{\rho} = 49$ ,  $\xi_{\rho} = 0.98$ , yielding a maximum deviation 0.0070 (right).



# 5.10 Conclusions

In this chapter we extended the rare-event simulation algorithm for the periodic  $GI_t/GI/1$  model in Chapter 3 to the periodic  $GI_t/GI_t/1$  model and applied the new algorithm to study methods to stabilize the expected (virtual) waiting time over time. We studied the modification in (5.4) of the rate-matching service-rate control in (4.1) to include a time lag  $\eta$  and a damping factor  $\xi$ . We developed and applied a simulation search algorithm to find optimal pairs of control parameters  $(\eta, \xi)$  for the control problems in (5.7) and (5.8). Thus, we obtained a practical solution to the open problem in Chapter 4 of developing an effective way to stabilize the expected waiting time in the periodic single-server model.

We also established supporting heavy-traffic limits for the general periodic  $G_t/G_t/1$  model and showed that the control problems in (5.7) and (5.8) converge to associated diffusion control parameters with appropriate scaling. The scaling involves the conventional heavy-traffic scaling associated in which  $\rho \uparrow 1$ , so that time is scaled by  $(1 - \rho)^{-2}$  while space is scaled by  $1 - \rho$ , but in addition to gain insight into the time-varying behavior, we identify and study three different scalings of the arrival rate function. As observed by Falin (1989), if the arrival-rate function is left unscaled, then the heavy-traffic limit is the same as if the periodicity were not present at all. A major conclusion is that important insight into the time-varying performance can be gained by scaling the arrival-rate function as well. Moreover, as illustrated by §5.6 and §5.8, there are two different natural scalings: First, there is the stronger scaling in the fluid scale in §5.6 as in Whitt (2015) and, second, there is the weaker scaling in the diffusion scale in §5.8 as in Whitt (2014) and §3.6. In the weaker scaling, the rate-matching control from Chapter 4 stabilizes both the queue length and the waiting time, but in the stronger fluid scaling we see significant differences, consistent with the simulation results in Figure 5.1. This insightful scaling in the fluid scale also yields a limiting diffusion control problem.

We conducted extensive simulation algorithms showing that the new  $(\eta, \xi)$  control is effective in stabilizing the expected waiting time. However, unlike the rate-matching control for the queue length process in Chapter 4, the new modified rate-matching control does not stabilize the expected waiting time perfectly, either for fixed  $\rho$  or in the heavy-traffic limit. However, Figures 5.1 and 5.3 shows that it stabilizes it remarkably well, while Figures 5.4, 5.5 and 5.6 show that it stabilizes the full waiting time distribution quite well too. We have shown the performance of the new control for mostly sinusoidal, but also piecewise-linear single-peak arrival rate functions, and it remains to run simulation experiments to see the performance for arrival rate functions with two or more peaks. More complicated methods may be needed in case the new control is not effective.

It is interesting to consider the performance impact of time-varying arrivals. In §5.1 we observed that the difference between the stable average waiting time in Figure 5.1 and the value  $\rho/(1-\rho)$ for the stationary model (4 on the left and 19 on the right) might be called "the average cost of periodicity," but Example 5.2.1 showed that the overall average expected waiting time with a service-rate control could be much less than in the stationary model. It remains to investigate more carefully.

Indeed, there remain many opportunities for future research, including the open problems mentioned in §5.1.5. It also remains to directly solve the diffusion control problems with objectives (5.7) and (5.8) resulting from Theorem 5.6.2. And there are other methods worth carefully studying, such as modifications of the iterated staffing alrgorithm (ISA) from Feldman et al. (2008) for single-server models.

Another future research topic is to generalize our results to a system with a fixed small number

of servers (more than one servers). The HT limits extend to this case as illustrated in Iglehart and Whitt (1970a,b), so the HT TV Little's Law established in theorem 5.6.2 is still valid. But it is worth careful studying how to find the optimal damped time-lag control and how it performs for this case.

Finally, we mention that the methods in this chapter generalize and can be applied to other problems. First, the rare-event simulation algorithm in §5.5 applies to any  $GI_t/GI_t/1$  model with other service-rate controls. Second, the heavy-traffic limits in §5.6 and §5.8 evidently extend to general  $G_t/G_t/1$  models with other service-rate controls. More generally, simulation of converging stochastic processes is a promising way to numerically solve complex diffusion control problems. Part III

Bibliography

# Bibliography

- Abate, J., Choudhury, G. L., and Whitt, W. (1993). Calculation of the GI/G/1 steady-state waiting-time distribution and its cumulants from Pollaczek's formula. Archiv für Elektronik und Übertragungstechnik, 47(5):311–321.
- Abate, J., Choudhury, G. L., and Whitt, W. (1994). Asymptotics for steady-state tail probabilities in structured Markov queueing models. *Stochastic Models*, 10(1):99–143.
- Abate, J. and Whitt, W. (1994). A heavy-traffic expansion for the asymptotic decay rates of tail probabilities in multi-channel queues. *Operations Research Letters*, 14(3):663–680.
- Abate, J. and Whitt, W. (1998). Calculating transient characteristics of the Erlang loss model by numerical transform inversion. *Stochastic Models*, 15:223–230.
- Adusumilli, K. M. and Hasenbein, J. J. (2010). Dynamic admission and service rate control of a queue. *Queueing Systems*, 66:131–154.
- Anantharam, V. (1988). How large delays build up in a GI/G/1 queue. Queueing Systems, 5:345–368.
- Anantharam, V., Heidelberger, P., and Tsoucas, P. (1990). Analysis of rare events in continuous time Markov chains via time reversal and fluid approximation. IBM Research Report RC 16280, Yorktown Heights, New York.
- Asmussen, S. (1982). Conditioned limit theorems relating a random walk to its associate. Advances in Applied Probability, 14:143–170.
- Asmussen, S. (1985). Conjugate processes and the simulation of ruin problems. *Stochastic Processes and their Applications*, 20:213–229.
- Asmussen, S. (2003). Applied Probability and Queues. Springer, New York, second edition.
- Asmussen, S. and Albrecher, H. (2010). Ruin Probabilities. World Scientific, Singapore, second edition.
- Asmussen, S., Binswanger, K., and Hojgaard, B. (2000). Rare events simulation for heavy-tailed distributions. Bernoulli, 6:303–322.
- Asmussen, S. and Glynn, P. W. (2007). Stochastic Simulation. Springer, New York, second edition.

- Asmussen, S. and Rolski, T. (1994). Risk theory in a periodic environment: the Cramer-Lundberg approximation and Lundberg's inequality. *Mathematics of Operations Research*, 19(2):410–433.
- Bassamboo, A., Juneja, S., and Zeevi, A. (2006). On the inefficiency of state-independent importance sampling in the presence of heavy-tails. *Operations Research Letters*, 34:521–531.
- Ben-Tal, A., El-Ghaoui, L., and Nemirovski, A. (2009). Robust Optimization. Princeton University Press, Princeton.
- Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. SIAM Review, 53(3):464–501.
- Bertsimas, D. and Mourtzinou, G. (1997). Transient laws of nonstationary queueing systems and their applications. *Queueing Systems*, 25:315–359.
- Beyer, H. G. and Sendhoff, B. (2007). Robust optimization a comprehensive survey. Computer Methods in Applied Mechanics and Engineering, 196(33-34):3190–3218.
- Billingsley, P. (1999). Convergence of Probability Measures. Wiley, New York.
- Blanchet, J. and Glynn, P. (2008). Efficient rare-event simulation for the maximum of a heavy-tailed random walk. The Annals of Applied Probability, 18(4):1351–1378.
- Blanchet, J., Glynn, P., and Liu, J. C. (2007). Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue, QUESTA, 57:99–113.
- Blanchet, J. and Lam, H. (2014). Rare-event simulation for many-server queues. Mathematics of Operations Research, 39(4):1142–1178.
- Blanchet, J. and Liu, J. C. (2008). State-dependent importance sampling for regularly varying random walks. Advances in Applied Probability, 40:1104–1128.
- Bucklew, J. (1990). Large Deviation Techniques in Decision, Simulation and Estimation. J. Wiley & Sons, Inc., New York.
- Budhiraja, A. and Lee, C. (2009). Stationary distribution convergence for generalized Jackson networks in heavy traffic. *Mathematics of Operations Research*, 34(1):45–56.
- Cinlar, E. (1975). Introduction to Stochastic Processes. Prentice-Hall, Englewood Cliffs, NJ.
- Chang, W. (1970). Single-server queuing processes in computing systems. IBM Systems Journal, 9(1):36–71.
- Chen, H. and Mandelbaum, A. (1991). Discrete flow networks: bottleneck analysis and fluid approximations. Math. Operations Research, 16(2):408–446.
- Chen, H. and Schmeiser, B. W. (1992). Simulation of Poisson processes with trigonometric rate. In *Proceed*ings of the 1992 Winter Simulation Conference, pages 609–617. ACM.

- Chen, H. and Schmeiser, B. W. (2015). I-SMOOTH: Iteratively smoothing mean-constrained and nonnegative piecewise-constant functions. *INFORMS Journal on Computing*, 25(3):432–445.
- Chen, H. and Schmeiser, B. W. (2017). MNO-PQRS: Max nonnegativity ordering piecewise-quadratic rate smoothing. ACM TOMACS, 27(3):1–19.
- Choudhury, G. L., Lucantoni, D. M., and Whitt, W. (1996). Squeezing the most out of ATM. *IEEE Transactions on Communications*, 44(2):203–217.
- Choudhury, G. L., Mandelbaum, A., Reiman, M. I., and Whitt, W. (1997). Fluid and diffusion limits for queues in slowly changing random environments. *Stochastic Models*, 13(1):121–146.
- Choudhury, G. L. and Whitt, W. (1994). Heavy-traffic asymptotic expansions for the asymptotic decay rates in the *BMAP/G/1* queue. *Stochastic Models*, 10(2):453–498.
- Cottrell, M., Fort, J.-C., and Malgouyres, G. (1983). Large deviations and rare events in the study of stochastic algorithms. *IEEE Transactions on Automatic Control*, 28:907–920.
- Daley, D. J. and Vere-Jones, D. (2008). An Introduction to the Theory of Point Processes, volume II. Springer, Oxford, U. K., second edition.
- Defraeye, M. and van Nieuwenhuyse, I. (2013). Controlling excessive waiting times in small service systems with time-varying demand: An extension of the ISA algorithm. *Decision Support Systems*, 54(4):1558– 1567.
- Devetsikiotis, M. and Townsend, K. (1992a). On the efficient simulation of large communication networks using importance sampling. In *Proceedings of IEEE Globecom '92*. IEEE Computer Society Press.
- Devetsikiotis, M. and Townsend, K. (1992b). Statistical optimization of dynamic importance sampling parameters for efficient simulation of communication networks. In *Proceedings of IEEE ICC '92.* IEEE Computer Society Press.
- Devetsikiotis, M. and Townsend, K. (1993). A dynamic importance sampling methodology for the efficient estimation of rare event probabilities in regenerative simulations of queueing systems. *IEEE/ACM Transactions on Networking*, 1(3):293–305.
- Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer, New York.
- Donsker, M. and Varadhan, S. (1975a). Asymptotic evaluation of certain Markov process expectations for large time. i. Communications on Pure and Applied Mathematics, 28:1–47.
- Donsker, M. and Varadhan, S. (1975b). Asymptotic evaluation of certain Markov process expectations for large time. ii. Communications on Pure and Applied Mathematics, 28:279–301.
- Dupuis, P., Leder, K., and Wang, H. (2006). Importance sampling for sums of random variables with regularly varying tails. ACM TOMACS, 17(3):Article No. 14.

- Dupuis, P. and Wang, H. (2004). Importance sampling, large deviations, and differential games. Stochastics and Stochastics Reports, 76:481–508.
- Edie, L. C. (1954). Traffic delays at toll booths. Operations Research, 2:107-138.
- Eick, S. G., Massey, W. A., and Whitt, W. (1993).  $M_t/G/\infty$  queues with sinusoidal arrival rates. Management Science, 39:241–252.
- Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.
- Falin, G. I. (1989). Periodic queues in heavy traffic. Advances in Applied Probability, 21:485-487.
- Feldman, Z., Mandelbaum, A., Massey, W. A., and Whitt, W. (2008). Staffing of time-varying queues to achieve time-stable performance. *Management Science*, 54(2):324–338.
- Feller, W. (1971). An Introduction to Probability Theory and its Applications. John Wiley, New York, second edition edition.
- Fralix, B. H. and Riano, G. (2010). A new look at transient versions of Little's law. Journal of Applied Probability, 47:459–473.
- Frater, M., Lenon, T., and Anderson, B. (1991). Optimally efficient estimation of the statistics of rare events in queueing networks. *IEEE Trans- actions on Automatic Control*, 36:1395–1405.
- Frater, M. R. and Anderson, B. (1989). Fast estimation of the statistics of excessive backlogs in tandem networks of queues. Australian Telecommunications Research, 23:49–55.
- Fu, M. C. (2015). Handbook of Simulation Optimization, volume 216 of International Series in Operations Research and Management Science. Springer, New York.
- George, J. M. and Harrison, J. M. (2001). Dynamic control of a queue with adjustable service rate. *Operations Research*, 49(5):720–731.
- Gerhardt, I. and Nelson, B. L. (2009). Transforming renewal processes for simulation of nonstationary arrival processes. *INFORMS Journal on Computing*, 21:630–640.
- Glasserman, P. and Kou, S. (1993). Overflow probabilities in Jackson networks. In Proceedings of the 32nd Conference on Decision and Control, pages 3178–3182. IEEE Press.
- Glasserman, P. and Kou, S. (1995). Analysis of an importance sampling estimator for tandem queues. TOMACS, 5:22–42.
- Glasserman, P. and Wang, Y. (1997). Counterexamples in importance sampling for large deviations probabilities. Annals of Applied Probability, 7:731–746.
- Glynn, P. and Iglehart, D. (1989). Importance sampling for stochastic simulations. Management Science, 35:1367–1392.

- Glynn, P. W. and Whitt, W. (1994). Logarithmic asymptotics for steady-state tail probabilities in a singleserver queue. In Studies in Applied Probability, Papers in Honour of Lajos Takàcs, pages 131–156. Applied Probability Trust, Sheffield, England.
- Green, L. V., Kolesar, P. J., and Whitt, W. (2007). Coping with time-varying demand when setting staffing requirements for a service system. *Production Operations Management*, 16:13–29.
- Hammersley, J. M. and Handscomb, D. C. (1964). Monte Carlo Methods. Methuen and Co., Ltd, London.
- Harrison, J. M. and Lemoine, A. J. (1977). Limit theorems for periodic queues. Journal of Applied Probability, 14:566–576.
- He, B., Liu, Y., and Whitt, W. (2016). Staffing a service system with non-Poisson nonstationary arrivals. Probability in the Engineering and Informational Sciences, 30(1):593–621.
- Heidelberger, P. (1995). Fast simulation of rare events in queueing and reliability models. ACM Transactions on Modeling and Computer Simulation, 5:43–85.
- Heyman, D. P. (1968). Optimal operating policies for M/G/1 queueing systems. Operations Research, 16(2):362–382.
- Heyman, D. P. and Whitt, W. (1984). The asymptoic behavior of queues with time-varying arrival. *Journal* of Applied Probability, 21(1):143–156.
- Iglehart, D. L. and Whitt, W. (1970a). Multiple channel queues in heavy traffic, I. Advances in Applied Probability, 2(1):150–177.
- Iglehart, D. L. and Whitt, W. (1970b). Multiple channel queues in heavy traffic, II: Sequences, networks and batches. *Advances in Applied Probability*, 2(2):355–369.
- Jacod, J. and Shiryaev, A. N. (1987). Limit Theorems for Stochastic Processes. Springer, New York.
- Jennings, O. B., Mandelbaum, A., Massey, W. A., and Whitt, W. (1996). Server staffing to meet time-varying demand. *Management Science*, 42:1383–1394.
- Jian, N. and Henderson, S. G. (2015). An introduction to simulation optimization. In Yilmaz, L., Chan, W. K. V., Moon, I., Roeder, T. M. K., Macal, C., and Rossetti, M. D., editors, *Proceedings of the 2015 Winter Simulation Conference, Huntington Beach, CA, December 6-9, 2015*, pages 1780–1794. ACM.
- Kim, S.-H., Vel, P., Whitt, W., and Cha, W. C. (2015). Poisson and non-Poisson properties of appointmentgenerated arrival processes: The case of an endocrinology clinic. *Operations Research Letters*, 43:247– 253.
- Kim, S.-H. and Whitt, W. (2014). Are call center and hospital arrivals well modeled by nonhomogeneous Poisson processes? *Manufacturing and Service Operations Management*, 16(3):464–480.
- Kleinrock, L. (1964). Communication Nets: Stochastic Message Flow and Delay. Dover, New York.

- Kolesar, P. J., Rider, P. J., Craybill, T. B., and Walker, W. E. (1975). A queueing-linear-programming approach to scheduling police patrol cars. *Operations Research*, 23:1045–1062.
- Koopman, B. O. (1972). Air-terminal queues under time-dependent conditions. Operations Research, 20:1089–1114.
- Kwon, S. and Gautam, N. (2016). Guaranteeing performance based on time stability for energy-efficient data centers. *IIE Transactions*, 48(9):812–825.
- L'Ecuyer, P. (2012). Random number generation. In Gentle, J. E., Hardle, J. K., and Mori, Y., editors, Handbook of Computational Statistics, chapter 3, pages 35–71. Springer, New York.
- Leemis, L. M. (1991). Nonparametric estimation of the cumulative intensity function for a nonhomogeneous poisson process. *Management Science*, 37:886–900.
- Lehtonen, T. and Nyrhinen, H. (1992). On asymptotically efficient simulation of ruin probabilities in a Markovian environment. Scandinavian Actuarial Journal, 1:60–75.
- Lemoine, A. J. (1981). On queues with periodic Poisson input. Journal of Applied Probability, 18:889–900.
- Lemoine, A. J. (1989). Waiting time and workload in queues with periodic Poisson input. *Journal of Applied* Probability, 26(2):390–397.
- Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. Naval Research Logistics Quarterly, 26(1):403–413.
- Li, A., Whitt, W., and Zhao, J. (2016). Staffing to stabilize blocking in loss models with non-Poisson nonstationary arrivals. *Queueing Systems*, 30(2):185–211.
- Liu, R., Kuhl, M. E., Liu, Y., and Wilson, J. R. (2018). Modeling and simulation of nonstationary non-Poisson arrival processes. *INFORMS Journal on Computing*. published online doi.org/10.1287/ijoc.2018.0828.
- Liu, Y. (2018). Staffing to stabilize the tail probability of delay in service systems with time-varying demand. Operations Research, 66(2). published online February 8, 2018.
- Liu, Y. and Whitt, W. (2011). A network of time-varying many-server fluid queues with customer abandonment. Operations Research, 59:835–846.
- Liu, Y. and Whitt, W. (2012a). The  $G_t/GI/s_t + GI$  many-server fluid queue. Queueing Systems, 71:405–444.
- Liu, Y. and Whitt, W. (2012b). Stabilizing customer abandonment in many-server queues with time-varying arrivals. *Operations Research*, 60(6):1551–1564.
- Loynes, R. (1962). The stability of a queue with non-independent inter-arrival and service times. *Mathe*matical Proceedings of the Cambridge Philosophical Society, 58(3):497–520.

- Ma, N. and Whitt, W. (2015a). Efficient simulation of non-Poisson non-stationary point processes to study queueing approximations. *Statistics and Probability Letters*, 102:202–207.
- Ma, N. and Whitt, W. (2015b). Using simulation to study service-rate controls to stabilize performance in a single-server queue with time-varying arrival rate. In Yilmaz, L., Chan, W. K. V., Moon, I., Roeder, T. M. K., Macal, C., and Rossetti, M. D., editors, *Proceedings of the 2015 Winter Simulation Conference, Huntington Beach, CA, December 6-9, 2015*, pages 1–12. ACM.
- Ma, N. and Whitt, W. (2018a). Minimizing the maximum expected waiting time in a periodic single-server queue with a service-rate control. *Stochastic Systems*, forthcoming; available at: http://www.columbia.edu/~ww2040/min\_102018.pdf.
- Ma, N. and Whitt, W. (2018b). A rare-event simulation algorithm for periodic single-server queues. IN-FORMS Journal on Computing, 30(1):71–89.
- Mandelbaum, A. and Massey, W. A. (1995). Strong approximations for time-dependent queues. Mathematics of Operations Research, 20:33–64.
- Massey, W. A. (1985). Asymptotic analysis of the time-varying M/M/1 queue. Mathematics of Operations Research, 10:305–327.
- Massey, W. A., Parker, G. A., and Whitt, W. (1996). Estimating the parameters of a nonhomogeneous Poisson process with linear rate. *Telecommunication Systems*, 5:361–388.
- Massey, W. A. and Whitt, W. (1994). Unstable asymptotics for nonstationary queues. Math. Operations Research, 19:267–291.
- May, A. D. and Keller, H. E. M. (1967). A deterministic queueing model. *Transportation Research*, 1:117–128.
- Morales, M. (2004). On a surplus process under a periodic environment: a simulation approach. North American Actuarial Journal, 8(4):76–89.
- Nelson, B. L. and Gerhardt, I. (2011). Modeling and simulating renewal nonstationary arrival processes to facilitate analysis. *Journal of Simulation*, 5:3–8.
- Newell, G. F. (1968a). Queues with time dependent arrival rates, i: the transition through saturation. Journal of Applied Probability, 5:436–451.
- Newell, G. F. (1968b). Queues with time dependent arrival rates, ii: the maximum queue and the return to equilibrium. *Journal of Applied Probability*, 5:579–590.
- Newell, G. F. (1968c). Queues with time dependent arrival rates, iii: a mild rush hour. *Journal of Applied Probability*, 5:591–606.
- Newell, G. F. (1982). Applications of Queueing Theory. Chapman and Hall/CRC, New York, second edition.

Nicol, D. M. and Leemis, L. M. (2014). A continuous piecewise-linear nhpp intensity-function estimator.

In *Proceedings of the 2014 Winter Simulation Conference*, pages 498–509. Institute of Electrical and Electronics Engineers, Piscataway, NJ.

- Nieuwenhuis, G. (1989). Equivalence of functional limit theorems for stationary point processes and their Palm distributions. *Probability and Related Fields*, 81:593–608.
- Oliver, R. M. and Samuel, A. H. (1962). Reducing letter delays in post offices. Operations Research, 10:839– 892.
- Ong, K. L. and Taaffe, M. R. (1989). Nonstationary queues with interrupted poisson arrivals and unreliable/ repairable servers. *Queueing Systems*, 4:27–46.
- Pang, G., Talreja, R., and Whitt, W. (2007). Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probability Surveys*, 4:193–267.
- Parekh, S. and Walrand, J. (1989). A quick simulation method for excessive backlogs in networks of queues. IEEE Transactions on Automatic Control, 34:54–56.
- Pender, J. and Massey, W. A. (2017). Approximating and stabilizing dynamic rate Jackson networks with abandonment. *Probability in the Engineering and Information Sciences*, 31:1–42.
- Porteus, E. (1989). The case analysis section: The national cranberry cooperative. Interfaces, 19:29–39.
- Porteus, E. (1993a). Case analysis: Analyses of the national cranberry cooperative -1. tactical options. Interfaces, 23:21–39.
- Porteus, E. (1993b). Case analysis: Analyses of the national cranberry cooperative -2. environmental changes and implementation. *Interfaces*, 23:81–92.
- Puhalskii, A. (1994). On the invariance principle for the first passage time. Mathematics of Operations Research, 19(4):946–954.
- Rojas-Nandayapa, L. and Asmussen, S. (2007). Efficient simulation of finite horizon problems in queueing and insurance risk. QUESTA, 57:85–97.
- Rolski, T. (1981). Queues with nonstationary input stream: Ross's conjecture. Advances in Applied Probability, 13:603–618.
- Rolski, T. (1989a). Queues with nonstationary inputs. Queueing Systems, 5:113–130.
- Rolski, T. (1989b). Relationships between characteristics in periodic Poisson queues. *Queueing Systems*, 4(1):17–26.
- Rothkopf, M. H. and Oren, S. S. (1979). A closure approximation for the nonstationary M/M/s queue. Management Science, 25(6):522–534.
- Sadowsky, J. and Bucklew, J. (1990). On large deviations theory and asymptotically efficient monte carlo estimation. *IEEE Transactions on Information Theory*, 36:579–588.

- Sadowsky, J. S. (1991). Large deviations and efficient simulation of excessive backlogs in a GI/G/m queue. IEEE Transactions on Automatic Control, 36:1383–1394.
- Siegmund, D. (1976). Importance sampling in the monte carlo study of sequential tests. The Annals of Statistics, 4:673–684.
- Sigman, K. (1995). Stationary Marked Point Processes: An Intuitive Approach. Chapman and Hall/CRC, New York.
- Sigman, K. and Whitt, W. (2018). Marked point processes in discrete time. Columbia University, available at: http://www.columbia.edu/~ww2040/allpapers.html.
- Stolletz, R. (2008). Approximation of the nonstationary M(t)/M(t)/c(t)-queue using stationary models: the stationary backlog-carryover approach. European Journal of Operations Research, 190(2):478–493.
- Sun, X. and Whitt, W. (2018). Delay-based service differentiation with many servers and time-varying arrival rates. Stochastic Systems. doi/10.1287/stsy.2018.0015.
- Suriadi, S., Wynn, M. T., Xu, J., van der Aalst, W. M. P., and ter Hofstede, A. H. M. (2017). Discovering work prioritization patterns from event logs. *Decision Support Systems*, 100(August):77–92.
- Taaffe, M. R. and Ong, K. L. (1987). Approximating Ph(t)/M(t)/S/C queueing systems. Annals of Operations Research, 8:103–116.
- Tsoucas, P. (1989). *Rare events in series of queues.* IBM Research Report RC 15530, Yorktown Heights, New York.
- Wein, L. M. (1989). Capacity allocation in generalized Jackson networks. Operations Research Letters, 8:143–146.
- Weiss, A. (1986). A new technique for analyzing large traffic systems. Advances in Applied Probability, 18:506–532.
- Whitt, W. (1971). Weak convergence theorems for priority queues: Preemptive-resume discipline. *Journal* of Applied Probability, 8(1):74–94.
- Whitt, W. (1980). Continuity of generalized semi-markov processes. *Mathematics of Operations Research*, 5:494–501.
- Whitt, W. (1982). Approximating a point process by a renewal process: Two basic methods. *Operations Research*, 30:125–147.
- Whitt, W. (1983). The queueing network analyzer. Bell System Technical Journal, 62(9):2779–2815.
- Whitt, W. (1984). Departures from a queue with many busy servers. *Mathematics of Operations Research*, 9:534–544.

- Whitt, W. (1999). Dynamic staffing in a telephone call center aiming to immediately answer all calls. Operations Research Letters, 24:205–212.
- Whitt, W. (2002). Stochastic-Process Limits. Springer, New York.
- Whitt, W. (2014). Heavy-traffic limits for queues with periodic arrival processes. *Operations Research Letters*, 42:458–461.
- Whitt, W. (2015). Stabilizing performance in a single-server queue with time-varying arrival rate. *Queueing* Systems, 81:341–378.
- Whitt, W. (2016a). Heavy-traffic limits for a single-server queue leading up to a critical point. Operations Research Letters, 44:796–800.
- Whitt, W. (2016b). A Poisson limit for the departure process from a queue with many slow server. *Operations Research Letters*, 44:602–606.
- Whitt, W. (2018). Time-varying queues. Queueing Models and Service Management, 1(2):79–164.
- Whitt, W. and You, W. (2018). Time-varying robust queueing. Columbia University, http://www.columbia.edu/~ww2040.
- Whitt, W. and Zhang, X. (2018). Periodic Little's Law. *Operations Research*, forthcoming; available at: http://www.columbia.edu/~ww2040/allpapers.html.
- Whitt, W. and Zhao, J. (2017). Many-server loss models with non-poisson time-varying arrivals. Naval Research Logistics, 64(3):177–202.
- Xiong, Y., Murdoch, D. J., and Stanford, D. A. (2015). Perfect sampling of a single server queue with periodic Poisson arrivals. *Queueing Systems*, 80:15–33.
- Yadin, M. and Naor, P. (1963). Queueing systems with a removable service station. Operational Research Quarterly, 14:393–405.
- Yom-Tov, G. and Mandelbaum, A. (2014). Erlang R: a time-varying queue with reentrant customers, in support of healthcare staffing. *Manufacturing and Service Operations Management*, 16(2):283–299.

Part IV

Appendices

# Appendix A

# Supplement to Chapter Three

# A.1 Introduction

This is a supplement to Chapter 3 of the main thesis. In  $\S$ A.2 we elaborate on  $\S$ 3.3 of the main thesis. First, we further discuss the tail asymptotics and the asymptotic decay rate needed in the simulation. At the end, we present a couple of additional bounds and approximations. In  $\S$ A.3 we report results of additional simulation experiments applying the algorithm developed in the main thesis.

# A.2 More Results on the Asymptotics

In this section we continue the discussion of the tail asymptotics in §3.3.2 of the main thesis. We start in §A.2.1 by conjecturing the asymptotic form of the periodic steady-state distribution of RPBM. Then in §A.2.2 we review an asymptotic expansion from Abate and Whitt (1994), which yields an approximation for the asymptotic decay rate needed in the simulation. In §A.2.4 we compare the approximate values of the asymptotic decay rate to exact values. Finally, in §A.2.4 we derive the exact form of the asymptotic decay rate for hyperexponential models.

### A.2.1 Tail Asymptotics for RPBM

It remains to establish tail asymptotics for the periodic steady-state distribution of RPBM. However, we can see the form that tail asymptotics should take from the heavy-traffic scaling and the tail asymptotics established for the  $G_t/G/1$  model in §3.3.2 of the main thesis.

Let  $Z_y(\infty; c_x)$  be the periodic steady-state distribution of RPBM with variability parameter  $c_x$ as in Theorem 3.6.1 of the main thesis. From Corollary 3.3.3, we are led to conjecture that

$$e^{\theta^* b} P(Z_y(\infty; c_x) > b) \to A_y \quad \text{as} \quad b \to \infty,$$
 (A.1)

for some constant  $A_y$  and

$$\theta^* = \lim_{\rho \uparrow 1} \theta^*_{\rho} / (1 - \rho), \tag{A.2}$$

where  $\theta_{\rho}^{*}$  is the associated asymptotic decay rate for a family of  $G_t/G/1$  models converging to RPBM. We remark that there is a limit-interchange problem for the tail probability asymptotics, closely paralleling the limit-interchange problem associated with the heavy-traffic limit discussed in §3.6 of the main thesis.

Moreover, the asymptotic decay rate of the steady-state distribution of RPBM should coincide with that of RBM, which directly has an exponential steady-state distribution, i.e.,  $P(Z(\infty; c_x) > b) = e^{-2b/c_x^2}$ . In the next section we provide support for (A.2). Our numerical results show how to compute the tail probability  $P(Z_y(\infty; c_x) > b)$  assuming that these limits are valid.

## A.2.2 Asymptotic Expansions for the Asymptotic Decay Rate

We can develop useful approximations for the asymptotic decay rate needed in the simulation and we can provide support for (A.2) making the connection to RPBM in §A.2.1 by applying asymptotic expansions established for the GI/GI/1 model (and more general multichannel queueing models) in Abate and Whitt (1994); corresponding asymptotic expansions for MAP/GI/1 queues were established in Choudhury and Whitt (1994). From (4) and (18) of Abate and Whitt (1994), we get the following result. As in the main thesis, we fix the service process and introduce the traffic intensity  $\rho$  by scaling time in a rate-1 arrival process. That produces a well-defined model as a function of the traffic intensity, where we only change the arrival rate, which we denote by the subscript  $\rho$ , as in the main thesis.

**Theorem A.2.1.** (asymptotic expansion from Abate and Whitt (1994)) For the GI/GI/1 model, and thus also the periodic  $GI_t/GI/1$  model,

$$\theta_{\rho}^{*} = \frac{2(1-\rho)}{c_{a}^{2}+c_{s}^{2}} + C_{\theta}(1-\rho)^{2} + O((1-\rho^{3})) \quad as \quad \rho \uparrow 1,$$
(A.3)

where  $C_{\theta}$  depends on the first three moments of the mean-1 interarrival time  $U_k$  and service time  $V_k$ , but not  $\rho$ , via

$$C_{\theta} \equiv C_{\theta}(c_a^2, d_a; c_s^2, d_s) \equiv \left(\frac{8(d_s - d_a)}{(c_a^2 + c_s^2)^3} - \frac{2(c_a^2 - c_s^2)}{(c_a^2 + c_s^2)^2}\right),\tag{A.4}$$

with  $d_s \equiv (E[V_k^3] - 3c_s^2(c_s^2 + 1) - 1)/6$ . and similarly for  $d_a$  using the interarrival time.

In §A.2.1, we have suggested that we can calculate the RPBM periodic steady-state tail probabilities  $P(Z_y(\infty; c_x) > b)$  by calculating associated tail probabilities  $P(W_y > b)$  for  $GI_t/GI/1$ queues. Now we show that we may be able to choose two different  $GI_t/GI/1$  queues that will bound the desired RPBM tail probabilities above and below, and thus bound the error. The following result only applies to the rates, but it explains what we have seen in numerical examples; see Table A.2 below and the ratios  $P(W_y > b)/P(W > b)$  in Tables 3.5 and 3.6 in the main thesis.

**Corollary A.2.1.** (switching interarrival-time and service-time distributions) If we switch the interarrival-time and service-time distributions without altering their mean values, and thus switch the pairs  $(c_a^2, d_a)$  and  $(c_s^2, d_s)$ , then  $C_{\theta}$  in (A.4) is unchanged except for its sign, which is reversed. Thus, the one-term asymptotic approximation for  $\theta^*(\rho)$  is bounded above and below by these special two-term approximations.

### A.2.3 Approximations for the Asymptotic Decay Rate

In §A.2.4 we discuss the exact values for the asymptotic decay rates in the special parametric cases in §3.3.2 of the main thesis. For  $M_t/M/1$ ,  $\theta^* \equiv \theta_{\rho}^* = 1 - \rho$ . For both  $M_t/H_2/1$  and  $(H_2)_t/M/1$ ,  $\theta^*$  is obtained as the solution of quadratic equations. Taylor series approximations produce asymptotic expansions that are consistent with (A.3).

Table A.2 compares the 1-term and 2-term approximations for the asymptotic decay rate  $\theta_{\rho}^{*}$  from the asymptotic expansion in (A.3) with the exact values for the  $M_t/H_2/1$  and  $(H_2)_t/M/1$  models, where the  $H_2$  distribution has  $c^2 = 2.0$  and balanced means. The scaled value  $\theta_{\rho}^{*}/(1-\rho)$  is shown for 6 values of  $1-\rho$ . The asymptotic decay rate for RBM and RPBM are obtained directly from the first term. Table A.2 shows that the 2-term approximation can serve as an explicit formula for  $\theta_{\rho}^{*}$  provided that  $\rho$  is not too small.

Assuming appropriate limit interchanges are valid, the asymptotic decay rate for RPBM is the same as for RBM, and that common value can be obtained directly from the first term in (A.3).

Assuming that limits for the steady-state quantities follow from the process limits in the HT FCLT in Theorem 3.6.1 of the main thesis,  $(1-\rho)W_{\rho,y} \Rightarrow Z_y(\infty; c^2)$ , where  $Z_y(\infty; c^2)$  has the steady-state distribution of RPBM. Assuming that the decay rates converge, we should have

$$\theta^* = \lim_{\rho \uparrow 1} \theta^*_{\rho} / (1 - \rho) = 2 / (c_a^2 + c_s^2)$$
(A.5)

from (A.3). For ordinary RBM, this is immediate because RBM has an exponential steady-state distribution. Since the asymptotic decay rate of  $(1 - \rho)W_{\rho,y}$  and  $(1 - \rho)W_{\rho}$  agrees for all  $\rho$ , the same will be true for the limits, provided that the limit interchange is valid.

**Table A.1:** A comparison of the 1-term and 2-term approximations for the asymptotic decay rate  $\theta_{\rho}^{*}$  from the asymptotic expansion in (A.3) with the exact values for the  $M_t/H_2/1$  and  $(H_2)_t/M/1$  models, where the  $H_2$  distribution has  $c^2 = 2.0$  and balanced means: The scaled value  $\theta_{\rho}^{*}/(1-\rho)$  is shown for 6 values of  $1-\rho$ .

	$1 - \rho = 0.16$	$1-\rho=0.08$	$1-\rho=0.04$	$1-\rho=0.02$	$1-\rho=0.01$	$1 - \rho = 0.005$
$M_t/H_2/1$ queue						
exact	0.62934	0.64843	0.65766	0.66219	0.66444	0.66555
first term	0.66667	0.66667	0.66667	0.66667	0.66667	0.66667
first two terms	0.63111	0.64889	0.65778	0.66222	0.66444	0.66556
$(H_2)_t/M/1$ queue						
exact	0.70619	0.68542	0.67580	0.67117	0.66890	0.66778
first term	0.66667	0.66667	0.66667	0.66667	0.66667	0.66667
first two terms	0.70222	0.68444	0.67556	0.67111	0.66889	0.66778

### A.2.4 Exact Values for the Asymptotic Decay Rate

We now give the exact values for the asymptotic decay rates in the special parametric cases considered in §3.3.2 and §3.4.1 of the main thesis. First, for  $M_t/M/1$ ,  $\theta^* \equiv \theta_{\rho}^* = 1 - \rho$ . For both  $M_t/H2/1$  and  $(H_2)_t/M/1$ ,  $\theta^*$  is obtained as the solution of quadratic equations. The other cases are:  $(M + D)_t/M/1$ ,  $M_t/M + D/1$  and  $(M + D)_t/(M + D)/1$ . The final one is important to treat cases with  $c_a^2 + c_s^2 < 1$ . The first two cover  $1 < c_a^2 + c_s^2 < 2$ . We may also want others such as  $(H_2)_t/H_2/1$ .

Table A.2 compares the 1-term and 2-term approximations for the asymptotic decay rate  $\theta_{\rho}^*$ from the asymptotic expansion with the exact values for the  $M_t/H_2/1$  and  $(H_2)_t/M/1$  models, where the  $H_2$  distribution has  $c^2 = 2.0$  and balanced means. The scaled value  $\theta_{\rho}^*/(1-\rho)$  is shown for 6 values of  $1-\rho$ . the asymptotic decay rate for RBM and RPBM are obtained directly from the first term. Table A.2 shows that the 2-term approximation can serve as an explicit formula for  $\theta_{\rho}^*$  provided that  $\rho$  is not too small.

**Table A.2:** A comparison of the 1-term and 2-term approximations for the asymptotic decay rate  $\theta_{\rho}^{*}$  from the asymptotic expansion in (A.3) with the exact values for the  $M_t/H_2/1$  and  $(H_2)_t/M/1$  models, where the  $H_2$  distribution has  $c^2 = 2.0$  and balanced means: The scaled value  $\theta_{\rho}^{*}/(1-\rho)$  is shown for 6 values of  $1-\rho$ .

	$1 - \rho = 0.16$	$1-\rho=0.08$	$1-\rho=0.04$	$1-\rho=0.02$	$1-\rho=0.01$	$1-\rho=0.005$
$M_t/H_2/1$ queue						
$\theta^*$						
exact	0.10069	0.05187	0.02631	0.01324	0.006644	0.003328
first term	0.10667	0.05333	0.02667	0.01333	0.006667	0.003333
first two terms	0.10098	0.05191	0.02631	0.01324	0.006644	0.003328
$(H_2)_t/M/1$ queue						
$\theta^*$						
exact	0.11299	0.05483	0.02703	0.01342	0.006689	0.003339
first term	0.10667	0.05333	0.02667	0.01333	0.006667	0.003333
first two terms	0.11236	0.05476	0.02702	0.01342	0.006689	0.003339
$M_t/H_2/1$ queue						
$\theta^*/(1-\rho)$						
exact	0.62934	0.64843	0.65766	0.66219	0.66444	0.66555
first term	0.66667	0.66667	0.66667	0.66667	0.66667	0.66667
first two terms	0.63111	0.64889	0.65778	0.66222	0.66444	0.66556
$(H_2)_t/M/1$ queue						
$\theta^*/(1-\rho)$						
exact	0.70619	0.68542	0.67580	0.67117	0.66890	0.66778
first term	0.66667	0.66667	0.66667	0.66667	0.66667	0.66667
first two terms	0.70222	0.68444	0.67556	0.67111	0.66889	0.66778

We now discuss the exact values for asymptotic decay rates in the special parametric cases in §3.3.2 of the main thesis. In the  $GI_t/GI/1$  model, let  $\lambda$  be the average arrival rate and  $\mu$  be the service rate, then the optimal  $\theta^*$  is the same as for the GI/GI/1 model with rate- $\lambda$  i.i.d. inter-arrival times  $U_k$  and rate- $\mu$  i.i.d. service times  $V_k$ . First, for  $M_t/M/1$ ,  $\theta^* = \mu - \lambda$  and  $\theta^* \equiv \theta_{\rho}^* = 1 - \rho$  as a function of  $\rho$  if we let  $\mu = 1$ .

For both the  $M_t/H_2/1$  and  $(H_2)_t/M/1$  models,  $\theta^*$  is obtained as the solution of quadratic
equations. In the  $M_t/H_2/1$  model, let  $V_k$  has density  $h(x) = p_1\mu_1 e^{-\mu_1 x} + p_2\mu_2 e^{-\mu_2 x}$ , where  $(p_1/\mu_1) + (p_2/\mu_2) = 1/\mu$ . We solve

$$E[e^{\theta^* V}]E[e^{-\theta^* U}] = 1$$

for  $\theta^*$ , so that

$$E[e^{\theta^*V}]E[e^{-\theta^*U}] = (p_1\mu_1/(\mu_1 - \theta^*) + p_2\mu_2/(\mu_2 - \theta^*))(\lambda/(\lambda + \theta^*)) = 1.$$

This reduces to the quadratic equation

$$(\theta^*)^2 + (\lambda - \mu_1 - \mu_2)\theta^* + (\mu_1\mu_2 - p_2\mu_1\lambda - p_1\mu_2\lambda) = 0$$

or

$$(\theta^*)^2 + (\lambda - \mu_1 - \mu_2)\theta^* + (1 - \rho)\mu_1\mu_2 = 0$$

Hence,

$$\theta^* = \left[(\mu_1 + \mu_2 - \lambda) \pm \sqrt{\lambda^2 - 2(\mu_1 + \mu_2)\lambda + \mu_1^2 + \mu_2^2 + (4\rho - 2)\mu_1\mu_2}\right]/2,$$

where we choose the value that is appropriate, i.e., satisfying  $\mu_1 - \theta^* > 0$ ,  $\mu_2 - \theta^* > 0$ ,  $\lambda + \theta^* > 0$ .

Similarly for the  $(H_2)_t/M/1$  model, let  $U_k$  has density  $g(x) = p_1\lambda_1e^{-\lambda_1x} + p_2\lambda_2e^{-\lambda_2x}$ , with  $(p_1/\lambda_1) + (p_2/\lambda_2) = 1/\lambda$ . Thus, we solve

$$E[e^{\theta^* V}]E[e^{-\theta^* U}] = (\mu/(\mu - \theta^*))(p_1\lambda_1/(\lambda_1 + \theta^*) + p_2\lambda_2/(\lambda_2 + \theta^*)) = 1,$$

which reduces to

$$(\theta^*)^2 + (\lambda_1 + \lambda_2 - \mu)\theta^* + (\lambda_1\lambda_2 - p_2\lambda_1\mu - p_1\lambda_2\mu) = 0$$

or

$$(\theta^*)^2 + (\lambda_1 + \lambda_2 - \mu)\theta^* + \lambda_1\lambda_2(1 - 1/\rho) = 0$$

which has solution

$$\theta^* = \left[ -(\lambda_1 + \lambda_2 - \mu) \pm \sqrt{\mu^2 - 2(\lambda_1 + \lambda_2)\mu + \lambda_1^2 + \lambda_2^2 + (4/\rho - 2)\lambda_1\lambda_2} \right]/2,$$

where we choose the value that is appropriate.

We now briefly discuss other cases, namely,  $(M + D)_t/M/1$ ,  $M_t/M + D/1$  and  $(M + D)_t/(M + D)/1$ . The final one is important to treat cases with  $c_a^2 + c_s^2 < 1$ . The first two cover  $1 < c_a^2 + c_s^2 < 2$ . In all of these cases, we need to solve transcendental equations to get  $\theta^*$ , which is done numerically using Newton's or bisection method. For example, in  $(M+D)_t/M/1$  queue, let  $U_k$  have parameter pair  $(d, \lambda')$  such that

$$\frac{e^{\lambda' d}}{\lambda'} = \frac{1}{\lambda}$$

. We solve

$$E[e^{\theta^*V}]E[e^{-\theta^*U}] = \frac{\mu}{\mu - \theta^*}e^{-\theta^*d}\frac{\lambda'}{\lambda' + \theta^*} = 1,$$

or

$$(\theta^*)^2 - (\mu - \lambda')\theta^* + \mu\lambda'(e^{-\theta^*d} - 1) = 0.$$

We obtain the following proposition when we compare the exact values of  $\theta^*$  with the aymptotic expansion in (A.3).

**Proposition A.2.1.** The exact values of  $\theta^*$  for  $M_t/H_2/1$  and  $(H_2)_t/M/1$  models are consistent with the two-term asymptotic expansion in (A.3).

**Proof.** We only do this proof for  $M_t/H_2/1$  model here; the proof for  $(H_2)_t/M/1$  model is similar. For the  $M_t/H_2/1$  model, the interarrival time is exponential with  $c_a^2 = 1$  and  $E[U_k^3] = 6$ , then the first term in (A.3) becomes  $2(1 - \rho)/(1 + c_2^2)$ . From (A.4), the coefficient of the second term is

$$C = \frac{-4(E[V_k^3] - 3c_s^2(c_s^2 + 1) - E[U_k^3] + 3c_a^2(c_a^2 + 1)) - 6((c_s^2)^2 - (c_a^2)^2)}{3(c_a^2 + c_s^2)^3}$$
  
=  $\frac{-4E[V_k^3] + 6(c_s^2)^2 + 12c_s^2 + 6}{3(1 + c_s^2)^3}$  (A.6)

Without loss of generality, we let  $\mu = 1$  and thus  $\lambda = \rho$ . For  $M_t/H_2/1$  queue,

$$\theta^* = \left[(\mu_1 + \mu_2 - \rho) - \sqrt{\rho^2 - 2(\mu_1 + \mu_2)\rho + \mu_1^2 + \mu_2^2 + (4\rho - 2)\mu_1\mu_2}\right]/2 \tag{A.7}$$

We use a change of variable with  $x = 1 - \rho$  and substitute  $\rho$  with 1 - x in (A.7):

$$\theta^* = \frac{1}{2}(\mu_1 + \mu_2 - 1 + x) - \frac{1}{2}\sqrt{x^2 + (2(\mu_1 + \mu_2) - 2 - 4\mu_1\mu_2)x + (\mu_1 + \mu_2 - 1)^2} \\
\equiv \frac{1}{2}(\mu_1 + \mu_2 - 1 + x) - \frac{1}{2}f(x) \\
= \frac{1}{2}(\mu_1 + \mu_2 - 1 + x) - \frac{1}{2}(f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + O(x^3)),$$
(A.8)

where we define the function f(x) and do taylor series expansion to get the first two terms of f(x).

First, we look at the constant term of  $\theta^*$  in (A.8), it equals

$$\frac{1}{2}(\mu_1 + \mu_2 - 1 - f(0)) = \frac{1}{2}(\mu_1 + \mu_2 - 1 - |\mu_1 + \mu_2 - 1|) = 0$$

Because  $(p_1/\mu_1) + (p_2/\mu_2) = 1$ , we have  $(p_1/\mu_1) < 1$  and  $(p_2/\mu_2) < 1$ . Hence,  $\mu_1 + \mu_2 > p_1 + p_2 = 1$ and  $|\mu_1 + \mu_2 - 1| = \mu_1 + \mu_2 - 1$ . This is consistent with (A.3) which has no constant term.

Second, we consider the first-order term in the Taylor expansion of  $\theta^*$  in (A.8). It equals

$$\frac{1}{2}(1-f'(0)) = \frac{1}{2}(1-\frac{1}{2}f(x)^{-\frac{1}{2}}(2x+2(\mu_1+\mu_2)-2-4\mu_1\mu_2)|_{x=0}) \\
= \frac{1}{2}(1-\frac{(\mu_1+\mu_2)-1-2\mu_1\mu_2}{\mu_1+\mu_2-1}) \\
= \frac{\mu_1\mu_2}{\mu_1+\mu_2-1} \\
= \frac{\mu_1^2\mu_2^2}{(\mu_1+\mu_2-1)\mu_1\mu_2} \\
= \frac{\mu_1^2\mu_2^2}{(\mu_1+\mu_2-1)(p_1\mu_2+p_2\mu_1)} \\
= \frac{\mu_1^2\mu_2^2}{p_1\mu_2^2+p_2\mu_1^2} \\
= \frac{2}{1+c_s^2},$$

where  $p_1\mu_2 + p_2\mu_1 = \mu_1\mu_2$  and  $p_1\mu_2^2 + p_2\mu_1^2 = ((c_s^2 + 1)/2)\mu_1^2\mu_2^2$  follow from the first two moments of  $V_k$ . Hence, we see that this first-order coefficient is consistent with the first term in (A.3).

Finally, we examine the second-order term in the expansion of  $\theta^*$ , which equals

$$\begin{aligned} -\frac{1}{4}f''(0) &= -\frac{1}{4}(-\frac{1}{4}f(x)^{-\frac{3}{2}}(2x+2(\mu_1+\mu_2)-2-4\mu_1\mu_2)^2 + \frac{1}{2}f(x)^{-\frac{1}{2}}2)|_{x=0} \\ &= \frac{((\mu_1+\mu_2)-1-2\mu_1\mu_2)^2}{4(\mu_1+\mu_2-1)^3} - \frac{1}{4(\mu_1+\mu_2-1)} \\ &= \frac{\mu_1^2\mu_2^2 - \mu_1\mu_2(\mu_1+\mu_2-1)}{(\mu_1+\mu_2-1)^3} \\ &= \frac{(p_1\mu_2+p_2\mu_1)^2 - (p_1\mu_2^2+p_2\mu_1^2)}{(\mu_1+\mu_2-1)^3}, \end{aligned}$$

where we used  $\mu_1\mu_2(\mu_1 + \mu_2 - 1) = p_1\mu_2^2 + p_2\mu_1^2$  that is derived in the last paragraph. We have also derived previously that

$$\frac{\mu_1\mu_2}{\mu_1+\mu_2-1} = \frac{2}{1+c_s^2}$$

. Hence, by substituting  $c_s^2$  in (A.6), we can write  $C_\theta$  as

$$C_{\theta} = \frac{-24p_{1}/\mu_{1}^{3} - 24p_{2}/\mu_{2}^{3} + 6(\frac{2(\mu_{1}+\mu_{2}-1)}{\mu_{1}\mu_{2}} - 1)^{2} + 12(\frac{2(\mu_{1}+\mu_{2}-1)}{\mu_{1}\mu_{2}} - 1) + 6}{24\frac{(\mu_{1}+\mu_{2}-1)^{3}}{\mu_{1}^{3}\mu_{2}^{3}}}$$

$$= \frac{-p_{1}\mu_{2}^{3} - p_{2}\mu_{1}^{3} + (\mu_{1} + \mu_{2} - 1)^{2}\mu_{1}\mu_{2}}{(\mu_{1} + \mu_{2} - 1)^{3}}$$

$$= \frac{-p_{1}\mu_{2}^{3} - p_{2}\mu_{1}^{3} + (p_{1}\mu_{2}^{2} + p_{2}\mu_{1}^{2})(\mu_{1} + \mu_{2} - 1)}{(\mu_{1} + \mu_{2} - 1)^{3}}$$

$$= \frac{(p_{1}\mu_{1}\mu_{2}^{2} + p_{2}\mu_{1}^{2}\mu_{2}) - (p_{1}\mu_{2}^{2} + p_{2}\mu_{1}^{2})}{(\mu_{1} + \mu_{2} - 1)^{3}}$$

$$= \frac{(p_{1}\mu_{2} + p_{2}\mu_{1})^{2} - (p_{1}\mu_{2}^{2} + p_{2}\mu_{1}^{2})}{(\mu_{1} + \mu_{2} - 1)^{3}}.$$

Therefore, we conclude that this second-order term coefficient in the exact  $\theta^*$  is consistent with that in (A.3).

As noted in Corollary A.2.1, the two-term approximations for  $\theta^*$  in the  $M_t/H_2/1$  and  $(H_2)_t/M/1$ models approach the one-term approximation in (A.3) from opposite sides.

Table A.2 compares the 1-term and 2-term approximations for the asymptotic decay rate  $\theta_{\rho}^{*}$  from the asymptotic expansion in (A.3) with the exact values for the  $M_t/H_2/1$  and  $(H_2)_t/M/1$  models, where the  $H_2$  distribution has  $c^2 = 2.0$  and balanced means. The scaled value  $\theta_{\rho}^{*}/(1-\rho)$  is shown for 6 values of  $1-\rho$ . the asymptotic decay rate for RBM and RPBM are obtained directly from the first term. Table A.2 shows that the 2-term approximation can serve as an explicit formula for  $\theta_{\rho}^{*}$  provided that  $\rho$  is not too small.

In this specific case, the asymptotic expansion (A.3) for  $\theta^*$  have the following expressions for  $M_t/H_2/1$  and  $(H_2)_t/M/1$  models respectively:

$$M_t/H_2/1: \quad \theta_{\rho}^* = \frac{2}{3}(1-\rho) - \frac{2}{9}(1-\rho)^2 + O(1-\rho)^3;$$
  
$$(H_2)_t/M/1: \quad \theta_{\rho}^* = \frac{2}{3}(1-\rho) + \frac{2}{9}(1-\rho)^2 + O(1-\rho)^3.$$

#### A.2.5 More Bounds

To obtain further bounds, consider the common case in which  $\lambda(t) \geq \overline{\lambda}$ ,  $0 \leq t \leq pC$  while  $\lambda(t) \leq \overline{\lambda}$ ,  $pC \leq t \leq C$ , for some  $p, 0 . Then <math>\tilde{\Lambda}_C(t) = \Lambda(C) - \Lambda(C-t) \leq \overline{\lambda}t$ ,  $0 \leq t \leq C$ , while  $\tilde{\Lambda}_{pC}(t) = \Lambda(pC) - \Lambda(pC-t) \geq \overline{\lambda}t$ ,  $0 \leq t \leq C$ . As a consequence,  $\tilde{\Lambda}_C^{-1}(t) \geq \overline{\lambda}t$ ,  $0 \leq t \leq C$ , while  $\tilde{\Lambda}_{pC}^{-1}(t) \leq \bar{\lambda}t, \ 0 \leq t \leq C.$  Thus,

$$W_0 = W_C \le W \le W_{pC}.\tag{A.9}$$

It is natural to seek conditions under which  $P(W_y > b)$  is increasing in y from a minimum at y = 0to a maximum at y = pC and then is decreasing back to the minimum at y = C.

### A.2.6 Heuristic Approximations

Given Lemmas 3.3.1 and 3.3.2 and Corollary 3.3.4 of the main thesis, we propose the approximation

$$W_y \approx W - \omega_y,$$
 (A.10)

where

$$\omega_y \equiv \frac{-1}{\rho C} \int_0^C (\tilde{\Lambda}_y(s) - \rho s) \, ds. \tag{A.11}$$

For the sinusoidal case, from Corollary 2 of the main thesis, we obtain

$$\omega_y = \frac{\beta \cos\left(\gamma y\right)}{\gamma} = \frac{\zeta_y^+ + \zeta_y^-}{2}.$$
(A.12)

Unfortunately, we find that this approximation is not consistently accurate, but it does help us understand roughly how  $W_y$  depends on the parameters. In our examples, this approximation consistently underestimates the exact values. Intuitively, that makes sense because we expect the extrema to be larger than the time average.

## A.3 Simulation Results

For all experiments we use the sinusoidal arrival-rate function (2.4), where  $\beta$ ,  $0 < \beta < 1$ , is the relative amplitude and the cycle length is  $C = 2\pi/\gamma$ .

In §A.3.1 (Tables A.3-A.16) and §A.3.2 (Tables A.17-A.28) we report results on experiments to estimate the tail probabilities  $P(W_y > b)$  in the Markovian  $M_t/M/1$  model. In §A.3.3 (Tables A.29-A.38) and §A.3.4 (Tables A.39-A.49), respectively, we report results on experiments to estimate the tail probabilities  $P(W_y > b)$  in the  $(H_2)_t/M/1$  and  $M_t/H_2/1$  models. For non-exponential distributions, we use the  $H_2$  distribution (hyperexponential, mixture of two exponential distributions), with with probability density function (pdf)  $f(x) = p_1\mu_1e^{-\mu_1x} + p_2\mu_2e^{-\mu_2x}$ , with  $p_1 + p_2 = 1$ , having parameter triple  $(p_1, \mu_1, \mu_2)$ . To reduce the parameters to two (the mean and scv), we assume balanced means, i.e.,  $p_1/\mu_1 = p_2/\mu_2$ , as in (3.7) of Whitt (1982). In all examples, we let the squared coefficient of variation (scv, variance divided by the mean) be  $c^2 = 2.0$ .

In §A.3.5 we report additional results on experiments to estimate the mean  $E[W_y]$  and standard deviation  $SD[W_y]$  using §3.4.5 of the main thesis. Tables A.50-A.52 report results for the  $M_t/M/1$  model, while Tables A.53 and A.54 report results for the  $(H_2)_t/M/1$  and  $M_t/H_2/1$  models, respectively.

In §A.3.6 we display analogs of Tables 3.6 and 3.10 in the main thesis reporting estimates of tail probabilities for the  $(H_2)_t/M/1$  model, which requires the adjustment involving  $m_{X_1}(\theta^*)$  in (3.28) of the main thesis. That adjustment is required because the first interarrival time has the equilibrium lifetime distribution associated with the  $H_2$  interarrival-time distribution (which is a different  $H_2$  distribution). Tables A.55 and A.56 show the closely related values when that factor is omitted. These tables are closely related because the steady-state workload and waiting time coincide in the heavy-traffic limit. Table 3.1 in the main thesis shows that the steady-state workload and waiting time in the stationary  $H_2/M/1$  model are quite different for the low traffic intensity of  $\rho = 0.1$ .

# A.3.1 Tail Probabilities for the $M_t/M/1$ Periodic Queue with $\bar{\lambda} = 1$

Tables A.3-A.16 display simulation estimates of  $P(W_y > b)$  for the  $M_t/M/1$  model, scaled to have  $\bar{\lambda} = 1$  and  $\mu = 1/\rho$ . in subsequent tables, the scaling was changed to have  $\bar{\lambda} = \rho$  and  $\mu = 1$ , as in the main thesis. An approximation for  $A_y$  is shown; it is discussed in §A.2.6.

Tables A.3-A.7 show estimates for 12 values of *b* ranging from 5 to 90 to show that the simulation accuracy tends to be independent of *b*, as intended for rare-event simulation. To check the simulation algorithm and for a basis of comparison, Table A.3 shows simulation results for the M/M/1 queue, where the exact results are known. Then Tables A.4-A.7 show the estimates for 3 different values of  $\gamma$  in (2.4) ( $\gamma = 10$ ,  $\gamma = 1.0$ , and  $\gamma = 0.1$ ) and 4 different cases of *y*. Here the cycle length is chosen to be  $C = 2\pi/\gamma$ , so the four values of *y* are 0C,  $0.25C = \pi/2\gamma$ ,  $0.50C = \pi/\gamma$  and  $0.75C = 3\pi/2\gamma$ . All these examples have  $\rho = 0.8$  and  $\beta = 0.2$ . We regard this as our base model, and regard  $\gamma = 1.0$ and 0.1 as our base examples illustrating shorter and longer cycles, respectively.

There are 8 columns. The first column gives n, the number of replications. The second column gives the tail probability estimate  $\hat{p} \equiv P(W_y > b) \equiv A_y e - \theta^* b$  and then the third and fourth columns give the components  $\exp(-\theta^* b)$  and  $A \equiv A_y$ . The fifth column gives the standard error (s.e.), while the sixth and seventh columns give the lower bound (lb) and upper bound (ub) of the associated 95% confidence interval (CI). The final eight column gives the relative error (r.e.), which is the estimated s.e divided by the estimated value itself.

Tables A.8-A.16 show the estimates as a function of y for 40 values of y within the cycle in 9 different cases. As noted above, in all these cases  $\bar{\lambda} = 1$  and  $\mu = 1/\rho$ . Tables A.8-A.10 consider three values of the pair  $(\gamma, b)$  for fixed  $(\rho, \beta) = (0.8, 0.2)$ , in particular,  $(\gamma, b) = (10, 20)$ ,  $(\gamma, b) = (0.1, 50)$  and  $(\gamma, b) = (0.01, 300)$ . Tables A.11 and A.12 consider three values of the pair  $(\gamma, b)$  for fixed  $(\rho, \beta) = (0.9, 0.2)$ , in particular,  $(\gamma, b) = (1, 20)$  and  $(\gamma, b) = (0.1, 50)$ . Tables A.13-A.15 consider three values of the pair  $(\gamma, b)$  for fixed  $(\rho, \beta) = (0.9, 0.2)$ , in particular,  $(\gamma, b) = (1, 20)$  and  $(\gamma, b) = (0.1, 50)$ . Tables A.13-A.15 consider three values of the pair  $(\gamma, b)$  for fixed  $(\rho, \beta) = (0.8, 0.5)$ , in particular,  $(\gamma, b) = (10, 20)$ ,  $(\gamma, b) = (1.0, 20)$  and  $(\gamma, b) = (0.1, 100)$ . Finally, Table A.16 shows estimates as a function of y for 40 values of y within a small subinterval in the center of the cycle, in an attempt to verify that the maximum occurs in the middle of the cycle, i.e, at y = 0.5. Table A.16 has the parameter 4-tuple  $(\gamma, \beta\rho, b) = (0.1, 0.2, 0.8, 20)$ .

Tables A.3-A.16 display simulation estimates of  $P(W_y > b)$  for the  $M_t/M/1$  model, scaled to have  $\bar{\lambda} = 1$  and  $\mu = 1/\rho$ . in subsequent tables, the scaling was changed to have  $\bar{\lambda} = \rho$  and  $\mu = 1$ , as in the main thesis.

**Table A.3:** Estimates of  $\hat{p} \equiv P(W > b) \equiv Ae^{-\theta^* b}$  in the M/M/1 model with  $\rho = 0.8, \bar{\lambda} = 1, \mu = 1.25$  based on n = 5000 replications.

$\beta = 0$	b	n	$\hat{p}$	$\exp(-\theta^* b)$	A	s.e.	95% CI (lb)	(ub)	r.e.
	5	5000	0.229	0.287	0.799	6.60E-04	0.228	0.230	0.00289
	10	5000	0.0656	0.0821	0.799	1.90E-04	0.0652	0.0660	0.00289
	15	5000	0.0187	0.0235	0.797	5.48E-05	0.0186	0.0188	0.00293
	20	5000	0.00541	0.00674	0.803	1.52E-05	0.00538	0.00544	0.00280
	25	5000	1.54E-03	1.93E-03	0.797	4.51E-06	0.00153	0.00155	0.00293
	30	5000	4.43E-04	5.53E-04	0.800	1.29E-06	0.000440	0.000445	0.00290
	40	5000	3.64E-05	4.54E-05	0.802	1.05E-07	3.62E-05	3.66E-05	0.00288
	50	5000	2.98E-06	3.73E-06	0.800	8.51 E-09	2.97 E-06	3.00E-06	0.00285
	60	5000	2.45E-07	3.06E-07	0.800	6.97 E- 10	2.43E-07	2.46E-07	0.00285
	70	5000	2.01E-08	2.51E-08	0.802	5.75E-11	2.00E-08	2.02E-08	0.00286
	80	5000	1.65E-09	2.06E-09	0.798	4.73E-12	1.64E-09	1.65E-09	0.00287
	90	5000	1.35E-10	1.69E-10	0.795	4.04E-13	1.34E-10	1.35E-10	0.00300

	b	n	$\hat{p}$	$\exp(-\theta^* b)$	A	s.e.	95% CI (lb)	(ub)	r.e.
$\gamma = 10$	5	5000	0.228	0.287	0.797	6.55 E-04	0.227	0.230	0.00287
	10	5000	0.0654	0.0821	0.797	1.87E-04	0.0651	0.0658	0.00286
	15	5000	0.0188	0.0235	0.799	5.32E-05	0.0187	0.0189	0.00283
	20	5000	0.00537	0.00674	0.797	1.55E-05	0.00534	0.00540	0.00289
	25	5000	1.53E-03	1.93E-03	0.795	4.37E-06	0.00153	0.00154	0.00285
	30	5000	4.40E-04	5.53E-04	0.795	1.28E-06	4.37E-04	4.42E-04	0.00290
	40	5000	3.61E-05	4.54E-05	0.795	1.05E-07	3.59E-05	3.63E-05	0.00290
	50	5000	2.97 E-06	3.73E-06	0.796	8.59E-09	2.95E-06	2.99E-06	0.00289
	60	5000	2.44E-07	3.06E-07	0.798	7.02E-10	2.43E-07	2.45E-07	0.00288
	70	5000	2.01E-08	2.51E-08	0.799	5.67 E- 11	1.99E-08	2.02E-08	0.00283
	80	5000	1.64E-09	2.06E-09	0.796	4.82E-12	1.63E-09	1.65E-09	0.00294
	90	5000	1.35E-10	1.69E-10	0.797	3.88E-13	1.34E-10	1.36E-10	0.00288
$\gamma = 1$	5	5000	0.219	0.287	0.764	6.38E-04	0.218	0.220	0.00292
	10	5000	0.0628	0.0821	0.765	1.87E-04	0.0624	0.0632	0.00298
	15	5000	0.0179	0.0235	0.762	5.19E-05	0.0178	0.0180	0.00290
	20	5000	0.00516	0.00674	0.766	1.51E-05	0.00513	0.00519	0.00292
	25	5000	1.48E-03	1.93E-03	0.764	4.29E-06	0.00147	0.00148	0.00291
	30	5000	4.25E-04	5.53E-04	0.769	1.20E-06	4.23E-04	4.27E-04	0.00283
	40	5000	3.49E-05	4.54E-05	0.769	1.00E-07	3.47E-05	3.51E-05	0.00287
	50	5000	2.85E-06	3.73E-06	0.764	8.40E-09	2.83E-06	2.86E-06	0.00295
	60	5000	2.34E-07	3.06E-07	0.766	6.85E-10	2.33E-07	2.36E-07	0.00292
	70	5000	1.92E-08	2.51E-08	0.763	5.61E-11	1.90E-08	1.93E-08	0.00293
	80	5000	1.58E-09	2.06E-09	0.767	4.65E-12	1.57E-09	1.59E-09	0.00294
	90	5000	1.29E-10	1.69E-10	0.764	3.86E-13	1.28E-10	1.30E-10	0.00299
$\gamma=0.1$	5	5000	0.161	0.287	0.563	8.88E-04	0.160	0.163	0.00550
	10	5000	0.0413	0.0821	0.503	2.33E-04	0.0409	0.0418	0.00565
	15	5000	0.0122	0.0235	0.520	6.77 E-05	0.0121	0.0124	0.00554
	20	5000	0.00360	0.00674	0.535	1.98E-05	0.00356	0.00364	0.00550
	25	5000	1.06E-03	1.93E-03	0.551	5.72E-06	0.00105	0.00107	0.00538
	30	5000	3.04E-04	5.53E-04	0.550	1.66E-06	3.01E-04	3.08E-04	0.00546
	40	5000	2.50E-05	4.54E-05	0.551	1.37E-07	2.47E-05	2.53E-05	0.00548
	50	5000	2.04E-06	3.73E-06	0.547	1.10E-08	2.02E-06	2.06E-06	0.00538
	60	5000	1.67 E-07	3.06E-07	0.546	9.25E-10	1.65E-07	1.69E-07	0.00553
	70	5000	1.37E-08	2.51E-08	0.544	7.59E-11	1.35E-08	1.38E-08	0.00556
	80	5000	1.12E-09	2.06E-09	0.545	6.20E-12	1.11E-09	1.14E-09	0.00552
	90	5000	9.21E-11	1.69E-10	0.544	5.01E-13	9.11E-11	9.31E-11	0.00544

**Table A.4:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model for y = 0.0as a function of  $\gamma$  and b based on n = 5,000 replications:  $\rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.2$ 

	b	n	$\hat{p}$	$\exp(-\theta^* b)$	Α	s.e.	95% CI (lb)	(ub)	r.e.
$\gamma = 10$	5	5000	0.229	0.287	0.801	6.61E-04	0.228	0.231	0.00288
	10	5000	0.0659	0.0821	0.803	1.88E-04	0.0655	0.0663	0.00285
	15	5000	0.0187	0.0235	0.797	5.57 E-05	0.0186	0.0188	0.00297
	20	5000	0.00538	0.00674	0.799	1.59E-05	0.00535	0.00541	0.00296
	25	5000	1.55E-03	1.93E-03	0.801	4.45E-06	0.00154	0.00156	0.00288
	30	5000	4.42E-04	5.53E-04	0.800	1.28E-06	4.40E-04	4.45E-04	0.00288
	40	5000	3.64E-05	4.54E-05	0.802	1.05E-07	3.62E-05	3.66E-05	0.00288
	50	5000	3.00E-06	3.73E-06	0.806	8.51 E-09	2.99E-06	3.02E-06	0.00283
	60	5000	2.44E-07	3.06E-07	0.797	7.13E-10	2.43E-07	2.45E-07	0.00292
	70	5000	2.02E-08	2.51E-08	0.805	5.76E-11	2.01E-08	2.03E-08	0.00285
	80	5000	1.64E-09	2.06E-09	0.798	4.81E-12	1.64E-09	1.65E-09	0.00293
	90	5000	1.35E-10	1.69E-10	0.799	3.84E-13	1.34E-10	1.36E-10	0.00284
$\gamma = 1$	5	5000	0.230	0.287	0.804	6.81E-04	0.229	0.232	0.00295
	10	5000	0.0659	0.0821	0.803	1.92E-04	0.0655	0.0663	0.00292
	15	5000	0.0188	0.0235	0.801	5.67 E-05	0.0187	0.0189	0.00301
	20	5000	0.00540	0.00674	0.801	1.58E-05	0.00536	0.00543	0.00294
	25	5000	1.54E-03	1.93E-03	0.799	4.59E-06	0.00153	0.00155	0.00298
	30	5000	4.45E-04	5.53E-04	0.805	1.28E-06	4.43E-04	4.48E-04	0.00287
	40	5000	3.63E-05	4.54E-05	0.800	1.07E-07	3.61E-05	3.65E-05	0.00294
	50	5000	2.97E-06	3.73E-06	0.798	8.98E-09	2.96E-06	2.99E-06	0.00302
	60	5000	2.46E-07	3.06E-07	0.803	7.18E-10	2.44E-07	2.47E-07	0.00292
	70	5000	2.02E-08	2.51E-08	0.804	5.90E-11	2.01E-08	2.03E-08	0.00293
	80	5000	1.66E-09	2.06E-09	0.806	4.74E-12	1.65E-09	1.67 E-09	0.00285
	90	5000	1.36E-10	1.69E-10	0.804	4.00E-13	1.35E-10	1.37E-10	0.00294
$\gamma=0.1$	5	5000	0.293	0.287	1.024	1.24E-03	0.291	0.296	0.00421
	10	5000	0.0828	0.0821	1.008	4.06E-04	0.0820	0.0836	0.00491
	15	5000	0.0217	0.0235	0.924	1.20E-04	0.0215	0.0220	0.00553
	20	5000	0.00600	0.00674	0.891	3.37E-05	0.00594	0.00607	0.00561
	25	5000	1.71E-03	1.93E-03	0.887	9.53E-06	0.00169	0.00173	0.00556
	30	5000	4.95E-04	5.53E-04	0.895	2.76E-06	4.90E-04	5.00E-04	0.00558
	40	5000	4.13E-05	4.54E-05	0.910	2.23E-07	4.09E-05	4.18E-05	0.00539
	50	5000	3.37E-06	3.73E-06	0.904	1.86E-08	3.33E-06	3.40E-06	0.00551
	60	5000	2.73E-07	3.06E-07	0.893	1.51E-09	2.70E-07	$2.76\mathrm{E}\text{-}07$	0.00554
	70	5000	2.27E-08	2.51E-08	0.902	1.25E-10	2.24E-08	2.29E-08	0.00551
	80	5000	1.85E-09	2.06E-09	0.896	1.01E-11	1.83E-09	1.87 E-09	0.00547
	90	5000	1.52E-10	1.69E-10	0.900	8.36E-13	1.51E-10	1.54E-10	0.00549

**Table A.5:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model for  $y = \pi/2\gamma$  as a function of  $\gamma$  and b based on n = 5,000 replications:  $\rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.2$ 

_	b	n	$\hat{p}$	$\exp(-\theta^* b)$	Α	s.e.	95% CI (lb)	(ub)	r.e.
$\gamma = 10$	5	5000	0.232	0.287	0.808	6.64E-04	0.230	0.233	0.00286
	10	5000	0.0657	0.0821	0.800	1.93E-04	0.0653	0.0661	0.00294
	15	5000	0.0190	0.0235	0.807	5.39E-05	0.0189	0.0191	0.00284
	20	5000	0.00546	0.00674	0.810	1.53E-05	0.00543	0.00549	0.00281
	25	5000	1.55E-03	1.93E-03	0.804	4.49E-06	0.00154	0.00156	0.00289
	30	5000	4.46E-04	5.53E-04	0.807	1.28E-06	4.44E-04	4.49E-04	0.00286
	40	5000	3.64E-05	4.54E-05	0.802	1.06E-07	3.62E-05	3.66E-05	0.00291
	50	5000	3.00E-06	3.73E-06	0.804	8.59E-09	2.98E-06	3.01E-06	0.00286
	60	5000	2.46E-07	3.06E-07	0.803	7.21E-10	2.44E-07	2.47E-07	0.00294
	70	5000	2.02E-08	2.51E-08	0.804	5.76E-11	2.01E-08	2.03E-08	0.00285
	80	5000	1.65E-09	2.06E-09	0.803	4.79E-12	1.65E-09	1.66E-09	0.00289
	90	5000	1.36E-10	1.69E-10	0.805	3.88E-13	1.35E-10	1.37E-10	0.00285
$\gamma = 1$	5	5000	0.242	0.287	0.846	6.96E-04	0.241	0.244	0.00287
	10	5000	0.0691	0.0821	0.842	2.05E-04	0.0687	0.0695	0.00297
	15	5000	0.0198	0.0235	0.841	5.89E-05	0.0197	0.0199	0.00298
	20	5000	0.00570	0.00674	0.846	1.65E-05	0.00567	0.00573	0.00290
	25	5000	1.62E-03	1.93E-03	0.840	4.80E-06	0.00161	0.00163	0.00296
	30	5000	4.68E-04	5.53E-04	0.847	1.36E-06	4.66E-04	4.71E-04	0.00289
	40	5000	3.81E-05	4.54E-05	0.840	1.15E-07	3.79E-05	3.83E-05	0.00303
	50	5000	3.14E-06	3.73E-06	0.843	9.16E-09	3.12E-06	3.16E-06	0.00292
	60	5000	2.59E-07	3.06E-07	0.847	7.62E-10	2.58E-07	2.61E-07	0.00294
	70	5000	2.13E-08	2.51E-08	0.849	6.11E-11	2.12E-08	2.14E-08	0.00287
	80	5000	1.74E-09	2.06E-09	0.842	5.18E-12	1.73E-09	1.75E-09	0.00298
	90	5000	1.42E-10	1.69E-10	0.839	4.25E-13	1.41E-10	1.43E-10	0.00300
$\gamma=0.1$	5	5000	0.342	0.287	1.195	1.99E-03	0.338	0.346	0.00581
	10	5000	0.1181	0.0821	1.438	6.32E-04	0.1168	0.1193	0.00535
	15	5000	0.0366	0.0235	1.554	1.86E-04	0.0362	0.0369	0.00508
	20	5000	0.01038	0.00674	1.541	5.42E-05	0.01027	0.01049	0.00522
	25	5000	2.93E-03	1.93E-03	1.516	1.57 E-05	0.00289	0.00296	0.00536
	30	5000	8.22E-04	5.53E-04	1.485	4.51E-06	8.13E-04	8.30E-04	0.00549
	40	5000	6.67 E-05	4.54E-05	1.470	3.68E-07	6.60E-05	6.75 E- 05	0.00552
	50	5000	5.49E-06	3.73E-06	1.473	3.01E-08	5.43E-06	5.55E-06	0.00548
	60	5000	4.58E-07	3.06E-07	1.499	2.52E-09	4.54E-07	4.63E-07	0.00549
	70	5000	3.75 E-08	2.51E-08	1.495	2.06E-10	3.71E-08	3.79 E-08	0.00549
	80	5000	3.07 E-09	2.06E-09	1.490	1.69E-11	3.04E-09	3.10E-09	0.00552
	90	5000	2.49E-10	1.69E-10	1.474	1.38E-12	2.47E-10	2.52E-10	0.00554

**Table A.6:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model for  $y = \pi/\gamma$  as a function of  $\gamma$  and b based on n = 5,000 replications:  $\rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.2$ 

	b	n	$\hat{p}$	$\exp(-\theta^*b)$	Α	s.e.	95% CI (lb)	(ub)	r.e.
$\gamma = 10$	5	5000	0.229	0.287	0.798	6.66E-04	0.227	0.230	0.00291
	10	5000	0.0657	0.0821	0.801	1.89E-04	0.0654	0.0661	0.00287
	15	5000	0.0187	0.0235	0.794	5.49E-05	0.0186	0.0188	0.00294
	20	5000	0.00541	0.00674	0.803	1.54E-05	0.00538	0.00544	0.00284
	25	5000	1.55E-03	1.93E-03	0.801	4.43E-06	0.00154	0.00155	0.00286
	30	5000	4.43E-04	5.53E-04	0.801	1.28E-06	4.40E-04	4.45E-04	0.00290
	40	5000	3.63E-05	4.54E-05	0.800	1.05E-07	3.61E-05	3.65 E-05	0.00289
	50	5000	2.98E-06	3.73E-06	0.798	8.62 E- 09	2.96E-06	2.99E-06	0.00290
	60	5000	2.46E-07	3.06E-07	0.803	6.95E-10	2.44E-07	2.47 E-07	0.00283
	70	5000	2.01E-08	2.51E-08	0.799	5.81E-11	2.00E-08	2.02E-08	0.00289
	80	5000	1.66E-09	2.06E-09	0.803	4.74E-12	1.65E-09	1.67 E-09	0.00286
	90	5000	1.36E-10	1.69E-10	0.802	3.93E-13	1.35E-10	1.37E-10	0.00290
$\gamma = 1$	5	5000	0.231	0.287	0.807	6.63E-04	0.230	0.232	0.00287
	10	5000	0.0659	0.0821	0.803	1.92E-04	0.0655	0.0663	0.00291
	15	5000	0.0189	0.0235	0.803	5.53E-05	0.0188	0.0190	0.00293
	20	5000	0.00539	0.00674	0.800	1.58E-05	0.00536	0.00542	0.00294
	25	5000	1.55E-03	1.93E-03	0.801	4.60E-06	0.00154	0.00155	0.00298
	30	5000	4.44E-04	5.53E-04	0.803	1.29E-06	4.42E-04	4.47E-04	0.00290
	40	5000	3.66E-05	4.54E-05	0.807	1.06E-07	3.64E-05	3.68E-05	0.00290
	50	5000	2.98E-06	3.73E-06	0.798	8.82E-09	2.96E-06	2.99E-06	0.00296
	60	5000	2.45E-07	3.06E-07	0.800	7.21E-10	2.43E-07	2.46E-07	0.00294
	70	5000	2.01E-08	2.51E-08	0.802	5.91E-11	2.00E-08	2.03E-08	0.00293
	80	5000	1.66E-09	2.06E-09	0.803	4.90E-12	1.65E-09	1.67 E-09	0.00296
	90	5000	1.36E-10	1.69E-10	0.805	4.00E-13	1.35E-10	1.37E-10	0.00293
$\gamma = 0.1$	5	5000	0.201	0.287	0.701	1.14E-03	0.199	0.203	0.00568
	10	5000	0.0658	0.0821	0.801	3.83E-04	0.0650	0.0665	0.00581
	15	5000	0.0205	0.0235	0.872	1.15E-04	0.0203	0.0207	0.00562
	20	5000	0.00612	0.00674	0.908	3.30E-05	0.00605	0.00618	0.00540
	25	5000	1.77E-03	1.93E-03	0.918	9.62E-06	0.00175	0.00179	0.00543
	30	5000	5.01E-04	5.53E-04	0.906	2.72E-06	4.96E-04	5.06E-04	0.00543
	40	5000	4.10E-05	4.54E-05	0.903	$2.27\mathrm{E}\text{-}07$	4.06E-05	4.14E-05	0.00555
	50	5000	3.33E-06	3.73E-06	0.893	1.84E-08	3.29E-06	3.37 E-06	0.00552
	60	5000	2.76E-07	3.06E-07	0.901	1.51E-09	2.73E-07	2.79 E- 07	0.00549
	70	5000	2.28E-08	2.51E-08	0.908	1.24E-10	2.26E-08	2.30E-08	0.00544
	80	5000	1.87 E-09	2.06E-09	0.905	1.02E-11	1.84E-09	1.89E-09	0.00549
	90	5000	1.52E-10	1.69E-10	0.898	8.37E-13	1.50E-10	1.54E-10	0.00551

**Table A.7:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model for  $y = 3\pi/2\gamma$  as a function of  $\gamma$  and b based on n = 5,000 replications:  $\rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.2$ 

$\gamma = 10$	position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	10000	0.0053699	0.00674	0.797	0.796	1.08E-05	0.0053487	0.0053911	0.00202
	0.025	10000	0.0053537	0.00674	0.795	0.796	1.09E-05	0.0053323	0.0053751	0.00204
	0.050	10000	0.0053577	0.00674	0.795	0.796	1.11E-05	0.0053359	0.0053795	0.00208
	0.075	10000	0.0053619	0.00674	0.796	0.796	1.10E-05	0.0053403	0.0053835	0.00206
	0.100	10000	0.0053614	0.00674	0.796	0.797	1.09E-05	0.0053400	0.0053829	0.00204
	0.125	10000	0.0053859	0.00674	0.799	0.797	1.09E-05	0.0053646	0.0054073	0.00202
	0.150	10000	0.0053805	0.00674	0.799	0.798	1.09E-05	0.0053590	0.0054019	0.00203
	0.175	10000	0.0053653	0.00674	0.796	0.798	1.09E-05	0.0053439	0.0053867	0.00204
	0.200	10000	0.0053969	0.00674	0.801	0.799	1.09E-05	0.0053755	0.0054183	0.00202
	0.225	10000	0.0053956	0.00674	0.801	0.799	1.10E-05	0.0053740	0.0054172	0.00204
	0.250	10000	0.0053814	0.00674	0.799	0.800	1.10E-05	0.0053598	0.0054029	0.00204
	0.275	10000	0.0053804	0.00674	0.799	0.801	1.10E-05	0.0053588	0.0054020	0.00205
	0.300	10000	0.0053728	0.00674	0.797	0.801	1.11E-05	0.0053510	0.0053945	0.00207
	0.325	10000	0.0053793	0.00674	0.798	0.802	1.12E-05	0.0053574	0.0054012	0.00208
	0.350	10000	0.0054018	0.00674	0.802	0.802	1.12E-05	0.0053799	0.0054238	0.00207
	0.375	10000	0.0053946	0.00674	0.801	0.803	1.12E-05	0.0053727	0.0054165	0.00207
	0.400	10000	0.0054297	0.00674	0.806	0.803	1.10E-05	0.0054081	0.0054514	0.00203
	0.425	10000	0.0054067	0.00674	0.802	0.804	1.10E-05	0.0053851	0.0054283	0.00204
	0.450	10000	0.0054257	0.00674	0.805	0.804	1.11E-05	0.0054040	0.0054474	0.00204
	0.475	10000	0.0054453	0.00674	0.808	0.804	1.09E-05	0.0054238	0.0054667	0.00201
	0.500	10000	0.0054138	0.00674	0.803	0.804	1.11E-05	0.0053920	0.0054356	0.00206
	0.525	10000	0.0054315	0.00674	0.806	0.804	1.10E-05	0.0054099	0.0054532	0.00203
	0.550	10000	0.0054065	0.00674	0.802	0.804	1.12E-05	0.0053846	0.0054284	0.00206
	0.575	10000	0.0054207	0.00674	0.805	0.804	1.11E-05	0.0053990	0.0054425	0.00205
	0.600	10000	0.0054270	0.00674	0.805	0.803	1.09E-05	0.0054057	0.0054484	0.00201
	0.625	10000	0.0054153	0.00674	0.804	0.803	1.09E-05	0.0053938	0.0054367	0.00202
	0.650	10000	0.0054065	0.00674	0.802	0.802	1.10E-05	0.0053849	0.0054281	0.00204
	0.675	10000	0.0054121	0.00674	0.803	0.802	1.09E-05	0.0053908	0.0054334	0.00201
	0.700	10000	0.0054175	0.00674	0.804	0.801	1.10E-05	0.0053960	0.0054390	0.00202
	0.725	10000	0.0053797	0.00674	0.798	0.801	1.11E-05	0.0053580	0.0054014	0.00206
	0.750	10000	0.0053901	0.00674	0.800	0.800	1.10E-05	0.0053686	0.0054116	0.00203
	0.775	10000	0.0053580	0.00674	0.795	0.799	1.11E-05	0.0053361	0.0053798	0.00208
	0.800	10000	0.0053783	0.00674	0.798	0.799	1.10E-05	0.0053568	0.0053998	0.00204
	0.825	10000	0.0053843	0.00674	0.799	0.798	1.08E-05	0.0053630	0.0054056	0.00201
	0.850	10000	0.0053946	0.00674	0.801	0.798	1.09E-05	0.0053733	0.0054160	0.00202
	0.875	10000	0.0053783	0.00674	0.798	0.797	1.09E-05	0.0053569	0.0053997	0.00203
	0.900	10000	0.0053758	0.00674	0.798	0.797	1.10E-05	0.0053543	0.0053974	0.00205
	0.925	10000	0.0053714	0.00674	0.797	0.796	1.08E-05	0.0053502	0.0053926	0.00201
	0.950	10000	0.0053435	0.00674	0.793	0.796	1.10E-05	0.0053220	0.0053651	0.00206
	0.975	10000	0.0053681	0.00674	0.797	0.796	1.09E-05	0.0053468	0.0053895	0.00203

**Table A.8:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on n = 5,000 replications:  $\gamma = 10, b = 20, \rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.2$ 

$\gamma = 0.1$	position	n	$\hat{p}$	$\exp(-\theta^*b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	10000	2.04E-06	3.73E-06	0.548	0.485	0.294	0.800	7.95E-09	2.03E-06	2.06E-06	0.00389
	0.025	10000	2.05E-06	3.73E-06	0.551	0.488	0.296	0.805	7.94E-09	2.04E-06	2.07E-06	0.00387
	0.050	10000	2.09E-06	3.73E-06	0.560	0.497	0.302	0.820	8.08E-09	2.07 E-06	2.10E-06	0.00387
	0.075	10000	2.15E-06	3.73E-06	0.577	0.512	0.311	0.845	8.34E-09	2.13E-06	2.17E-06	0.00388
	0.100	10000	2.24E-06	3.73E-06	0.602	0.534	0.324	0.880	8.68E-09	2.22E-06	2.26E-06	0.00387
	0.125	10000	2.37E-06	3.73E-06	0.635	0.562	0.341	0.926	9.21E-09	2.35E-06	2.38E-06	0.00389
	0.150	10000	2.50E-06	3.73E-06	0.671	0.596	0.362	0.983	9.70E-09	2.48E-06	2.52E-06	0.00388
	0.175	10000	2.68E-06	3.73E-06	0.719	0.638	0.387	1.051	1.05E-08	2.66E-06	2.70E-06	0.00392
	0.200	10000	2.87E-06	3.73E-06	0.770	0.685	0.416	1.130	1.12E-08	2.85 E-06	2.89E-06	0.00392
	0.225	10000	3.10E-06	3.73E-06	0.832	0.740	0.449	1.220	1.20E-08	3.08E-06	3.12E-06	0.00387
	0.250	10000	3.36E-06	3.73E-06	0.902	0.800	0.485	1.319	1.31E-08	3.33E-06	3.39E-06	0.00390
	0.275	10000	3.63E-06	3.73E-06	0.974	0.865	0.525	1.426	1.42E-08	3.60E-06	3.66E-06	0.00390
	0.300	10000	3.90E-06	3.73E-06	1.045	0.934	0.566	1.539	1.51E-08	3.87E-06	3.93E-06	0.00389
	0.325	10000	4.19E-06	3.73E-06	1.126	1.004	0.609	1.655	1.63E-08	4.16E-06	4.23E-06	0.00389
	0.350	10000	4.50E-06	3.73E-06	1.208	1.073	0.651	1.770	1.76E-08	4.47E-06	4.54E-06	0.00391
	0.375	10000	4.80E-06	3.73E-06	1.289	1.139	0.691	1.878	1.84E-08	4.77 E-06	4.84E-06	0.00383
	0.400	10000	5.04E-06	3.73E-06	1.352	1.199	0.727	1.977	1.96E-08	5.00E-06	5.08E-06	0.00389
	0.425	10000	5.27E-06	3.73E-06	1.413	1.249	0.758	2.059	2.03E-08	5.23E-06	5.31E-06	0.00386
	0.450	10000	5.39E-06	3.73E-06	1.446	1.287	0.781	2.122	2.09E-08	5.35E-06	5.43E-06	0.00388
	0.475	10000	5.54E-06	3.73E-06	1.487	1.311	0.795	2.161	2.14E-08	5.50 E-06	5.58E-06	0.00387
	0.500	10000	5.54E-06	3.73E-06	1.488	1.319	0.800	2.175	2.15E-08	5.50 E-06	5.59E-06	0.00388
	0.525	10000	5.51E-06	3.73E-06	1.479	1.311	0.795	2.161	2.14E-08	5.47E-06	5.55 E-06	0.00388
	0.550	10000	5.46E-06	3.73E-06	1.466	1.287	0.781	2.122	2.11E-08	5.42E-06	5.50E-06	0.00386
	0.575	10000	5.22E-06	3.73E-06	1.401	1.249	0.758	2.059	2.04E-08	5.18E-06	5.26E-06	0.00390
	0.600	10000	5.04E-06	3.73E-06	1.353	1.199	0.727	1.977	1.95E-08	5.00E-06	5.08E-06	0.00387
	0.625	10000	4.77E-06	3.73E-06	1.281	1.139	0.691	1.878	1.87E-08	4.74 E-06	4.81E-06	0.00391
	0.650	10000	4.55E-06	3.73E-06	1.221	1.073	0.651	1.770	1.75E-08	4.52E-06	4.58E-06	0.00386
	0.675	10000	4.23E-06	3.73E-06	1.134	1.004	0.609	1.655	1.64E-08	4.19E-06	4.26E-06	0.00388
	0.700	10000	3.91E-06	3.73E-06	1.049	0.934	0.566	1.539	1.52E-08	3.88E-06	3.94E-06	0.00389
	0.725	10000	3.63E-06	3.73E-06	0.975	0.865	0.525	1.426	1.40E-08	3.61E-06	3.66E-06	0.00385
	0.750	10000	3.35E-06	3.73E-06	0.899	0.800	0.485	1.319	1.30E-08	3.33E-06	3.38E-06	0.00387
	0.775	10000	3.09E-06	3.73E-06	0.829	0.740	0.449	1.220	1.21E-08	3.07 E-06	3.11E-06	0.00390
	0.800	10000	2.86E-06	3.73E-06	0.768	0.685	0.416	1.130	1.12E-08	2.84E-06	2.88E-06	0.00392
	0.825	10000	2.68E-06	3.73E-06	0.719	0.638	0.387	1.051	1.04E-08	2.66 E- 06	2.70E-06	0.00388
	0.850	10000	2.49E-06	3.73E-06	0.669	0.596	0.362	0.983	9.65 E-09	2.47 E-06	2.51E-06	0.00387
	0.875	10000	2.35E-06	3.73E-06	0.630	0.562	0.341	0.926	9.06E-09	2.33E-06	2.37E-06	0.00386
	0.900	10000	2.24E-06	3.73E-06	0.602	0.534	0.324	0.880	8.72E-09	2.23E-06	2.26E-06	0.00389
	0.925	10000	2.16E-06	3.73E-06	0.578	0.512	0.311	0.845	8.38E-09	2.14E-06	2.17E-06	0.00389
	0.950	10000	2.08E-06	3.73E-06	0.557	0.497	0.302	0.820	8.08E-09	2.06E-06	2.09E-06	0.00389
	0.975	10000	2.04E-06	3.73E-06	0.548	0.488	0.296	0.805	7.99E-09	2.03E-06	2.06E-06	0.00391

**Table A.9:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on n = 5,000 replications:  $\gamma = 0.1, b = 50, \rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.2$ 

$\gamma = 0.01$	position	n	$\hat{p}$	$\exp(-\theta^*b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	10000	4.71E-34	2.68E-33	0.176	0.005	0.00004	0.800	7.10E-36	4.57E-34	4.85E-34	0.0151
	0.025	10000	5.02E-34	2.68E-33	0.187	0.006	0.00004	0.851	7.56E-36	4.87E-34	5.17E-34	0.0151
	0.050	10000	5.78E-34	2.68E-33	0.216	0.007	0.00005	1.022	8.95E-36	5.60E-34	5.95E-34	0.0155
	0.075	10000	7.64E-34	2.68E-33	0.285	0.009	0.00006	1.380	1.19E-35	7.41E-34	7.88E-34	0.0155
	0.100	10000	1.07E-33	2.68E-33	0.401	0.014	0.00009	2.079	1.75E-35	1.04E-33	1.11E-33	0.0163
	0.125	10000	1.72E-33	2.68E-33	0.642	0.023	0.00016	3.460	2.88E-35	1.66E-33	1.78E-33	0.0167
	0.150	10000	2.86E-33	2.68E-33	1.066	0.042	0.00029	6.284	4.98E-35	2.76E-33	2.95E-33	0.0175
	0.175	10000	5.36E-33	2.68E-33	2.003	0.083	0.00056	12.267	9.71E-35	5.17E-33	5.56E-33	0.0181
	0.200	10000	1.07E-32	2.68E-33	3.994	0.171	0.00115	25.324	1.98E-34	1.03E-32	1.11E-32	0.0185
	0.225	10000	2.30E-32	2.68E-33	8.587	0.366	0.00247	54.309	4.24E-34	2.22E-32	2.38E-32	0.0184
	0.250	10000	4.95E-32	2.68E-33	18.490	0.800	0.00539	118.731	9.29E-34	4.77E-32	5.13E-32	0.0188
	0.275	10000	1.12E-31	2.68E-33	41.970	1.749	0.01178	259.571	2.08E-33	1.08E-31	1.16E-31	0.0185
	0.300	10000	2.60E-31	2.68E-33	96.956	3.751	0.02527	556.653	4.65E-33	2.51E-31	2.69E-31	0.0179
	0.325	10000	5.72E-31	2.68E-33	213.720	7.743	0.05217	1149.186	9.85E-33	5.53E-31	5.92E-31	0.0172
	0.350	10000	1.21E-30	2.68E-33	453.072	15.116	0.10185	2243.478	1.98E-32	1.17E-30	1.25E-30	0.0163
	0.375	10000	2.35E-30	2.68E-33	875.663	27.451	0.18496	4074.040	3.72E-32	2.27 E-30	2.42E-30	0.0158
	0.400	10000	4.22E-30	2.68E-33	1574.049	45.693	0.30788	6781.417	6.33E-32	4.09E-30	4.34E-30	0.0150
	0.425	10000	6.54E-30	2.68E-33	2440.946	68.847	0.46389	10217.825	9.60E-32	6.35E-30	6.73E-30	0.0147
	0.450	10000	9.11E-30	2.68E-33	3399.220	92.957	0.62634	13796.070	1.30E-31	8.85E-30	9.36E-30	0.0143
	0.475	10000	1.13E-29	2.68E-33	4227.878	111.642	0.75224	16569.156	1.58E-31	1.10E-29	1.16E-29	0.0139
	0.500	10000	1.21E-29	2.68E-33	4505.760	118.731	0.80000	17621.173	1.67 E- 31	1.17E-29	1.24E-29	0.0138
	0.525	10000	1.13E-29	2.68E-33	4211.232	111.642	0.75224	16569.156	1.56E-31	1.10E-29	1.16E-29	0.0138
	0.550	10000	9.29E-30	2.68E-33	3469.383	92.957	0.62634	13796.070	1.29E-31	9.04E-30	9.55E-30	0.0138
	0.575	10000	6.88E-30	2.68E-33	2567.108	68.847	0.46389	10217.825	9.59E-32	6.69E-30	7.06E-30	0.0139
	0.600	10000	4.67E-30	2.68E-33	1744.227	45.693	0.30788	6781.417	6.36E-32	4.55 E-30	4.80E-30	0.0136
	0.625	10000	2.74E-30	2.68E-33	1023.988	27.451	0.18496	4074.040	3.78E-32	$2.67 \text{E}{-}30$	2.82E-30	0.0138
	0.650	10000	1.48E-30	2.68E-33	553.534	15.116	0.10185	2243.478	2.07E-32	1.44E-30	1.52E-30	0.0140
	0.675	10000	7.60E-31	2.68E-33	283.764	7.743	0.05217	1149.186	1.07E-32	7.39E-31	7.81E-31	0.0140
	0.700	10000	3.72E-31	2.68E-33	138.823	3.751	0.02527	556.653	5.19E-33	3.62E-31	3.82E-31	0.0140
	0.725	10000	1.78E-31	2.68E-33	66.314	1.749	0.01178	259.571	2.44E-33	1.73E-31	1.82E-31	0.0137
	0.750	10000	7.88E-32	2.68E-33	29.402	0.800	0.00539	118.731	1.10E-33	7.66E-32	8.09E-32	0.0140
	0.775	10000	3.58E-32	2.68E-33	13.368	0.366	0.00247	54.309	4.98E-34	3.48E-32	3.68E-32	0.0139
	0.800	10000	1.61E-32	2.68E-33	6.025	0.171	0.00115	25.324	2.32E-34	1.57E-32	1.66E-32	0.0144
	0.825	10000	7.97E-33	2.68E-33	2.977	0.083	0.00056	12.267	1.13E-34	7.75E-33	8.19E-33	0.0142
	0.850	10000	3.99E-33	2.68E-33	1.488	0.042	0.00029	6.284	5.76E-35	3.87E-33	4.10E-33	0.0144
	0.875	10000	2.20E-33	2.68E-33	0.820	0.023	0.00016	3.460	3.14E-35	2.14E-33	2.26E-33	0.0143
	0.900	10000	1.30E-33	2.68E-33	0.487	0.014	0.00009	2.079	1.89E-35	1.27E-33	1.34E-33	0.0145
	0.925	10000	8.79E-34	2.68E-33	0.328	0.009	0.00006	1.380	1.26E-35	8.54E-34	9.04E-34	0.0143
	0.950	10000	6.41E-34	2.68E-33	0.239	0.007	0.00005	1.022	9.26E-36	6.23E-34	6.59E-34	0.0145
	0.975	10000	5.19E-34	2.68E-33	0.194	0.006	0.00004	0.851	7.67E-36	5.04E-34	5.34E-34	0.0148

**Table A.10:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on n = 5,000 replications:  $\gamma = 0.01, b = 300, \rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.2$ 

$\gamma = 1$	position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ UB	$A_y LB$	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	10000	0.0954	0.108	0.881	0.880	0.861	0.900	9.60E-05	0.0953	0.0956	0.00101
	0.025	10000	0.0956	0.108	0.883	0.880	0.861	0.900	9.63E-05	0.0954	0.0958	0.00101
	0.050	10000	0.0957	0.108	0.883	0.881	0.862	0.901	9.73E-05	0.0955	0.0959	0.00102
	0.075	10000	0.0956	0.108	0.883	0.882	0.863	0.902	9.72 E- 05	0.0955	0.0958	0.00102
	0.100	10000	0.0961	0.108	0.887	0.884	0.865	0.904	9.68E-05	0.0959	0.0963	0.00101
	0.125	10000	0.0961	0.108	0.886	0.886	0.866	0.906	9.82E-05	0.0959	0.0963	0.00102
	0.150	10000	0.0963	0.108	0.889	0.888	0.869	0.908	$9.80\mathrm{E}\text{-}05$	0.0961	0.0965	0.00102
	0.175	10000	0.0968	0.108	0.893	0.891	0.871	0.911	9.79E-05	0.0966	0.0970	0.00101
	0.200	10000	0.0970	0.108	0.895	0.894	0.874	0.914	9.86E-05	0.0968	0.0972	0.00102
	0.225	10000	0.0973	0.108	0.898	0.897	0.877	0.917	9.84E-05	0.0972	0.0975	0.00101
	0.250	10000	0.0976	0.108	0.901	0.900	0.880	0.920	9.81E- $05$	0.0974	0.0978	0.00100
	0.275	10000	0.0980	0.108	0.904	0.903	0.883	0.923	1.00E-04	0.0978	0.0982	0.00102
	0.300	10000	0.0984	0.108	0.908	0.906	0.886	0.927	9.93E-05	0.0982	0.0986	0.00101
	0.325	10000	0.0985	0.108	0.909	0.909	0.889	0.930	1.01E-04	0.0983	0.0987	0.00103
	0.350	10000	0.0992	0.108	0.915	0.912	0.892	0.932	1.00E-04	0.0990	0.0994	0.00101
	0.375	10000	0.0993	0.108	0.916	0.914	0.894	0.935	9.93E-05	0.0991	0.0994	0.00100
	0.400	10000	0.0994	0.108	0.917	0.916	0.896	0.937	1.02E-04	0.0992	0.0996	0.00102
	0.425	10000	0.0997	0.108	0.920	0.918	0.898	0.939	1.02E-04	0.0995	0.0999	0.00102
	0.450	10000	0.0997	0.108	0.920	0.919	0.899	0.940	1.02E-04	0.0995	0.0999	0.00102
	0.475	10000	0.0998	0.108	0.921	0.920	0.900	0.941	1.01E-04	0.0996	0.1000	0.00102
	0.500	10000	0.0998	0.108	0.921	0.920	0.900	0.941	1.02E-04	0.0996	0.1000	0.00102
	0.525	10000	0.0997	0.108	0.920	0.920	0.900	0.941	1.02E-04	0.0995	0.0999	0.00102
	0.550	10000	0.0998	0.108	0.921	0.919	0.899	0.940	1.02E-04	0.0996	0.1000	0.00102
	0.575	10000	0.0997	0.108	0.920	0.918	0.898	0.939	1.01E-04	0.0995	0.0998	0.00101
	0.600	10000	0.0993	0.108	0.917	0.916	0.896	0.937	1.01E-04	0.0991	0.0995	0.00101
	0.625	10000	0.0993	0.108	0.916	0.914	0.894	0.935	1.02E-04	0.0991	0.0995	0.00103
	0.650	10000	0.0987	0.108	0.911	0.912	0.892	0.932	1.03E-04	0.0985	0.0989	0.00104
	0.675	10000	0.0988	0.108	0.912	0.909	0.889	0.930	9.90 E- 05	0.0986	0.0990	0.00100
	0.700	10000	0.0984	0.108	0.908	0.906	0.886	0.927	9.98E-05	0.0982	0.0986	0.00101
	0.725	10000	0.0979	0.108	0.904	0.903	0.883	0.923	1.00E-04	0.0978	0.0981	0.00102
	0.750	10000	0.0978	0.108	0.902	0.900	0.880	0.920	9.81E-05	0.0976	0.0980	0.00100
	0.775	10000	0.0974	0.108	0.899	0.897	0.877	0.917	9.85E-05	0.0972	0.0976	0.00101
	0.800	10000	0.0972	0.108	0.897	0.894	0.874	0.914	9.73E-05	0.0970	0.0973	0.00100
	0.825	10000	0.0964	0.108	0.890	0.891	0.871	0.911	1.02E-04	0.0962	0.0966	0.00106
	0.850	10000	0.0963	0.108	0.889	0.888	0.869	0.908	9.84E-05	0.0961	0.0965	0.00102
	0.875	10000	0.0961	0.108	0.887	0.886	0.866	0.906	9.93E-05	0.0960	0.0963	0.00103
	0.900	10000	0.0963	0.108	0.888	0.884	0.865	0.904	9.43E-05	0.0961	0.0964	0.00098
	0.925	10000	0.0958	0.108	0.884	0.882	0.863	0.902	9.82E-05	0.0956	0.0960	0.00103
	0.950	10000	0.0956	0.108	0.882	0.881	0.862	0.901	$9.70\mathrm{E}\text{-}05$	0.0954	0.0958	0.00101
	0.975	10000	0.0957	0.108	0.883	0.880	0.861	0.900	9.63E-05	0.0955	0.0959	0.00101

**Table A.11:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on n = 5,000 replications:  $\gamma = 1, b = 20, \rho = 0.9, \bar{\lambda} = 1, \mu = 1.11, \beta = 0.2$ 

$\gamma = 0.1$	position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ UB	$A_y LB$	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	10000	0.00292	0.00387	0.757	0.721	0.577	0.900	5.22E-06	0.00291	0.00293	0.00178
	0.025	10000	0.00293	0.00387	0.759	0.723	0.579	0.902	5.23E-06	0.00292	0.00294	0.00178
	0.050	10000	0.00296	0.00387	0.766	0.729	0.583	0.910	5.32E-06	0.00295	0.00297	0.00180
	0.075	10000	0.00300	0.00387	0.776	0.738	0.591	0.922	5.32E-06	0.00299	0.00301	0.00177
	0.100	10000	0.00306	0.00387	0.792	0.752	0.602	0.939	5.42E-06	0.00305	0.00307	0.00177
	0.125	10000	0.00312	0.00387	0.808	0.769	0.616	0.961	5.56E-06	0.00311	0.00314	0.00178
	0.150	10000	0.00321	0.00387	0.829	0.790	0.632	0.986	5.75E-06	0.00319	0.00322	0.00179
	0.175	10000	0.00331	0.00387	0.857	0.814	0.652	1.016	5.86E-06	0.00330	0.00332	0.00177
	0.200	10000	0.00342	0.00387	0.884	0.840	0.673	1.049	6.07E-06	0.00340	0.00343	0.00178
	0.225	10000	0.00353	0.00387	0.914	0.869	0.696	1.086	6.30E-06	0.00352	0.00355	0.00178
	0.250	10000	0.00365	0.00387	0.944	0.900	0.721	1.124	6.55E-06	0.00364	0.00366	0.00180
	0.275	10000	0.00378	0.00387	0.978	0.932	0.746	1.164	6.74E-06	0.00377	0.00379	0.00178
	0.300	10000	0.00392	0.00387	1.014	0.964	0.772	1.204	7.02E-06	0.00391	0.00394	0.00179
	0.325	10000	0.00404	0.00387	1.045	0.996	0.797	1.243	7.20E-06	0.00403	0.00405	0.00178
	0.350	10000	0.00417	0.00387	1.078	1.026	0.821	1.281	7.45E-06	0.00415	0.00418	0.00179
	0.375	10000	0.00428	0.00387	1.108	1.053	0.843	1.315	7.71E-06	0.00427	0.00430	0.00180
	0.400	10000	0.00438	0.00387	1.132	1.077	0.863	1.345	7.86E-06	0.00436	0.00439	0.00180
	0.425	10000	0.00445	0.00387	1.152	1.097	0.878	1.370	7.89E-06	0.00444	0.00447	0.00177
	0.450	10000	0.00452	0.00387	1.168	1.112	0.890	1.388	8.10E-06	0.00450	0.00453	0.00179
	0.475	10000	0.00454	0.00387	1.174	1.121	0.898	1.400	8.11E-06	0.00452	0.00455	0.00179
	0.500	10000	0.00456	0.00387	1.179	1.124	0.900	1.404	8.24E-06	0.00454	0.00457	0.00181
	0.525	10000	0.00455	0.00387	1.177	1.121	0.898	1.400	8.09E-06	0.00453	0.00457	0.00178
	0.550	10000	0.00452	0.00387	1.170	1.112	0.890	1.388	8.01E-06	0.00451	0.00454	0.00177
	0.575	10000	0.00446	0.00387	1.153	1.097	0.878	1.370	7.94E-06	0.00444	0.00447	0.00178
	0.600	10000	0.00437	0.00387	1.131	1.077	0.863	1.345	7.79E-06	0.00436	0.00439	0.00178
	0.625	10000	0.00427	0.00387	1.106	1.053	0.843	1.315	7.65E-06	0.00426	0.00429	0.00179
	0.650	10000	0.00416	0.00387	1.077	1.026	0.821	1.281	7.43E-06	0.00415	0.00418	0.00179
	0.675	10000	0.00405	0.00387	1.047	0.996	0.797	1.243	7.16E-06	0.00403	0.00406	0.00177
	0.700	10000	0.00391	0.00387	1.013	0.964	0.772	1.204	7.05E-06	0.00390	0.00393	0.00180
	0.725	10000	0.00378	0.00387	0.977	0.932	0.746	1.164	6.78E-06	0.00376	0.00379	0.00179
	0.750	10000	0.00366	0.00387	0.946	0.900	0.721	1.124	6.51E-06	0.00365	0.00367	0.00178
	0.775	10000	0.00353	0.00387	0.914	0.869	0.696	1.086	6.27E-06	0.00352	0.00355	0.00177
	0.800	10000	0.00341	0.00387	0.882	0.840	0.673	1.049	6.16E-06	0.00340	0.00342	0.00181
	0.825	10000	0.00329	0.00387	0.850	0.814	0.652	1.016	5.94E-06	0.00328	0.00330	0.00181
	0.850	10000	0.00320	0.00387	0.829	0.790	0.632	0.986	5.75E-06	0.00319	0.00321	0.00180
	0.875	10000	0.00312	0.00387	0.806	0.769	0.616	0.961	5.52E-06	0.00311	0.00313	0.00177
	0.900	10000	0.00305	0.00387	0.789	0.752	0.602	0.939	5.43E-06	0.00304	0.00306	0.00178
	0.925	10000	0.00300	0.00387	0.776	0.738	0.591	0.922	5.33E-06	0.00299	0.00301	0.00178
	0.950	10000	0.00295	0.00387	0.764	0.729	0.583	0.910	5.28E-06	0.00294	0.00296	0.00179
	0.975	10000	0.00294	0.00387	0.760	0.723	0.579	0.902	5.23E-06	0.00293	0.00295	0.00178

**Table A.12:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on n = 5,000 replications:  $\gamma = 0.1, b = 50, \rho = 0.9, \bar{\lambda} = 1, \mu = 1.11, \beta = 0.2$ 

$\gamma = 10$	position	n	$\hat{p}$	$\exp(-\theta^* b)$	Α	$A_y$ approx	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	10000	0.005312	0.00674	0.788	0.790	1.10E-05	0.005290	0.005333	0.00206
	0.025	10000	0.005315	0.00674	0.789	0.790	1.09E-05	0.005294	0.005337	0.00204
	0.050	10000	0.005323	0.00674	0.790	0.791	1.08E-05	0.005302	0.005345	0.00203
	0.075	10000	0.005347	0.00674	0.794	0.791	1.08E-05	0.005326	0.005369	0.00202
	0.100	10000	0.005318	0.00674	0.789	0.792	1.10E-05	0.005297	0.005340	0.00208
	0.125	10000	0.005334	0.00674	0.792	0.793	1.10E-05	0.005312	0.005355	0.00206
	0.150	10000	0.005356	0.00674	0.795	0.794	1.10E-05	0.005334	0.005377	0.00205
	0.175	10000	0.005373	0.00674	0.797	0.795	1.09E-05	0.005351	0.005394	0.00203
	0.200	10000	0.005383	0.00674	0.799	0.797	1.09E-05	0.005362	0.005405	0.00203
	0.225	10000	0.005387	0.00674	0.799	0.798	1.10E-05	0.005365	0.005408	0.00205
	0.250	10000	0.005409	0.00674	0.803	0.800	1.09E-05	0.005388	0.005430	0.00201
	0.275	10000	0.005417	0.00674	0.804	0.802	1.11E-05	0.005396	0.005439	0.00204
	0.300	10000	0.005408	0.00674	0.803	0.803	1.10E-05	0.005386	0.005429	0.00204
	0.325	10000	0.005427	0.00674	0.805	0.805	1.09E-05	0.005405	0.005448	0.00200
	0.350	10000	0.005432	0.00674	0.806	0.806	1.10E-05	0.005410	0.005453	0.00202
	0.375	10000	0.005449	0.00674	0.809	0.807	1.12E-05	0.005427	0.005471	0.00205
	0.400	10000	0.005437	0.00674	0.807	0.808	1.12E-05	0.005415	0.005459	0.00206
	0.425	10000	0.005467	0.00674	0.811	0.809	1.10E-05	0.005445	0.005489	0.00202
	0.450	10000	0.005453	0.00674	0.809	0.810	1.12E-05	0.005431	0.005475	0.00206
	0.475	10000	0.005462	0.00674	0.811	0.810	1.11E-05	0.005440	0.005484	0.00204
	0.500	10000	0.005451	0.00674	0.809	0.810	1.11E-05	0.005429	0.005472	0.00203
	0.525	10000	0.005440	0.00674	0.807	0.810	1.11E-05	0.005418	0.005462	0.00205
	0.550	10000	0.005443	0.00674	0.808	0.810	1.13E-05	0.005421	0.005465	0.00208
	0.575	10000	0.005475	0.00674	0.813	0.809	1.09E-05	0.005454	0.005497	0.00200
	0.600	10000	0.005445	0.00674	0.808	0.808	1.12E-05	0.005423	0.005467	0.00205
	0.625	10000	0.005434	0.00674	0.806	0.807	1.12E-05	0.005412	0.005456	0.00206
	0.650	10000	0.005440	0.00674	0.807	0.806	1.10E-05	0.005418	0.005462	0.00203
	0.675	10000	0.005424	0.00674	0.805	0.805	1.12E-05	0.005402	0.005446	0.00206
	0.700	10000	0.005400	0.00674	0.801	0.803	1.12E-05	0.005378	0.005422	0.00207
	0.725	10000	0.005408	0.00674	0.803	0.802	1.11E-05	0.005386	0.005430	0.00205
	0.750	10000	0.005375	0.00674	0.798	0.800	1.11E-05	0.005353	0.005397	0.00207
	0.775	10000	0.005374	0.00674	0.798	0.798	1.10E-05	0.005352	0.005396	0.00205
	0.800	10000	0.005404	0.00674	0.802	0.797	1.08E-05	0.005383	0.005426	0.00200
	0.825	10000	0.005361	0.00674	0.796	0.795	1.09E-05	0.005340	0.005383	0.00203
	0.850	10000	0.005351	0.00674	0.794	0.794	1.10E-05	0.005329	0.005372	0.00205
	0.875	10000	0.005358	0.00674	0.795	0.793	1.08E-05	0.005337	0.005380	0.00201
	0.900	10000	0.005349	0.00674	0.794	0.792	1.08E-05	0.005328	0.005371	0.00202
	0.925	10000	0.005346	0.00674	0.793	0.791	1.08E-05	0.005325	0.005367	0.00203
	0.950	10000	0.005323	0.00674	0.790	0.791	1.08E-05	0.005302	0.005344	0.00203
	0.975	10000	0.005319	0.00674	0.789	0.790	1.10E-05	0.005297	0.005340	0.00206

**Table A.13:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on n = 5,000 replications:  $\gamma = 10, b = 20, \rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.5$ 

$\gamma = 1$	position	n	$\hat{p}$	$\exp(-\theta^*b)$	A	$A_y$ approx	$A_y$ UB	$A_y LB$	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	10000	0.00486	0.00674	0.721	0.706	0.623	0.800	1.07E-05	0.00484	0.00488	0.00219
	0.025	10000	0.00485	0.00674	0.720	0.707	0.624	0.801	1.09E-05	0.00483	0.00487	0.00224
	0.050	10000	0.00490	0.00674	0.727	0.710	0.627	0.805	1.08E-05	0.00488	0.00492	0.00220
	0.075	10000	0.00492	0.00674	0.731	0.716	0.632	0.811	1.09E-05	0.00490	0.00495	0.00221
	0.100	10000	0.00497	0.00674	0.738	0.723	0.638	0.819	1.11E-05	0.00495	0.00500	0.00222
	0.125	10000	0.00505	0.00674	0.750	0.732	0.646	0.830	1.11E-05	0.00503	0.00507	0.00220
	0.150	10000	0.00512	0.00674	0.759	0.743	0.656	0.842	1.15E-05	0.00509	0.00514	0.00224
	0.175	10000	0.00521	0.00674	0.773	0.756	0.667	0.857	1.16E-05	0.00518	0.00523	0.00223
	0.200	10000	0.00528	0.00674	0.784	0.770	0.679	0.872	1.17E-05	0.00526	0.00531	0.00222
	0.225	10000	0.00540	0.00674	0.801	0.785	0.692	0.889	1.19E-05	0.00537	0.00542	0.00221
	0.250	10000	0.00550	0.00674	0.816	0.800	0.706	0.907	1.22E-05	0.00547	0.00552	0.00221
	0.275	10000	0.00563	0.00674	0.835	0.816	0.720	0.924	1.25E-05	0.00560	0.00565	0.00222
	0.300	10000	0.00576	0.00674	0.854	0.832	0.734	0.942	1.25E-05	0.00573	0.00578	0.00218
	0.325	10000	0.00583	0.00674	0.865	0.847	0.747	0.959	1.28E-05	0.00580	0.00585	0.00220
	0.350	10000	0.00593	0.00674	0.880	0.861	0.760	0.976	1.29E-05	0.00590	0.00595	0.00218
	0.375	10000	0.00601	0.00674	0.892	0.874	0.771	0.990	1.34E-05	0.00598	0.00603	0.00222
	0.400	10000	0.00605	0.00674	0.898	0.885	0.781	1.003	1.36E-05	0.00602	0.00608	0.00225
	0.425	10000	0.00617	0.00674	0.916	0.894	0.789	1.013	1.35E-05	0.00615	0.00620	0.00219
	0.450	10000	0.00618	0.00674	0.917	0.901	0.795	1.021	1.40E-05	0.00615	0.00621	0.00226
	0.475	10000	0.00624	0.00674	0.926	0.905	0.799	1.026	1.37 E-05	0.00621	0.00626	0.00220
	0.500	10000	0.00621	0.00674	0.922	0.907	0.800	1.027	1.38E-05	0.00619	0.00624	0.00222
	0.525	10000	0.00624	0.00674	0.927	0.905	0.799	1.026	1.37 E-05	0.00622	0.00627	0.00219
	0.550	10000	0.00620	0.00674	0.920	0.901	0.795	1.021	1.38E-05	0.00617	0.00623	0.00222
	0.575	10000	0.00615	0.00674	0.912	0.894	0.789	1.013	1.39E-05	0.00612	0.00618	0.00225
	0.600	10000	0.00609	0.00674	0.904	0.885	0.781	1.003	1.35E-05	0.00606	0.00611	0.00222
	0.625	10000	0.00600	0.00674	0.890	0.874	0.771	0.990	1.35E-05	0.00597	0.00602	0.00224
	0.650	10000	0.00593	0.00674	0.881	0.861	0.760	0.976	1.31E-05	0.00591	0.00596	0.00220
	0.675	10000	0.00582	0.00674	0.864	0.847	0.747	0.959	1.30E-05	0.00579	0.00585	0.00223
	0.700	10000	0.00571	0.00674	0.847	0.832	0.734	0.942	1.27E-05	0.00568	0.00573	0.00223
	0.725	10000	0.00562	0.00674	0.834	0.816	0.720	0.924	1.25E-05	0.00559	0.00564	0.00222
	0.750	10000	0.00551	0.00674	0.818	0.800	0.706	0.907	1.23E-05	0.00548	0.00553	0.00223
	0.775	10000	0.00541	0.00674	0.804	0.785	0.692	0.889	1.19E-05	0.00539	0.00544	0.00220
	0.800	10000	0.00527	0.00674	0.782	0.770	0.679	0.872	1.18E-05	0.00525	0.00529	0.00224
	0.825	10000	0.00520	0.00674	0.772	0.756	0.667	0.857	1.16E-05	0.00518	0.00523	0.00223
	0.850	10000	0.00510	0.00674	0.757	0.743	0.656	0.842	1.14E-05	0.00508	0.00513	0.00223
	0.875	10000	0.00505	0.00674	0.749	0.732	0.646	0.830	1.12E-05	0.00503	0.00507	0.00222
	0.900	10000	0.00497	0.00674	0.738	0.723	0.638	0.819	1.09E-05	0.00495	0.00500	0.00219
	0.925	10000	0.00493	0.00674	0.732	0.716	0.632	0.811	1.10E-05	0.00491	0.00495	0.00223
	0.950	10000	0.00487	0.00674	0.723	0.710	0.627	0.805	1.09E-05	0.00485	0.00489	0.00224
	0.975	10000	0.00488	0.00674	0.724	0.707	0.624	0.801	1.08E-05	0.00485749	0.004899702	0.00221

**Table A.14:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on n = 5,000 replications:  $\gamma = 1, b = 20, \rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.5$ 

**Table A.15:** Summary of simulation results for a fixed b and differing y in a cycle:  $\gamma = 0.1, b = 100, \rho = 0.8, \overline{\lambda} = 1, \mu = 1.25, \beta = 0.5$ 

$\gamma = 0.1$	position	n	$\hat{p}$	$\exp(-\theta^* b)$	Α	$A_y$ approx	$A_y$ UB	$A_y LB$	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	10000	5.89E-12	1.39E-11	0.424	0.229	0.066	0.800	3.74E-14	5.82E-12	5.96E-12	0.00635
	0.025	10000	5.99E-12	1.39E-11	0.431	0.233	0.067	0.812	3.75E-14	5.92E-12	6.06E-12	0.00627
	0.050	10000	6.26E-12	1.39E-11	0.451	0.244	0.070	0.850	3.93E-14	6.18E-12	6.34E-12	0.00628
	0.075	10000	6.71E-12	1.39E-11	0.483	0.263	0.075	0.917	4.24E-14	6.63E-12	6.80E-12	0.00632
	0.100	10000	7.44E-12	1.39E-11	0.536	0.291	0.083	1.016	4.72E-14	7.35E-12	7.53E-12	0.00634
	0.125	10000	8.50E-12	1.39E-11	0.612	0.331	0.095	1.154	5.34E-14	8.39E-12	8.60E-12	0.00629
	0.150	10000	9.87E-12	1.39E-11	0.711	0.384	0.110	1.339	6.20E-14	9.75E-12	9.99E-12	0.00628
	0.175	10000	1.16E-11	1.39E-11	0.832	0.454	0.130	1.583	7.29E-14	1.14E-11	1.17E-11	0.00631
	0.200	10000	1.39E-11	1.39E-11	1.004	0.544	0.156	1.898	8.76E-14	1.38E-11	1.41E-11	0.00628
	0.225	10000	1.69E-11	1.39E-11	1.217	0.658	0.188	2.296	1.06E-13	1.67E-11	1.71E-11	0.00628
	0.250	10000	2.05E-11	1.39E-11	1.474	0.800	0.229	2.792	1.29E-13	2.02E-11	2.07E-11	0.00628
	0.275	10000	2.49E-11	1.39E-11	1.796	0.973	0.279	3.395	1.59E-13	2.46E-11	2.52E-11	0.00636
	0.300	10000	3.04E-11	1.39E-11	2.185	1.177	0.337	4.109	1.91E-13	3.00E-11	3.07E-11	0.00630
	0.325	10000	3.61E-11	1.39E-11	2.599	1.411	0.404	4.925	2.27E-13	3.56E-11	3.65E-11	0.00630
	0.350	10000	4.33E-11	1.39E-11	3.117	1.668	0.478	5.822	2.72E-13	4.28E-11	4.38E-11	0.00628
	0.375	10000	4.96E-11	1.39E-11	3.568	1.936	0.555	6.758	3.12E-13	4.89E-11	5.02E-11	0.00629
	0.400	10000	5.61E-11	1.39E-11	4.040	2.199	0.630	7.676	3.56E-13	5.54E-11	5.68E-11	0.00634
	0.425	10000	6.24E-11	1.39E-11	4.492	2.437	0.698	8.505	3.94E-13	6.16E-11	6.32E-11	0.00631
	0.450	10000	6.78E-11	1.39E-11	4.879	2.627	0.753	9.168	4.27E-13	6.69E-11	6.86E-11	0.00630
	0.475	10000	7.05E-11	1.39E-11	5.076	2.750	0.788	9.597	4.43E-13	6.96E-11	7.14E-11	0.00629
	0.500	10000	7.11E-11	1.39E-11	5.122	2.792	0.800	9.746	4.49E-13	7.02E-11	7.20E-11	0.00631
	0.525	10000	7.05E-11	1.39E-11	5.076	2.750	0.788	9.597	4.46E-13	6.96E-11	7.14E-11	0.00633
	0.550	10000	6.80E-11	1.39E-11	4.895	2.627	0.753	9.168	4.26E-13	6.72E-11	6.88E-11	0.00627
	0.575	10000	6.24E-11	1.39E-11	4.492	2.437	0.698	8.505	3.96E-13	6.16E-11	6.32E-11	0.00635
	0.600	10000	5.59E-11	1.39E-11	4.022	2.199	0.630	7.676	3.54E-13	5.52E-11	5.66E-11	0.00633
	0.625	10000	4.98E-11	1.39E-11	3.588	1.936	0.555	6.758	3.15E-13	4.92E-11	5.04E-11	0.00631
	0.650	10000	4.27E-11	1.39E-11	3.078	1.668	0.478	5.822	2.70E-13	4.22E-11	4.33E-11	0.00631
	0.675	10000	3.64E-11	1.39E-11	2.624	1.411	0.404	4.925	2.29E-13	3.60E-11	3.69E-11	0.00629
	0.700	10000	3.03E-11	1.39E-11	2.180	1.177	0.337	4.109	1.91E-13	2.99E-11	3.06E-11	0.00630
	0.725	10000	2.47E-11	1.39E-11	1.782	0.973	0.279	3.395	1.58E-13	2.44E-11	2.51E-11	0.00638
	0.750	10000	2.05E-11	1.39E-11	1.476	0.800	0.229	2.792	1.30E-13	2.02E-11	2.08E-11	0.00635
	0.775	10000	1.69E-11	1.39E-11	1.218	0.658	0.188	2.296	1.05E-13	1.67E-11	1.71E-11	0.00623
	0.800	10000	1.39E-11	1.39E-11	1.000	0.544	0.156	1.898	8.80E-14	1.37E-11	1.41E-11	0.00634
	0.825	10000	1.16E-11	1.39E-11	0.835	0.454	0.130	1.583	7.35E-14	1.15E-11	1.17E-11	0.00633
	0.850	10000	9.80E-12	1.39E-11	0.706	0.384	0.110	1.339	6.25E-14	9.68E-12	9.93E-12	0.00638
	0.875	10000	8.56E-12	1.39E-11	0.617	0.331	0.095	1.154	5.39E-14	8.46E-12	8.67E-12	0.00630
	0.900	10000	7.43E-12	1.39E-11	0.535	0.291	0.083	1.016	4.68E-14	7.34E-12	7.52E-12	0.00630
	0.925	10000	6.77E-12	1.39E-11	0.487	0.263	0.075	0.917	4.24E-14	6.68E-12	6.85E-12	0.00626
	0.950	10000	6.18E-12	1.39E-11	0.445	0.244	0.070	0.850	3.96E-14	6.10E-12	6.25E-12	0.00641
	0.975	10000	6.03E-12	1.39E-11	0.434	0.233	0.067	0.812	3.80E-14	5.96E-12	6.11E-12	0.00629

\_

**Table A.16:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y, for y in the small interval [0.45, 0.55] based on n = 5,000 replications:  $\gamma = 0.1, b =$  $20, \rho = 0.8, \bar{\lambda} = 1, \mu = 1.25, \beta = 0.2$ 

$\gamma = 0.1$	position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	s.e.	95% CI (lb)	(ub)	r.e.
	0.4500	10000	0.0102042	0.00674	1.514	3.81E-05	0.0101296	0.0102789	0.00373
	0.4525	10000	0.0102480	0.00674	1.521	3.85E-05	0.0101726	0.0103234	0.00375
	0.4550	10000	0.0102694	0.00674	1.524	3.83E-05	0.0101944	0.0103445	0.00373
	0.4575	10000	0.0102476	0.00674	1.521	3.84E-05	0.0101723	0.0103229	0.00375
	0.4600	10000	0.0103224	0.00674	1.532	3.86E-05	0.0102467	0.0103981	0.00374
	0.4625	10000	0.0101856	0.00674	1.512	3.85E-05	0.0101101	0.0102611	0.00378
	0.4650	10000	0.0103051	0.00674	1.529	3.87E-05	0.0102292	0.0103810	0.00376
	0.4675	10000	0.0102580	0.00674	1.522	3.84E-05	0.0101826	0.0103333	0.00375
	0.4700	10000	0.0103188	0.00674	1.531	3.83E-05	0.0102438	0.0103938	0.00371
	0.4725	10000	0.0103469	0.00674	1.536	3.87E-05	0.0102711	0.0104227	0.00374
	0.4750	10000	0.0102930	0.00674	1.528	3.83E-05	0.0102179	0.0103681	0.00372
	0.4775	10000	0.0103730	0.00674	1.539	3.90E-05	0.0102966	0.0104495	0.00376
	0.4800	10000	0.0103778	0.00674	1.540	3.82E-05	0.0103029	0.0104528	0.00369
	0.4825	10000	0.0103410	0.00674	1.535	3.88 E-05	0.0102649	0.0104172	0.00376
	0.4850	10000	0.0103687	0.00674	1.539	3.88 E-05	0.0102926	0.0104448	0.00374
	0.4875	10000	0.0104335	0.00674	1.548	3.88 E-05	0.0103574	0.0105097	0.00372
	0.4900	10000	0.0103590	0.00674	1.537	3.86E-05	0.0102833	0.0104346	0.00373
	0.4925	10000	0.0103960	0.00674	1.543	3.89E-05	0.0103197	0.0104723	0.00374
	0.4950	10000	0.0103041	0.00674	1.529	3.87 E-05	0.0102282	0.0103800	0.00376
	0.4975	10000	0.0104239	0.00674	1.547	3.92E-05	0.0103472	0.0105007	0.00376
	0.5000	10000	0.0104064	0.00674	1.544	3.89E-05	0.0103300	0.0104827	0.00374
	0.5025	10000	0.0103887	0.00674	1.542	3.88E-05	0.0103125	0.0104648	0.00374
	0.5050	10000	0.0104046	0.00674	1.544	3.90E-05	0.0103281	0.0104811	0.00375
	0.5075	10000	0.0103907	0.00674	1.542	3.89E-05	0.0103144	0.0104670	0.00375
	0.5100	10000	0.0103596	0.00674	1.538	3.88E-05	0.0102835	0.0104357	0.00375
	0.5125	10000	0.0103260	0.00674	1.533	3.83E-05	0.0102509	0.0104010	0.00371
	0.5150	10000	0.0104469	0.00674	1.550	3.87 E-05	0.0103711	0.0105226	0.00370
	0.5175	10000	0.0103561	0.00674	1.537	3.85E-05	0.0102806	0.0104316	0.00372
	0.5200	10000	0.0104290	0.00674	1.548	3.86E-05	0.0103534	0.0105047	0.00370
	0.5225	10000	0.0103480	0.00674	1.536	3.84E-05	0.0102727	0.0104232	0.00371
	0.5250	10000	0.0103970	0.00674	1.543	3.84E-05	0.0103218	0.0104723	0.00369
	0.5275	10000	0.0102753	0.00674	1.525	3.83E-05	0.0102004	0.0103503	0.00372
	0.5300	10000	0.0102461	0.00674	1.521	3.86E-05	0.0101706	0.0103217	0.00376
	0.5325	10000	0.0102789	0.00674	1.526	3.82E-05	0.0102039	0.0103538	0.00372
	0.5350	10000	0.0102817	0.00674	1.526	3.82E-05	0.0102067	0.0103566	0.00372
	0.5375	10000	0.0102432	0.00674	1.520	3.80E-05	0.0101688	0.0103176	0.00371
	0.5400	10000	0.0102212	0.00674	1.517	3.81E-05	0.0101465	0.0102958	0.00373
	0.5425	10000	0.0102356	0.00674	1.519	3.78 E- 05	0.0101615	0.0103097	0.00369
	0.5450	10000	0.0102132	0.00674	1.516	3.81E-05	0.0101386	0.0102878	0.00373
	0.5475	10000	0.0101162	0.00674	1.501	3.77 E-05	0.0100424	0.0101901	0.00372
	0.5500	10000	0.0101235	0.00674	1.502	3.78E-05	0.0100495	0.0101976	0.00373

### A.3.2 Tail Probabilities for the $M_t/M/1$ Periodic Queue with $\mu = 1$

Table A.17 displays simulation results for what we regard as our base case, having parameter 4tuple  $\rho, \beta, \gamma, b$  = (0.8, 0.2, 0.1, 20), which corresponds to the general framework in (3.39) of the main thesis, i.e.,

$$(\bar{\lambda}_{\rho}, \beta_{\rho}, \gamma_{\rho}, b_{\rho}) = (\rho, (1-\rho)\beta, (1-\rho)^2\gamma, (1-\rho)^{-1}b),$$
 (A.13)

where  $(\beta, \gamma, b)$  is a feasible base triple of positive constants with  $\beta < 1$  when the base triple is  $(\beta, \gamma, b) = (1, 2.5, 4.0)$  as in (3.40) of the main thesis. The bounds for  $A_y$  are discussed in Corollary 4 of the main thesis, while the approximation is discussed at the end here in §A.2.6.

Tables A.18-A.25 give results for the framework in (3.39) for the base triple  $(\beta, \gamma, b) = (1, 25, 4.0)$ . The results for different  $\rho$  ranging from  $\rho = 0.84$  to  $\rho = 0.99$  are summarized in Tables A.26, A.27 and A.28. These summaries strongly supports the heavy-traffic scaling in (3.39).

$\gamma = 1$	position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	5000	1.05E-02	1.83E-02	0.571	0.536	0.359	0.800	5.00E-05	1.04E-02	1.06E-02	4.78E-03
	0.025	5000	1.05E-02	1.83E-02	0.572	0.539	0.361	0.804	5.06E-05	1.04E-02	1.06E-02	4.83E-03
	0.050	5000	1.06E-02	1.83E-02	0.580	0.547	0.367	0.816	5.08E-05	1.05E-02	1.07E-02	4.79E-03
	0.075	5000	1.09E-02	1.83E-02	0.593	0.560	0.375	0.836	5.27 E-05	1.08E-02	1.10E-02	4.85E-03
	0.100	5000	1.13E-02	1.83E-02	0.616	0.579	0.388	0.864	5.47 E-05	1.12E-02	1.14E-02	4.85E-03
	0.125	5000	1.17E-02	1.83E-02	0.639	0.603	0.404	0.899	5.62 E- 05	1.16E-02	1.18E-02	4.80E-03
	0.150	5000	1.24E-02	1.83E-02	0.678	0.632	0.424	0.943	5.99E-05	1.23E-02	1.25E-02	4.82E-03
	0.175	5000	1.30E-02	1.83E-02	0.711	0.667	0.447	0.995	6.29 E- 05	1.29E-02	1.31E-02	4.83E-03
	0.200	5000	1.40E-02	1.83E-02	0.762	0.707	0.474	1.055	6.79 E- 05	1.38E-02	1.41E-02	4.87E-03
	0.225	5000	1.48E-02	1.83E-02	0.806	0.751	0.504	1.121	7.21E-05	1.46E-02	1.49E-02	4.88E-03
	0.250	5000	1.60E-02	1.83E-02	0.873	0.800	0.536	1.193	7.67 E-05	1.58E-02	1.61E-02	4.79E-03
	0.275	5000	1.72E-02	1.83E-02	0.938	0.852	0.571	1.271	8.32E-05	1.70E-02	1.73E-02	4.84E-03
	0.300	5000	1.83E-02	1.83E-02	1.001	0.905	0.607	1.350	8.74E-05	1.82E-02	1.85E-02	4.77 E-03
	0.325	5000	1.95E-02	1.83E-02	1.067	0.959	0.643	1.431	9.28E-05	1.94E-02	1.97E-02	4.75E-03
	0.350	5000	2.10E-02	1.83E-02	1.146	1.012	0.678	1.510	9.69E-05	2.08E-02	2.12E-02	4.62E-03
	0.375	5000	2.17E-02	1.83E-02	1.184	1.062	0.712	1.584	1.01E-04	2.15E-02	2.19E-02	4.68E-03
	0.400	5000	2.27E-02	1.83E-02	1.238	1.106	0.741	1.649	1.05E-04	2.25E-02	2.29E-02	4.65 E- 03
	0.425	5000	2.35E-02	1.83E-02	1.285	1.143	0.766	1.704	1.10E-04	2.33E-02	2.38E-02	4.66E-03
	0.450	5000	2.42E-02	1.83E-02	1.323	1.170	0.784	1.746	1.12E-04	2.40E-02	2.45E-02	4.61E-03
	0.475	5000	2.45E-02	1.83E-02	1.337	1.188	0.796	1.772	1.13E-04	2.43E-02	2.47E-02	4.61E-03
	0.500	5000	2.47E-02	1.83E-02	1.350	1.193	0.800	1.780	1.13E-04	2.45E-02	2.49E-02	4.56E-03
	0.525	5000	2.43E-02	1.83E-02	1.326	1.188	0.796	1.772	1.12E-04	2.41E-02	2.45E-02	4.62E-03
	0.550	5000	2.40E-02	1.83E-02	1.309	1.170	0.784	1.746	1.10E-04	2.38E-02	2.42E-02	4.58E-03
	0.575	5000	2.34E-02	1.83E-02	1.278	1.143	0.766	1.704	1.08E-04	2.32E-02	2.36E-02	4.63E-03
	0.600	5000	2.26E-02	1.83E-02	1.234	1.106	0.741	1.649	1.04E-04	2.24E-02	2.28E-02	4.61E-03
	0.625	5000	2.15E-02	1.83E-02	1.174	1.062	0.712	1.584	1.01E-04	2.13E-02	2.17E-02	4.68E-03
	0.650	5000	2.04E-02	1.83E-02	1.116	1.012	0.678	1.510	9.51E-05	2.02E-02	2.06E-02	4.66E-03
	0.675	5000	1.94E-02	1.83E-02	1.061	0.959	0.643	1.431	8.93E-05	1.93E-02	1.96E-02	4.60E-03
	0.700	5000	1.81E-02	1.83E-02	0.988	0.905	0.607	1.350	$8.47\mathrm{E}\text{-}05$	1.79E-02	1.83E-02	4.68E-03
	0.725	5000	1.71E-02	1.83E-02	0.934	0.852	0.571	1.271	8.01E-05	1.69E-02	1.73E-02	4.68E-03
	0.750	5000	1.60E-02	1.83E-02	0.873	0.800	0.536	1.193	7.54E-05	1.58E-02	1.61E-02	4.72E-03
	0.775	5000	1.50E-02	1.83E-02	0.817	0.751	0.504	1.121	7.14E-05	1.48E-02	1.51E-02	4.77 E-03
	0.800	5000	1.40E-02	1.83E-02	0.764	0.707	0.474	1.055	6.71 E- 05	1.39E-02	1.41E-02	4.79E-03
	0.825	5000	1.31E-02	1.83E-02	0.718	0.667	0.447	0.995	6.22E-05	1.30E-02	1.33E-02	4.73E-03
	0.850	5000	1.25E-02	1.83E-02	0.683	0.632	0.424	0.943	6.00E-05	1.24E-02	1.26E-02	4.80E-03
	0.875	5000	1.19E-02	1.83E-02	0.652	0.603	0.404	0.899	5.69E-05	1.18E-02	1.21E-02	4.77 E-03
	0.900	5000	1.15E-02	1.83E-02	0.625	0.579	0.388	0.864	5.48E-05	1.13E-02	1.16E-02	4.79E-03
	0.925	5000	1.10E-02	1.83E-02	0.601	0.560	0.375	0.836	5.31E-05	1.09E-02	1.11E-02	4.82E-03
	0.950	5000	1.07E-02	1.83E-02	0.586	0.547	0.367	0.816	5.17E-05	1.06E-02	1.08E-02	4.81E-03
	0.975	5000	1.05E-02	1.83E-02	0.575	0.539	0.361	0.804	5.11E-05	1.04E-02	1.06E-02	4.86E-03

**Table A.17:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.1, b = 20, \rho = 0.8, \bar{\lambda} = 0.8, \mu = 1, \beta = 0.2$ 

$\gamma = 1$	position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	5000	0.014162	0.0183	0.773	0.769	0.738	0.800	4.07E-05	0.01408	0.01424	0.00288
	0.025	5000	0.014104	0.0183	0.770	0.769	0.739	0.800	4.12E-05	0.01402	0.01419	0.00292
	0.050	5000	0.014038	0.0183	0.766	0.770	0.740	0.802	4.22E-05	0.01396	0.01412	0.00301
	0.075	5000	0.014227	0.0183	0.777	0.772	0.742	0.803	4.04E-05	0.01415	0.01431	0.00284
	0.100	5000	0.014197	0.0183	0.775	0.775	0.744	0.806	4.11E-05	0.01412	0.01428	0.00289
	0.125	5000	0.014289	0.0183	0.780	0.778	0.747	0.809	4.11E-05	0.01421	0.01437	0.00287
	0.150	5000	0.014311	0.0183	0.781	0.781	0.751	0.813	4.21E-05	0.01423	0.01439	0.00294
	0.175	5000	0.014465	0.0183	0.790	0.786	0.755	0.818	4.18E-05	0.01438	0.01455	0.00289
	0.200	5000	0.014520	0.0183	0.793	0.790	0.759	0.822	4.21E-05	0.01444	0.01460	0.00290
	0.225	5000	0.014620	0.0183	0.798	0.795	0.764	0.827	4.24E-05	0.01454	0.01470	0.00290
	0.250	5000	0.014725	0.0183	0.804	0.800	0.769	0.833	4.22E-05	0.01464	0.01481	0.00286
	0.275	5000	0.014810	0.0183	0.809	0.805	0.773	0.838	4.28E-05	0.01473	0.01489	0.00289
	0.300	5000	0.014879	0.0183	0.812	0.810	0.778	0.843	4.36E-05	0.01479	0.01496	0.00293
	0.325	5000	0.014961	0.0183	0.817	0.815	0.783	0.848	4.35E-05	0.01488	0.01505	0.00291
	0.350	5000	0.015099	0.0183	0.824	0.819	0.787	0.852	4.36E-05	0.01501	0.01518	0.00289
	0.375	5000	0.015093	0.0183	0.824	0.823	0.791	0.857	4.43E-05	0.01501	0.01518	0.00293
	0.400	5000	0.015156	0.0183	0.827	0.826	0.794	0.860	4.44E-05	0.01507	0.01524	0.00293
	0.425	5000	0.015162	0.0183	0.828	0.829	0.797	0.863	4.45E-05	0.01508	0.01525	0.00293
	0.450	5000	0.015274	0.0183	0.834	0.831	0.798	0.865	4.46E-05	0.01519	0.01536	0.00292
	0.475	5000	0.015280	0.0183	0.834	0.832	0.800	0.866	4.39E-05	0.01519	0.01537	0.00287
	0.500	5000	0.015332	0.0183	0.837	0.833	0.800	0.867	4.43E-05	0.01524	0.01542	0.00289
	0.525	5000	0.015291	0.0183	0.835	0.832	0.800	0.866	4.48E-05	0.01520	0.01538	0.00293
	0.550	5000	0.015307	0.0183	0.836	0.831	0.798	0.865	4.42E-05	0.01522	0.01539	0.00289
	0.575	5000	0.015218	0.0183	0.831	0.829	0.797	0.863	4.41E-05	0.01513	0.01530	0.00290
	0.600	5000	0.015178	0.0183	0.829	0.826	0.794	0.860	4.32E-05	0.01509	0.01526	0.00284
	0.625	5000	0.015175	0.0183	0.829	0.823	0.791	0.857	4.30E-05	0.01509	0.01526	0.00283
	0.650	5000	0.015123	0.0183	0.826	0.819	0.787	0.852	4.30E-05	0.01504	0.01521	0.00284
	0.675	5000	0.015090	0.0183	0.824	0.815	0.783	0.848	4.29E-05	0.01501	0.01517	0.00284
	0.700	5000	0.014905	0.0183	0.814	0.810	0.778	0.843	4.28E-05	0.01482	0.01499	0.00287
	0.725	5000	0.014770	0.0183	0.806	0.805	0.773	0.838	4.31E-05	0.01469	0.01485	0.00292
	0.750	5000	0.014647	0.0183	0.800	0.800	0.769	0.833	4.30E-05	0.01456	0.01473	0.00294
	0.775	5000	0.014614	0.0183	0.798	0.795	0.764	0.827	4.26E-05	0.01453	0.01470	0.00291
	0.800	5000	0.014500	0.0183	0.792	0.790	0.759	0.822	4.29E-05	0.01442	0.01458	0.00296
	0.825	5000	0.014415	0.0183	0.787	0.786	0.755	0.818	4.23E-05	0.01433	0.01450	0.00294
	0.850	5000	0.014291	0.0183	0.780	0.781	0.751	0.813	4.29E-05	0.01421	0.01437	0.00300
	0.875	5000	0.014214	0.0183	0.776	0.778	0.747	0.809	4.17E-05	0.01413	0.01430	0.00294
	0.900	5000	0.014238	0.0183	0.777	0.775	0.744	0.806	4.12E-05	0.01416	0.01432	0.00289
	0.925	5000	0.014138	0.0183	0.772	0.772	0.742	0.803	4.16E-05	0.01406	0.01422	0.00294
	0.950	5000	0.014165	0.0183	0.773	0.770	0.740	0.802	4.06E-05	0.01409	0.01424	0.00287
	0.975	5000	0.014140	0.0183	0.772	0.769	0.739	0.800	4.11E-05	0.01406	0.01422	0.00291

**Table A.18:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on 5,000 replications:  $\gamma = 1, b = 20, \rho = 0.8, \bar{\lambda} = 0.8, \mu = 1, \beta = 0.2$ 

$\gamma=0.25$	position	n	$\hat{p}$	$\exp(-\theta^*b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	5000	0.015877	0.0183	0.867	0.865	0.831	0.900	2.36E-05	0.01583	0.01592	0.00148
	0.025	5000	0.015880	0.0183	0.867	0.865	0.831	0.900	2.33E-05	0.01583	0.01593	0.00147
	0.050	5000	0.015923	0.0183	0.869	0.866	0.832	0.902	2.32E-05	0.01588	0.01597	0.00146
	0.075	5000	0.015942	0.0183	0.870	0.868	0.834	0.904	2.36E-05	0.01590	0.01599	0.00148
	0.100	5000	0.015996	0.0183	0.873	0.871	0.837	0.907	2.41E-05	0.01595	0.01604	0.00150
	0.125	5000	0.016072	0.0183	0.878	0.875	0.841	0.911	2.30E-05	0.01603	0.01612	0.00143
	0.150	5000	0.016184	0.0183	0.884	0.879	0.845	0.915	2.35E-05	0.01614	0.01623	0.00145
	0.175	5000	0.016232	0.0183	0.886	0.884	0.849	0.920	2.38E-05	0.01618	0.01628	0.00147
	0.200	5000	0.016293	0.0183	0.890	0.889	0.854	0.925	2.45E-05	0.01625	0.01634	0.00150
	0.225	5000	0.016422	0.0183	0.897	0.894	0.859	0.931	2.43E-05	0.01637	0.01647	0.00148
	0.250	5000	0.016556	0.0183	0.904	0.900	0.865	0.937	2.36E-05	0.01651	0.01660	0.00142
	0.275	5000	0.016641	0.0183	0.909	0.906	0.870	0.943	2.43E-05	0.01659	0.01669	0.00146
	0.300	5000	0.016743	0.0183	0.914	0.911	0.875	0.948	2.45E-05	0.01669	0.01679	0.00147
	0.325	5000	0.016778	0.0183	0.916	0.916	0.881	0.954	2.49E-05	0.01673	0.01683	0.00149
	0.350	5000	0.016989	0.0183	0.928	0.921	0.885	0.959	2.48E-05	0.01694	0.01704	0.00146
	0.375	5000	0.017009	0.0183	0.929	0.926	0.890	0.964	2.50E-05	0.01696	0.01706	0.00147
	0.400	5000	0.017058	0.0183	0.931	0.930	0.893	0.968	2.52E-05	0.01701	0.01711	0.00148
	0.425	5000	0.017128	0.0183	0.935	0.933	0.896	0.971	2.55 E-05	0.01708	0.01718	0.00149
	0.450	5000	0.017153	0.0183	0.937	0.935	0.898	0.973	2.50E-05	0.01710	0.01720	0.00146
	0.475	5000	0.017128	0.0183	0.935	0.936	0.900	0.974	2.60E-05	0.01708	0.01718	0.00152
	0.500	5000	0.017247	0.0183	0.942	0.937	0.900	0.975	2.46E-05	0.01720	0.01730	0.00143
	0.525	5000	0.017189	0.0183	0.938	0.936	0.900	0.974	2.52E-05	0.01714	0.01724	0.00147
	0.550	5000	0.017176	0.0183	0.938	0.935	0.898	0.973	2.53E-05	0.01713	0.01723	0.00147
	0.575	5000	0.017126	0.0183	0.935	0.933	0.896	0.971	2.56E-05	0.01708	0.01718	0.00150
	0.600	5000	0.017047	0.0183	0.931	0.930	0.893	0.968	2.51E-05	0.01700	0.01710	0.00147
	0.625	5000	0.016999	0.0183	0.928	0.926	0.890	0.964	2.56E-05	0.01695	0.01705	0.00150
	0.650	5000	0.016917	0.0183	0.924	0.921	0.885	0.959	2.54E-05	0.01687	0.01697	0.00150
	0.675	5000	0.016858	0.0183	0.920	0.916	0.881	0.954	2.44E-05	0.01681	0.01691	0.00144
	0.700	5000	0.016752	0.0183	0.915	0.911	0.875	0.948	2.46E-05	0.01670	0.01680	0.00147
	0.725	5000	0.016642	0.0183	0.909	0.906	0.870	0.943	2.45E-05	0.01659	0.01669	0.00147
	0.750	5000	0.016563	0.0183	0.904	0.900	0.865	0.937	2.39E-05	0.01652	0.01661	0.00145
	0.775	5000	0.016440	0.0183	0.898	0.894	0.859	0.931	2.37E-05	0.01639	0.01649	0.00144
	0.800	5000	0.016280	0.0183	0.889	0.889	0.854	0.925	2.50E-05	0.01623	0.01633	0.00154
	0.825	5000	0.016219	0.0183	0.886	0.884	0.849	0.920	2.40E-05	0.01617	0.01627	0.00148
	0.850	5000	0.016130	0.0183	0.881	0.879	0.845	0.915	2.42E-05	0.01608	0.01618	0.00150
	0.875	5000	0.016051	0.0183	0.876	0.875	0.841	0.911	2.43E-05	0.01600	0.01610	0.00151
	0.900	5000	0.015954	0.0183	0.871	0.871	0.837	0.907	2.44E-05	0.01591	0.01600	0.00153
	0.925	5000	0.015943	0.0183	0.870	0.868	0.834	0.904	2.40E-05	0.01590	0.01599	0.00150
	0.950	5000	0.015877	0.0183	0.867	0.866	0.832	0.902	2.38E-05	0.01583	0.01592	0.00150
	0.975	5000	0.015857	0.0183	0.866	0.865	0.831	0.900	2.37 E-05	0.01581	0.01590	0.00149

**Table A.19:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.25, b = 40, \rho = 0.9, \bar{\lambda} = 0.9, \mu = 1, \beta = 0.1$ 

$\gamma = \frac{1}{16}$	position	n	$\hat{p}$	$\exp(-\theta^*b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
	0.000	5000	0.01676	0.0183	0.915	0.913	0.877	0.950	1.35E-05	0.01674	0.01679	8.03E-04
	0.025	5000	0.01678	0.0183	0.916	0.913	0.877	0.950	1.36E-05	0.01675	0.01680	8.13E-04
	0.050	5000	0.01681	0.0183	0.918	0.915	0.879	0.952	1.36E-05	0.01678	0.01684	8.06E-04
	0.075	5000	0.01685	0.0183	0.920	0.917	0.881	0.954	1.37E-05	0.01682	0.01687	8.12E-04
	0.100	5000	0.01689	0.0183	0.922	0.920	0.884	0.957	1.40E-05	0.01686	0.01692	8.31E-04
	0.125	5000	0.01697	0.0183	0.926	0.924	0.887	0.961	1.37E-05	0.01694	0.01700	8.07E-04
	0.150	5000	0.01701	0.0183	0.929	0.928	0.892	0.966	1.42E-05	0.01698	0.01704	8.32E-04
	0.175	5000	0.01713	0.0183	0.935	0.933	0.896	0.971	1.41E-05	0.01710	0.01716	8.22E-04
	0.200	5000	0.01725	0.0183	0.942	0.938	0.902	0.977	1.40E-05	0.01722	0.01727	8.09E-04
	0.225	5000	0.01732	0.0183	0.946	0.944	0.907	0.983	1.40E-05	0.01730	0.01735	8.09E-04
	0.250	5000	0.01744	0.0183	0.952	0.950	0.913	0.989	1.42E-05	0.01741	0.01747	8.14E-04
	0.275	5000	0.01754	0.0183	0.958	0.956	0.918	0.995	1.44E-05	0.01752	0.01757	8.19E-04
	0.300	5000	0.01769	0.0183	0.966	0.962	0.924	1.001	1.42E-05	0.01766	0.01771	8.03E-04
	0.325	5000	0.01778	0.0183	0.971	0.967	0.929	1.007	1.44E-05	0.01775	0.01781	8.12E-04
	0.350	5000	0.01787	0.0183	0.976	0.973	0.934	1.012	1.46E-05	0.01784	0.01790	8.18E-04
	0.375	5000	0.01794	0.0183	0.980	0.977	0.939	1.017	1.47E-05	0.01791	0.01797	8.20E-04
	0.400	5000	0.01801	0.0183	0.983	0.981	0.943	1.021	1.46E-05	0.01798	0.01804	8.11E-04
	0.425	5000	0.01809	0.0183	0.988	0.984	0.946	1.025	1.48E-05	0.01806	0.01812	8.20E-04
	0.450	5000	0.01813	0.0183	0.990	0.987	0.948	1.027	1.48E-05	0.01810	0.01816	8.16E-04
	0.475	5000	0.01817	0.0183	0.992	0.988	0.950	1.029	1.46E-05	0.01814	0.01820	8.06E-04
	0.500	5000	0.01815	0.0183	0.991	0.989	0.950	1.029	1.50E-05	0.01812	0.01817	8.26E-04
	0.525	5000	0.01817	0.0183	0.992	0.988	0.950	1.029	1.46E-05	0.01814	0.01819	8.05E-04
	0.550	5000	0.01813	0.0183	0.990	0.987	0.948	1.027	1.48E-05	0.01810	0.01816	8.14E-04
	0.575	5000	0.01811	0.0183	0.989	0.984	0.946	1.025	1.44E-05	0.01808	0.01814	7.94E-04
	0.600	5000	0.01804	0.0183	0.985	0.981	0.943	1.021	1.44E-05	0.01801	0.01806	8.00E-04
	0.625	5000	0.01796	0.0183	0.980	0.977	0.939	1.017	1.49E-05	0.01793	0.01798	8.33E-04
	0.650	5000	0.01786	0.0183	0.975	0.973	0.934	1.012	1.44E-05	0.01783	0.01788	8.08E-04
	0.675	5000	0.01775	0.0183	0.969	0.967	0.929	1.007	1.47E-05	0.01772	0.01778	8.26E-04
	0.700	5000	0.01767	0.0183	0.965	0.962	0.924	1.001	1.43E-05	0.01764	0.01769	8.09E-04
	0.725	5000	0.01757	0.0183	0.959	0.956	0.918	0.995	1.40E-05	0.01754	0.01759	8.00E-04
	0.750	5000	0.01743	0.0183	0.952	0.950	0.913	0.989	1.45E-05	0.01740	0.01746	8.34E-04
	0.775	5000	0.01734	0.0183	0.947	0.944	0.907	0.983	1.42E-05	0.01731	0.01737	8.20E-04
	0.800	5000	0.01724	0.0183	0.941	0.938	0.902	0.977	1.41E-05	0.01721	0.01726	8.20E-04
	0.825	5000	0.01713	0.0183	0.935	0.933	0.896	0.971	1.37E-05	0.01710	0.01715	8.00E-04
	0.850	5000	0.01705	0.0183	0.931	0.928	0.892	0.966	1.37 E-05	0.01702	0.01708	8.06E-04
	0.875	5000	0.01697	0.0183	0.926	0.924	0.887	0.961	1.37E-05	0.01694	0.01699	8.08E-04
	0.900	5000	0.01689	0.0183	0.922	0.920	0.884	0.957	1.39E-05	0.01686	0.01692	8.21E-04
	0.925	5000	0.01685	0.0183	0.920	0.917	0.881	0.954	1.40E-05	0.01682	0.01688	8.32E-04
	0.950	5000	0.01678	0.0183	0.916	0.915	0.879	0.952	1.36E-05	0.01675	0.01681	8.11E-04
	0.975	5000	0.01676	0.0183	0.915	0.913	0.877	0.950	1.37E-05	0.01674	0.01679	8.18E-04

**Table A.20:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on 5,000 replications:  $\gamma = \frac{1}{16}, b = 80, \rho = 0.95, \bar{\lambda} = 0.95, \mu = 1, \beta = 0.05$ 

position	n	$\hat{p}$	$\exp(-\theta^*b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
0.000	5000	0.014826	0.0183	0.809	0.807	0.775	0.840	3.46E-05	0.01476	0.01489	0.00233
0.025	5000	0.014831	0.0183	0.810	0.807	0.776	0.840	3.47E-05	0.01476	0.01490	0.00234
0.050	5000	0.014861	0.0183	0.811	0.809	0.777	0.842	3.46E-05	0.01479	0.01493	0.00233
0.075	5000	0.014898	0.0183	0.813	0.811	0.779	0.844	3.48E-05	0.01483	0.01497	0.00233
0.100	5000	0.014968	0.0183	0.817	0.813	0.781	0.846	3.48E-05	0.01490	0.01504	0.00233
0.125	5000	0.015019	0.0183	0.820	0.817	0.785	0.850	3.50E-05	0.01495	0.01509	0.00233
0.150	5000	0.015084	0.0183	0.824	0.820	0.788	0.854	3.52E-05	0.01502	0.01515	0.00234
0.175	5000	0.015164	0.0183	0.828	0.825	0.793	0.859	3.54E-05	0.01509	0.01523	0.00234
0.200	5000	0.015243	0.0183	0.832	0.830	0.797	0.864	3.55E-05	0.01517	0.01531	0.00233
0.225	5000	0.015338	0.0183	0.837	0.835	0.802	0.869	3.57E-05	0.01527	0.01541	0.00233
0.250	5000	0.015455	0.0183	0.844	0.840	0.807	0.874	3.59E-05	0.01538	0.01553	0.00232
0.275	5000	0.015555	0.0183	0.849	0.845	0.812	0.880	3.62E-05	0.01548	0.01563	0.00232
0.300	5000	0.015650	0.0183	0.854	0.850	0.817	0.885	3.63E-05	0.01558	0.01572	0.00232
0.325	5000	0.015733	0.0183	0.859	0.855	0.822	0.890	3.65 E-05	0.01566	0.01580	0.00232
0.350	5000	0.015816	0.0183	0.863	0.860	0.826	0.895	3.68E-05	0.01574	0.01589	0.00232
0.375	5000	0.015889	0.0183	0.868	0.864	0.830	0.899	3.70E-05	0.01582	0.01596	0.00233
0.400	5000	0.015947	0.0183	0.871	0.868	0.834	0.903	3.71E-05	0.01587	0.01602	0.00233
0.425	5000	0.016011	0.0183	0.874	0.870	0.836	0.906	3.70E-05	0.01594	0.01608	0.00231
0.450	5000	0.016053	0.0183	0.876	0.873	0.838	0.908	3.71E-05	0.01598	0.01613	0.00231
0.475	5000	0.016065	0.0183	0.877	0.874	0.840	0.910	3.73E-05	0.01599	0.01614	0.00232
0.500	5000	0.016088	0.0183	0.878	0.874	0.840	0.910	3.73E-05	0.01601	0.01616	0.00232
0.525	5000	0.016074	0.0183	0.878	0.874	0.840	0.910	3.73E-05	0.01600	0.01615	0.00232
0.550	5000	0.016037	0.0183	0.876	0.873	0.838	0.908	3.73E-05	0.01596	0.01611	0.00233
0.575	5000	0.016004	0.0183	0.874	0.870	0.836	0.906	3.71E-05	0.01593	0.01608	0.00232
0.600	5000	0.015944	0.0183	0.871	0.868	0.834	0.903	3.70E-05	0.01587	0.01602	0.00232
0.625	5000	0.015879	0.0183	0.867	0.864	0.830	0.899	3.68E-05	0.01581	0.01595	0.00232
0.650	5000	0.015802	0.0183	0.863	0.860	0.826	0.895	3.65 E-05	0.01573	0.01587	0.00231
0.675	5000	0.015720	0.0183	0.858	0.855	0.822	0.890	3.63E-05	0.01565	0.01579	0.00231
0.700	5000	0.015624	0.0183	0.853	0.850	0.817	0.885	3.60E-05	0.01555	0.01569	0.00230
0.725	5000	0.015530	0.0183	0.848	0.845	0.812	0.880	3.57E-05	0.01546	0.01560	0.00230
0.750	5000	0.015420	0.0183	0.842	0.840	0.807	0.874	3.57E-05	0.01535	0.01549	0.00232
0.775	5000	0.015316	0.0183	0.836	0.835	0.802	0.869	3.56E-05	0.01525	0.01539	0.00233
0.800	5000	0.015214	0.0183	0.831	0.830	0.797	0.864	3.56E-05	0.01514	0.01528	0.00234
0.825	5000	0.015126	0.0183	0.826	0.825	0.793	0.859	3.53E-05	0.01506	0.01520	0.00234
0.850	5000	0.015050	0.0183	0.822	0.820	0.788	0.854	3.51E-05	0.01498	0.01512	0.00233
0.875	5000	0.014981	0.0183	0.818	0.817	0.785	0.850	3.49E-05	0.01491	0.01505	0.00233
0.900	5000	0.014927	0.0183	0.815	0.813	0.781	0.846	3.46E-05	0.01486	0.01499	0.00232
0.925	5000	0.014869	0.0183	0.812	0.811	0.779	0.844	3.46E-05	0.01480	0.01494	0.00233
0.950	5000	0.014835	0.0183	0.810	0.809	0.777	0.842	3.46E-05	0.01477	0.01490	0.00233
0.975	5000	0.014820	0.0183	0.809	0.807	0.776	0.840	3.46E-05	0.01475	0.01489	0.00233

**Table A.21:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.64, b = 25, \rho = 0.84, \bar{\lambda} = 0.84, \mu = 1, \beta = 0.16$ 

position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
0.000	5000	0.016264	0.0183	0.888	0.884	0.849	0.920	1.95E-05	0.01623	0.01630	0.00120
0.025	5000	0.016268	0.0183	0.888	0.884	0.850	0.920	1.96E-05	0.01623	0.01631	0.00120
0.050	5000	0.016297	0.0183	0.890	0.886	0.851	0.922	1.97 E- 05	0.01626	0.01634	0.00121
0.075	5000	0.016327	0.0183	0.891	0.888	0.853	0.924	1.97 E- 05	0.01629	0.01637	0.00121
0.100	5000	0.016389	0.0183	0.895	0.891	0.856	0.927	1.97 E- 05	0.01635	0.01643	0.00120
0.125	5000	0.016460	0.0183	0.899	0.894	0.859	0.931	1.98E-05	0.01642	0.01650	0.00120
0.150	5000	0.016532	0.0183	0.903	0.899	0.863	0.935	1.98E-05	0.01649	0.01657	0.00120
0.175	5000	0.016615	0.0183	0.907	0.903	0.868	0.940	1.98E-05	0.01658	0.01665	0.00119
0.200	5000	0.016718	0.0183	0.913	0.909	0.873	0.946	2.00E-05	0.01668	0.01676	0.00119
0.225	5000	0.016825	0.0183	0.919	0.914	0.878	0.952	2.01E-05	0.01679	0.01686	0.00119
0.250	5000	0.016926	0.0183	0.924	0.920	0.884	0.958	2.02E-05	0.01689	0.01697	0.00119
0.275	5000	0.017021	0.0183	0.929	0.926	0.889	0.964	2.03E-05	0.01698	0.01706	0.00119
0.300	5000	0.017132	0.0183	0.935	0.931	0.895	0.969	2.03E-05	0.01709	0.01717	0.00118
0.325	5000	0.017229	0.0183	0.941	0.937	0.900	0.975	2.04E-05	0.01719	0.01727	0.00118
0.350	5000	0.017318	0.0183	0.946	0.942	0.905	0.980	2.06E-05	0.01728	0.01736	0.00119
0.375	5000	0.017400	0.0183	0.950	0.946	0.909	0.985	2.09E-05	0.01736	0.01744	0.00120
0.400	5000	0.017465	0.0183	0.954	0.950	0.913	0.989	2.09E-05	0.01742	0.01751	0.00120
0.425	5000	0.017529	0.0183	0.957	0.953	0.916	0.992	2.09E-05	0.01749	0.01757	0.00119
0.450	5000	0.017573	0.0183	0.959	0.956	0.918	0.995	2.10E-05	0.01753	0.01761	0.00120
0.475	5000	0.017604	0.0183	0.961	0.957	0.920	0.996	2.11E-05	0.01756	0.01765	0.00120
0.500	5000	0.017604	0.0183	0.961	0.958	0.920	0.997	2.10E-05	0.01756	0.01764	0.00119
0.525	5000	0.017595	0.0183	0.961	0.957	0.920	0.996	2.10E-05	0.01755	0.01764	0.00119
0.550	5000	0.017559	0.0183	0.959	0.956	0.918	0.995	2.10E-05	0.01752	0.01760	0.00119
0.575	5000	0.017512	0.0183	0.956	0.953	0.916	0.992	2.09E-05	0.01747	0.01755	0.00119
0.600	5000	0.017445	0.0183	0.952	0.950	0.913	0.989	2.07E-05	0.01740	0.01749	0.00119
0.625	5000	0.017372	0.0183	0.948	0.946	0.909	0.985	2.06E-05	0.01733	0.01741	0.00119
0.650	5000	0.017298	0.0183	0.944	0.942	0.905	0.980	2.05E-05	0.01726	0.01734	0.00119
0.675	5000	0.017212	0.0183	0.940	0.937	0.900	0.975	2.05E-05	0.01717	0.01725	0.00119
0.700	5000	0.017114	0.0183	0.934	0.931	0.895	0.969	2.04E-05	0.01707	0.01715	0.00119
0.725	5000	0.017014	0.0183	0.929	0.926	0.889	0.964	2.03E-05	0.01697	0.01705	0.00119
0.750	5000	0.016918	0.0183	0.924	0.920	0.884	0.958	2.01E-05	0.01688	0.01696	0.00119
0.775	5000	0.016822	0.0183	0.918	0.914	0.878	0.952	1.98E-05	0.01678	0.01686	0.00118
0.800	5000	0.016727	0.0183	0.913	0.909	0.873	0.946	1.97 E- 05	0.01669	0.01677	0.00118
0.825	5000	0.016626	0.0183	0.908	0.903	0.868	0.940	1.97 E- 05	0.01659	0.01666	0.00118
0.850	5000	0.016535	0.0183	0.903	0.899	0.863	0.935	1.95E-05	0.01650	0.01657	0.00118
0.875	5000	0.016462	0.0183	0.899	0.894	0.859	0.931	1.95E-05	0.01642	0.01650	0.00119
0.900	5000	0.016390	0.0183	0.895	0.891	0.856	0.927	1.96E-05	0.01635	0.01643	0.00119
0.925	5000	0.016332	0.0183	0.892	0.888	0.853	0.924	1.95E-05	0.01629	0.01637	0.00120
0.950	5000	0.016291	0.0183	0.889	0.886	0.851	0.922	1.95E-05	0.01625	0.01633	0.00120
0.975	5000	0.016271	0.0183	0.888	0.884	0.850	0.920	1.95E-05	0.01623	0.01631	0.00120

**Table A.22:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.16, b = 50, \rho = 0.92, \bar{\lambda} = 0.92, \mu = 1, \beta = 0.08$ 

position	n	$\hat{p}$	$\exp(-\theta^*b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
0.000	5000	0.01695	0.0183	0.926	0.922	0.886	0.960	1.18E-05	0.01693	0.01698	6.96E-04
0.025	5000	0.01696	0.0183	0.926	0.923	0.887	0.960	1.18E-05	0.01693	0.01698	6.98E-04
0.050	5000	0.01698	0.0183	0.927	0.924	0.888	0.962	1.19E-05	0.01695	0.01700	7.02E-04
0.075	5000	0.01702	0.0183	0.929	0.926	0.890	0.964	1.19E-05	0.01699	0.01704	6.97E-04
0.100	5000	0.01707	0.0183	0.932	0.929	0.893	0.967	1.20E-05	0.01705	0.01710	7.00E-04
0.125	5000	0.01714	0.0183	0.936	0.933	0.897	0.971	1.19E-05	0.01712	0.01717	6.97E-04
0.150	5000	0.01722	0.0183	0.940	0.938	0.901	0.976	1.21E-05	0.01719	0.01724	7.01E-04
0.175	5000	0.01731	0.0183	0.945	0.943	0.906	0.981	1.21E-05	0.01729	0.01733	7.00E-04
0.200	5000	0.01741	0.0183	0.951	0.948	0.911	0.987	1.22E-05	0.01739	0.01743	6.99E-04
0.225	5000	0.01752	0.0183	0.956	0.954	0.917	0.993	1.23E-05	0.01749	0.01754	7.04E-04
0.250	5000	0.01763	0.0183	0.962	0.960	0.922	0.999	1.22E-05	0.01760	0.01765	6.94E-04
0.275	5000	0.01774	0.0183	0.968	0.966	0.928	1.005	1.24E-05	0.01771	0.01776	6.98E-04
0.300	5000	0.01784	0.0183	0.974	0.972	0.934	1.012	1.26E-05	0.01782	0.01787	7.05E-04
0.325	5000	0.01794	0.0183	0.979	0.978	0.939	1.017	1.27E-05	0.01791	0.01796	7.07E-04
0.350	5000	0.01804	0.0183	0.985	0.983	0.944	1.023	1.28E-05	0.01801	0.01806	7.08E-04
0.375	5000	0.01812	0.0183	0.989	0.988	0.949	1.028	1.27E-05	0.01809	0.01814	7.02E-04
0.400	5000	0.01820	0.0183	0.993	0.992	0.953	1.032	1.27E-05	0.01817	0.01822	6.99E-04
0.425	5000	0.01826	0.0183	0.997	0.995	0.956	1.035	1.27E-05	0.01824	0.01829	6.95E-04
0.450	5000	0.01830	0.0183	0.999	0.997	0.958	1.038	1.29E-05	0.01827	0.01833	7.03E-04
0.475	5000	0.01833	0.0183	1.001	0.999	0.960	1.039	1.30E-05	0.01830	0.01835	7.07E-04
0.500	5000	0.01834	0.0183	1.001	0.999	0.960	1.040	1.30E-05	0.01831	0.01836	7.09E-04
0.525	5000	0.01834	0.0183	1.001	0.999	0.960	1.039	1.29E-05	0.01831	0.01836	7.06E-04
0.550	5000	0.01832	0.0183	1.000	0.997	0.958	1.038	1.28E-05	0.01829	0.01834	7.00E-04
0.575	5000	0.01828	0.0183	0.998	0.995	0.956	1.035	1.27E-05	0.01825	0.01830	6.93E-04
0.600	5000	0.01822	0.0183	0.995	0.992	0.953	1.032	1.27E-05	0.01819	0.01824	6.95E-04
0.625	5000	0.01815	0.0183	0.991	0.988	0.949	1.028	1.26E-05	0.01813	0.01818	6.95E-04
0.650	5000	0.01807	0.0183	0.986	0.983	0.944	1.023	1.25E-05	0.01804	0.01809	6.92E-04
0.675	5000	0.01797	0.0183	0.981	0.978	0.939	1.017	1.24E-05	0.01795	0.01799	6.89E-04
0.700	5000	0.01788	0.0183	0.976	0.972	0.934	1.012	1.23E-05	0.01785	0.01790	6.87E-04
0.725	5000	0.01777	0.0183	0.970	0.966	0.928	1.005	1.22E-05	0.01774	0.01779	6.89E-04
0.750	5000	0.01766	0.0183	0.964	0.960	0.922	0.999	1.22E-05	0.01763	0.01768	6.94E-04
0.775	5000	0.01755	0.0183	0.958	0.954	0.917	0.993	1.22E-05	0.01752	0.01757	6.93E-04
0.800	5000	0.01744	0.0183	0.952	0.948	0.911	0.987	1.22E-05	0.01741	0.01746	6.97E-04
0.825	5000	0.01734	0.0183	0.947	0.943	0.906	0.981	1.20E-05	0.01731	0.01736	6.92E-04
0.850	5000	0.01724	0.0183	0.941	0.938	0.901	0.976	1.20E-05	0.01722	0.01727	6.98E-04
0.875	5000	0.01717	0.0183	0.937	0.933	0.897	0.971	1.19E-05	0.01714	0.01719	6.93E-04
0.900	5000	0.01710	0.0183	0.933	0.929	0.893	0.967	1.18E-05	0.01707	0.01712	6.90E-04
0.925	5000	0.01704	0.0183	0.930	0.926	0.890	0.964	1.18E-05	0.01701	0.01706	6.93E-04
0.950	5000	0.01699	0.0183	0.928	0.924	0.888	0.962	1.18E-05	0.01697	0.01701	6.96E-04
0.975	5000	0.01696	0.0183	0.926	0.923	0.887	0.960	1.18E-05	0.01694	0.01699	6.98E-04

**Table A.23:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.04, b = 100, \rho = 0.96, \bar{\lambda} = 0.96, \mu = 1, \beta = 0.04$ 

position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
0.000	5000	0.017295	0.0183	0.944	0.942	0.905	0.980	8.47E-06	0.01728	0.01731	4.90E-04
0.025	5000	0.017303	0.0183	0.945	0.942	0.905	0.980	8.47E-06	0.01729	0.01732	4.89E-04
0.050	5000	0.017327	0.0183	0.946	0.943	0.906	0.982	8.51E-06	0.01731	0.01734	4.91E-04
0.075	5000	0.017370	0.0183	0.948	0.946	0.909	0.984	8.51E-06	0.01735	0.01739	4.90E-04
0.100	5000	0.017426	0.0183	0.951	0.949	0.912	0.988	8.52E-06	0.01741	0.01744	4.89E-04
0.125	5000	0.017499	0.0183	0.955	0.953	0.915	0.992	8.50E-06	0.01748	0.01752	4.86E-04
0.150	5000	0.017588	0.0183	0.960	0.957	0.920	0.996	8.52E-06	0.01757	0.01761	4.85E-04
0.175	5000	0.017676	0.0183	0.965	0.962	0.925	1.002	8.64E-06	0.01766	0.01769	4.89E-04
0.200	5000	0.017778	0.0183	0.971	0.968	0.930	1.007	8.72E-06	0.01776	0.01779	4.90E-04
0.225	5000	0.017885	0.0183	0.976	0.974	0.936	1.014	8.79E-06	0.01787	0.01790	4.91E-04
0.250	5000	0.017994	0.0183	0.982	0.980	0.942	1.020	8.85E-06	0.01798	0.01801	4.92E-04
0.275	5000	0.018114	0.0183	0.989	0.986	0.947	1.026	8.85E-06	0.01810	0.01813	4.89E-04
0.300	5000	0.018221	0.0183	0.995	0.992	0.953	1.033	8.95E-06	0.01820	0.01824	4.91E-04
0.325	5000	0.018322	0.0183	1.000	0.998	0.959	1.039	8.99E-06	0.01830	0.01834	4.91E-04
0.350	5000	0.018420	0.0183	1.006	1.003	0.964	1.044	9.05 E-06	0.01840	0.01844	4.91E-04
0.375	5000	0.018515	0.0183	1.011	1.008	0.969	1.049	9.09E-06	0.01850	0.01853	4.91E-04
0.400	5000	0.018592	0.0183	1.015	1.012	0.973	1.054	9.05 E-06	0.01857	0.01861	4.87E-04
0.425	5000	0.018647	0.0183	1.018	1.016	0.976	1.057	9.06E-06	0.01863	0.01866	4.86E-04
0.450	5000	0.018691	0.0183	1.021	1.018	0.978	1.060	9.09E-06	0.01867	0.01871	4.86E-04
0.475	5000	0.018717	0.0183	1.022	1.019	0.980	1.061	9.20E-06	0.01870	0.01873	4.92E-04
0.500	5000	0.018729	0.0183	1.023	1.020	0.980	1.062	9.20E-06	0.01871	0.01875	4.91E-04
0.525	5000	0.018720	0.0183	1.022	1.019	0.980	1.061	9.15 E-06	0.01870	0.01874	4.89E-04
0.550	5000	0.018688	0.0183	1.020	1.018	0.978	1.060	9.13E-06	0.01867	0.01871	4.89E-04
0.575	5000	0.018647	0.0183	1.018	1.016	0.976	1.057	9.14E-06	0.01863	0.01866	4.90E-04
0.600	5000	0.018589	0.0183	1.015	1.012	0.973	1.054	9.17 E-06	0.01857	0.01861	4.93E-04
0.625	5000	0.018515	0.0183	1.011	1.008	0.969	1.049	9.13E-06	0.01850	0.01853	4.93E-04
0.650	5000	0.018432	0.0183	1.006	1.003	0.964	1.044	9.05E-06	0.01841	0.01845	4.91E-04
0.675	5000	0.018331	0.0183	1.001	0.998	0.959	1.039	8.94E-06	0.01831	0.01835	4.88E-04
0.700	5000	0.018222	0.0183	0.995	0.992	0.953	1.033	8.81E-06	0.01821	0.01824	4.83E-04
0.725	5000	0.018117	0.0183	0.989	0.986	0.947	1.026	8.71E-06	0.01810	0.01813	4.81E-04
0.750	5000	0.017999	0.0183	0.983	0.980	0.942	1.020	8.61E-06	0.01798	0.01802	4.78E-04
0.775	5000	0.017891	0.0183	0.977	0.974	0.936	1.014	$8.65 \text{E}{-}06$	0.01787	0.01791	4.84E-04
0.800	5000	0.017786	0.0183	0.971	0.968	0.930	1.007	8.62E-06	0.01777	0.01780	4.84E-04
0.825	5000	0.017674	0.0183	0.965	0.962	0.925	1.002	8.62E-06	0.01766	0.01769	4.88E-04
0.850	5000	0.017581	0.0183	0.960	0.957	0.920	0.996	8.59E-06	0.01756	0.01760	4.89E-04
0.875	5000	0.017493	0.0183	0.955	0.953	0.915	0.992	8.62E-06	0.01748	0.01751	4.93E-04
0.900	5000	0.017422	0.0183	0.951	0.949	0.912	0.988	8.60E-06	0.01741	0.01744	4.94E-04
0.925	5000	0.017366	0.0183	0.948	0.946	0.909	0.984	8.56E-06	0.01735	0.01738	4.93E-04
0.950	5000	0.017327	0.0183	0.946	0.943	0.906	0.982	8.48E-06	0.01731	0.01734	4.89E-04
0.975	5000	0.017299	0.0183	0.944	0.942	0.905	0.980	8.50E-06	0.01728	0.01732	4.91E-04

**Table A.24:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.01, b = 200, \rho = 0.98, \bar{\lambda} = 0.98, \mu = 1, \beta = 0.02$ 

position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
0.000	5000	0.017470	0.0183	0.954	0.951	0.914	0.990	7.39E-06	0.01746	0.01748	4.23E-04
0.025	5000	0.017482	0.0183	0.954	0.952	0.914	0.990	7.30E-06	0.01747	0.01750	4.18E-04
0.050	5000	0.017506	0.0183	0.956	0.953	0.916	0.992	7.47E-06	0.01749	0.01752	4.27E-04
0.075	5000	0.017559	0.0183	0.959	0.955	0.918	0.994	7.48E-06	0.01754	0.01757	4.26E-04
0.100	5000	0.017617	0.0183	0.962	0.958	0.921	0.998	7.31E-06	0.01760	0.01763	4.15E-04
0.125	5000	0.017676	0.0183	0.965	0.962	0.925	1.002	7.45E-06	0.01766	0.01769	4.21E-04
0.150	5000	0.017760	0.0183	0.970	0.967	0.929	1.006	7.53E-06	0.01775	0.01777	4.24E-04
0.175	5000	0.017868	0.0183	0.976	0.972	0.934	1.012	7.56E-06	0.01785	0.01788	4.23E-04
0.200	5000	0.017955	0.0183	0.980	0.978	0.939	1.018	7.59E-06	0.01794	0.01797	4.23E-04
0.225	5000	0.018087	0.0183	0.987	0.984	0.945	1.024	7.68E-06	0.01807	0.01810	4.25E-04
0.250	5000	0.018188	0.0183	0.993	0.990	0.951	1.030	7.68E-06	0.01817	0.01820	4.22E-04
0.275	5000	0.018299	0.0183	0.999	0.996	0.957	1.037	7.64E-06	0.01828	0.01831	4.18E-04
0.300	5000	0.018407	0.0183	1.005	1.002	0.963	1.043	7.83E-06	0.01839	0.01842	4.25E-04
0.325	5000	0.018526	0.0183	1.011	1.008	0.969	1.049	7.78E-06	0.01851	0.01854	4.20E-04
0.350	5000	0.018632	0.0183	1.017	1.014	0.974	1.055	7.86E-06	0.01862	0.01865	4.22E-04
0.375	5000	0.018704	0.0183	1.021	1.018	0.978	1.060	7.89E-06	0.01869	0.01872	4.22E-04
0.400	5000	0.018775	0.0183	1.025	1.023	0.982	1.064	7.90E-06	0.01876	0.01879	4.21E-04
0.425	5000	0.018842	0.0183	1.029	1.026	0.986	1.068	7.97 E-06	0.01883	0.01886	4.23E-04
0.450	5000	0.018875	0.0183	1.031	1.028	0.988	1.070	7.90E-06	0.01886	0.01889	4.19E-04
0.475	5000	0.018911	0.0183	1.033	1.030	0.990	1.072	8.03E-06	0.01890	0.01893	4.25E-04
0.500	5000	0.018931	0.0183	1.034	1.030	0.990	1.072	7.97 E-06	0.01892	0.01895	4.21E-04
0.525	5000	0.018921	0.0183	1.033	1.030	0.990	1.072	8.04E-06	0.01891	0.01894	4.25E-04
0.550	5000	0.018884	0.0183	1.031	1.028	0.988	1.070	7.91E-06	0.01887	0.01890	4.19E-04
0.575	5000	0.018839	0.0183	1.029	1.026	0.986	1.068	7.99E-06	0.01882	0.01885	4.24E-04
0.600	5000	0.018780	0.0183	1.025	1.023	0.982	1.064	7.93E-06	0.01876	0.01880	4.22E-04
0.625	5000	0.018713	0.0183	1.022	1.018	0.978	1.060	7.80E-06	0.01870	0.01873	4.17E-04
0.650	5000	0.018621	0.0183	1.017	1.014	0.974	1.055	7.79E-06	0.01861	0.01864	4.19E-04
0.675	5000	0.018513	0.0183	1.011	1.008	0.969	1.049	7.85 E-06	0.01850	0.01853	4.24E-04
0.700	5000	0.018408	0.0183	1.005	1.002	0.963	1.043	7.71E-06	0.01839	0.01842	4.19E-04
0.725	5000	0.018294	0.0183	0.999	0.996	0.957	1.037	7.73E-06	0.01828	0.01831	4.23E-04
0.750	5000	0.018176	0.0183	0.992	0.990	0.951	1.030	7.71E-06	0.01816	0.01819	4.24E-04
0.775	5000	0.018061	0.0183	0.986	0.984	0.945	1.024	7.60E-06	0.01805	0.01808	4.21E-04
0.800	5000	0.017962	0.0183	0.981	0.978	0.939	1.018	7.55E-06	0.01795	0.01798	4.20E-04
0.825	5000	0.017862	0.0183	0.975	0.972	0.934	1.012	7.60 E-06	0.01785	0.01788	4.25E-04
0.850	5000	0.017768	0.0183	0.970	0.967	0.929	1.006	7.52E-06	0.01775	0.01778	4.23E-04
0.875	5000	0.017687	0.0183	0.966	0.962	0.925	1.002	7.45E-06	0.01767	0.01770	4.21E-04
0.900	5000	0.017608	0.0183	0.961	0.958	0.921	0.998	7.38E-06	0.01759	0.01762	4.19E-04
0.925	5000	0.017547	0.0183	0.958	0.955	0.918	0.994	7.42E-06	0.01753	0.01756	4.23E-04
0.950	5000	0.017499	0.0183	0.955	0.953	0.916	0.992	7.43E-06	0.01748	0.01751	4.25E-04
0.975	5000	0.017482	0.0183	0.954	0.952	0.914	0.990	7.40E-06	0.01747	0.01750	4.23E-04

**Table A.25:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/M/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.0025, b = 400, \rho = 0.99, \bar{\lambda} = 0.99, \mu = 1, \beta = 0.01$ 

position	$\rho=0.84$	$\rho=0.92$	$\rho=0.96$	$\rho=0.98$	$\rho=0.99$
0.000	0.96364	0.96523	0.96424	0.96357	0.96344
0.025	0.96397	0.96543	0.96436	0.96398	0.96412
0.050	0.96596	0.96718	0.96545	0.96531	0.96545
0.075	0.96832	0.96896	0.96778	0.96771	0.96839
0.100	0.97289	0.97264	0.97102	0.97086	0.97156
0.125	0.97619	0.97686	0.97504	0.97493	0.97482
0.150	0.98044	0.98109	0.97919	0.97989	0.97945
0.175	0.98562	0.98605	0.98442	0.98475	0.98539
0.200	0.99074	0.99215	0.99018	0.99043	0.99019
0.225	0.99693	0.99851	0.99614	0.99642	0.99747
0.250	1.00456	1.00450	1.00255	1.00251	1.00305
0.275	1.01102	1.01015	1.00875	1.00918	1.00918
0.300	1.01721	1.01669	1.01482	1.01513	1.01514
0.325	1.02258	1.02247	1.02006	1.02079	1.02171
0.350	1.02797	1.02776	1.02572	1.02622	1.02757
0.375	1.03278	1.03264	1.03035	1.03152	1.03152
0.400	1.03650	1.03649	1.03484	1.03582	1.03546
0.425	1.04065	1.04030	1.03871	1.03886	1.03911
0.450	1.04342	1.04286	1.04079	1.04134	1.04094
0.475	1.04420	1.04475	1.04237	1.04276	1.04294
0.500	1.04565	1.04470	1.04278	1.04346	1.04405
0.525	1.04475	1.04420	1.04296	1.04291	1.04351
0.550	1.04236	1.04204	1.04183	1.04118	1.04142
0.575	1.04021	1.03925	1.03955	1.03886	1.03895
0.600	1.03634	1.03532	1.03597	1.03564	1.03569
0.625	1.03213	1.03096	1.03230	1.03150	1.03204
0.650	1.02712	1.02655	1.02745	1.02690	1.02695
0.675	1.02178	1.02146	1.02203	1.02124	1.02099
0.700	1.01554	1.01565	1.01682	1.01521	1.01517
0.725	1.00944	1.00971	1.01054	1.00935	1.00888
0.750	1.00225	1.00404	1.00425	1.00277	1.00241
0.775	0.99551	0.99831	0.99785	0.99674	0.99604
0.800	0.98888	0.99266	0.99178	0.99088	0.99059
0.825	0.98318	0.98671	0.98604	0.98464	0.98509
0.850	0.97825	0.98128	0.98059	0.97950	0.97991
0.875	0.97371	0.97696	0.97629	0.97457	0.97545
0.900	0.97022	0.97270	0.97228	0.97063	0.97108
0.925	0.96647	0.96926	0.96887	0.96748	0.96772
0.950	0.96422	0.96681	0.96626	0.96531	0.96507
0.975	0.96328	0.96564	0.96472	0.96377	0.96414
avg diff w.r.t. last column	0.00037	0.00112	0.00015	-0.00019	0.00000
avg. abs. diff w.r.t. last column	0.00099	0.00121	0.00081	0.00039	0.00000
rmse w.r.t. last column	0.00116	0.00134	0.00096	0.00049	0.00000

**Table A.26:** Comparison of ratio  $P(W_y > b)/P(W > b) = A_y/\rho$  for different  $\rho$ 's with base parameter  $(\beta, \gamma, b) = (1, 25, 4)$  using the scaling in (3.39) of the main thesis

**Table A.27:** Summary of simulation results for  $M_t/M/1$  queue at y = 0 as a function of  $1 - \rho$  with base parameter  $(\beta, \gamma, b) = (1, 25, 4)$  using the scaling in (3.39) of the main thesis:

	$1 - \rho = 0.16$	$1-\rho=0.08$	$1-\rho=0.04$	$1-\rho=0.02$	$1-\rho=0.01$	$1-\rho=0.005$	$1 - \rho = 0.0025$
n	5000	5000	5000	5000	5000	5000	5000
$\hat{p}$	0.014834	0.016239	0.016941	0.017298	0.017462	0.017566	0.017596
$e^{-\theta^* b}$	0.0183	0.0183	0.0183	0.0183	0.0183	0.0183	0.0183
$A_y$	0.810	0.887	0.925	0.944	0.953	0.959	0.961
$A_y$ approxi	0.807	0.884	0.922	0.942	0.951	0.956	0.958
$A_y$ LB	0.775	0.849	0.886	0.905	0.914	0.919	0.921
$A_y$ UB	0.840	0.920	0.960	0.980	0.990	0.995	0.998
s.e.	3.42E-05	1.99E-05	1.16E-05	8.35E-06	7.38E-06	7.09E-06	7.02E-06
95% CI (lb)	0.01477	0.01620	0.01692	0.01728	0.01745	0.01755	0.01758
(ub)	0.01490	0.01628	0.01696	0.01731	0.01748	0.01758	0.01761
r.e.	0.002303	0.001222	0.000685	0.000483	0.000422	0.000403	0.000399
$P(W_y > b) / P(W > b)$							
ratio	0.96419	0.96375	0.96349	0.96370	0.96301	0.96386	0.96312
diff	-0.00107	-0.00062	-0.00037	-0.00058	0.00011	-0.00074	0.00000
abs diff	0.00107	0.00062	0.00037	0.00058	0.00011	0.00074	0.00000

**Table A.28:** Summary of simulation results for  $M_t/M/1$  queue at y = 0 and at y = 0.5 as a function of  $1 - \rho$  with base parameter  $(\beta, \gamma, b) = (1, 2.5, 4)$  using the scaling in (3.39) of the main thesis

	$1 - \rho = 0.16$	$1 - \rho = 0.08$	$1 - \rho = 0.04$	$1 - \rho = 0.02$	$1 - \rho = 0.01$
n	40000	40000	40000	40000	40000
y = 0					
$\hat{p}$	0.011053	0.012192	0.012814	0.013122	0.013263
$e^{-\theta^* b}$	0.0183	0.0183	0.0183	0.0183	0.0183
$A_y$	0.604	0.666	0.700	0.716	0.724
$A_y$ approxi	0.563	0.617	0.644	0.657	0.664
$A_y$ LB	0.377	0.413	0.431	0.440	0.445
$A_y$ UB	0.840	0.920	0.960	0.980	0.990
s.e.	1.75E-05	1.69E-05	1.71E-05	1.73E-05	1.74E-05
95% CI (lb)	0.01102	0.01216	0.01278	0.01309	0.01323
(ub)	0.01109	0.01223	0.01285	0.01316	0.01330
r.e.	0.001582	0.001387	0.001333	0.001319	0.001313
$P(W_y > b) / P(W > b)$					
ratio	0.71845	0.72356	0.72879	0.73103	0.73144
diff w.r.t. last column	0.01298	0.00788	0.00264	0.00041	0.00000
abs diff	0.01298	0.00788	0.00264	0.00041	0.00000
y = 0.5					
$\hat{p}$	0.025888	0.028396	0.029551	0.030110	0.030430
$e^{-\theta^* b}$	0.0183	0.0183	0.0183	0.0183	0.0183
$A_y$	1.413	1.550	1.613	1.644	1.661
$A_y$ approxi	1.253	1.372	1.432	1.462	1.477
$A_y$ LB	0.840	0.920	0.960	0.980	0.990
$A_y$ UB	1.869	2.047	2.137	2.181	2.203
s.e.	3.87E-05	3.74E-05	3.80E-05	3.86E-05	3.89E-05
95% CI (lb)	0.02581	0.02832	0.02948	0.03003	0.03035
(ub)	0.02596	0.02847	0.02963	0.03019	0.03051
r.e.	0.001496	0.001318	0.001286	0.001281	0.001279
$P(W_y > b) / P(W > b)$					
ratio	1 69966	1.68517	1.68068	1.67751	1.67821
14010	1.06200				
diff w.r.t. last column	-0.00445	-0.00696	-0.00247	0.00071	0.00000

# A.3.3 Tail Probabilities for the $(H_2)_t/M/1$ Periodic Queue

Tables A.29-A.38 present results for the  $(H_2)_t/M/1$  model paralleling the results for the  $M_t/M/1$  model in Tables A.18-A.28.

position	n	$\hat{p}$	$\exp(-\theta^*b)$	$A_y$	$A_y/A$ approx	$A_y/A$ LB	$A_y/A$ UB	s.e.	95% CI (lb)	(ub)	r.e.
0.000	5000	0.025326	0.0317	0.799	0.966	0.933	1.000	6.42E-05	0.02520	0.02545	0.00254
0.025	5000	0.025266	0.0317	0.797	0.966	0.934	1.000	6.35E-05	0.02514	0.02539	0.00251
0.050	5000	0.025358	0.0317	0.800	0.968	0.935	1.002	6.40E-05	0.02523	0.02548	0.00252
0.075	5000	0.025503	0.0317	0.805	0.970	0.937	1.004	6.39E-05	0.02538	0.02563	0.00251
0.100	5000	0.025516	0.0317	0.805	0.972	0.939	1.007	6.45E-05	0.02539	0.02564	0.00253
0.125	5000	0.025714	0.0317	0.811	0.976	0.943	1.010	6.39E-05	0.02559	0.02584	0.00248
0.150	5000	0.025790	0.0317	0.814	0.980	0.947	1.014	6.43E-05	0.02566	0.02592	0.00249
0.175	5000	0.025819	0.0317	0.815	0.984	0.951	1.019	6.43E-05	0.02569	0.02595	0.00249
0.200	5000	0.026039	0.0317	0.822	0.989	0.956	1.024	6.38E-05	0.02591	0.02616	0.00245
0.225	5000	0.026254	0.0317	0.828	0.995	0.961	1.030	6.54E-05	0.02613	0.02638	0.00249
0.250	5000	0.026343	0.0317	0.831	1.000	0.966	1.035	6.56E-05	0.02621	0.02647	0.00249
0.275	5000	0.026580	0.0317	0.839	1.005	0.971	1.041	6.42E-05	0.02645	0.02671	0.00241
0.300	5000	0.026623	0.0317	0.840	1.011	0.976	1.046	6.57 E-05	0.02649	0.02675	0.00247
0.325	5000	0.026778	0.0317	0.845	1.016	0.981	1.051	6.71E-05	0.02665	0.02691	0.00251
0.350	5000	0.026745	0.0317	0.844	1.020	0.986	1.056	6.79E-05	0.02661	0.02688	0.00254
0.375	5000	0.026965	0.0317	0.851	1.025	0.990	1.061	6.68E-05	0.02683	0.02710	0.00248
0.400	5000	0.027035	0.0317	0.853	1.028	0.993	1.064	6.72E-05	0.02690	0.02717	0.00248
0.425	5000	0.027097	0.0317	0.855	1.031	0.996	1.067	6.84E-05	0.02696	0.02723	0.00253
0.450	5000	0.027116	0.0317	0.856	1.033	0.998	1.070	6.71E-05	0.02698	0.02725	0.00247
0.475	5000	0.027069	0.0317	0.854	1.035	1.000	1.071	6.82E-05	0.02694	0.02720	0.00252
0.500	5000	0.027280	0.0317	0.861	1.035	1.000	1.071	6.79E-05	0.02715	0.02741	0.00249
0.525	5000	0.027020	0.0317	0.853	1.035	1.000	1.071	6.96E-05	0.02688	0.02716	0.00257
0.550	5000	0.027095	0.0317	0.855	1.033	0.998	1.070	6.87 E-05	0.02696	0.02723	0.00253
0.575	5000	0.026990	0.0317	0.852	1.031	0.996	1.067	6.80E-05	0.02686	0.02712	0.00252
0.600	5000	0.027078	0.0317	0.854	1.028	0.993	1.064	6.78 E- 05	0.02695	0.02721	0.00250
0.625	5000	0.026855	0.0317	0.847	1.025	0.990	1.061	6.82E-05	0.02672	0.02699	0.00254
0.650	5000	0.026811	0.0317	0.846	1.020	0.986	1.056	6.78 E- 05	0.02668	0.02694	0.00253
0.675	5000	0.026697	0.0317	0.842	1.016	0.981	1.051	6.72E-05	0.02657	0.02683	0.00252
0.700	5000	0.026616	0.0317	0.840	1.011	0.976	1.046	6.48E-05	0.02649	0.02674	0.00243
0.725	5000	0.026456	0.0317	0.835	1.005	0.971	1.041	6.62E-05	0.02633	0.02659	0.00250
0.750	5000	0.026376	0.0317	0.832	1.000	0.966	1.035	6.46E-05	0.02625	0.02650	0.00245
0.775	5000	0.026222	0.0317	0.827	0.995	0.961	1.030	6.41E-05	0.02610	0.02635	0.00245
0.800	5000	0.025962	0.0317	0.819	0.989	0.956	1.024	6.52E-05	0.02583	0.02609	0.00251
0.825	5000	0.025856	0.0317	0.816	0.984	0.951	1.019	6.53E-05	0.02573	0.02598	0.00252
0.850	5000	0.025724	0.0317	0.812	0.980	0.947	1.014	6.53E-05	0.02560	0.02585	0.00254
0.875	5000	0.025635	0.0317	0.809	0.976	0.943	1.010	6.60E-05	0.02551	0.02576	0.00257
0.900	5000	0.025539	0.0317	0.806	0.972	0.939	1.007	6.43E-05	0.02541	0.02566	0.00252
0.925	5000	0.025462	0.0317	0.803	0.970	0.937	1.004	6.30E-05	0.02534	0.02559	0.00247
0.950	5000	0.025424	0.0317	0.802	0.968	0.935	1.002	6.36E-05	0.02530	0.02555	0.00250
0.975	5000	0.025442	0.0317	0.803	0.966	0.934	1.000	6.29E-05	0.02532	0.02557	0.00247

**Table A.29:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $(H_2)_t / M / 1$  model as a function of y based on 5,000 replications:  $\gamma = 1, b = 20, \rho = 0.8, \bar{\lambda} = 0.8, \mu = 1, \beta = 0.2, \theta^* = 0.173$
**Table A.30:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $(H_2)_t / M / 1$  model as a function of y based on 5,000 replications:  $\gamma = 0.25, b = 40, \rho = 0.9, \bar{\lambda} = 0.9, \mu = 1, \beta = 0.1, \theta^* = 0.0761$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y/A$ approx	$A_y/A$ LB	$A_y/A$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.042734	0.0477	0.897	0.970	0.941	1.000	4.96E-05	0.04264	0.04283	0.00116
0.025	5000	0.042861	0.0477	0.899	0.970	0.941	1.000	4.79E-05	0.04277	0.04295	0.00112
0.050	5000	0.042816	0.0477	0.898	0.971	0.942	1.001	4.75E-05	0.04272	0.04291	0.00111
0.075	5000	0.042915	0.0477	0.901	0.973	0.944	1.003	4.85E-05	0.04282	0.04301	0.00113
0.100	5000	0.043023	0.0477	0.903	0.976	0.946	1.006	4.79E-05	0.04293	0.04312	0.00111
0.125	5000	0.043116	0.0477	0.905	0.979	0.949	1.009	4.93E-05	0.04302	0.04321	0.00114
0.150	5000	0.043219	0.0477	0.907	0.982	0.953	1.013	4.94E-05	0.04312	0.04332	0.00114
0.175	5000	0.043533	0.0477	0.914	0.986	0.957	1.017	4.85E-05	0.04344	0.04363	0.00111
0.200	5000	0.043765	0.0477	0.918	0.991	0.961	1.021	4.93E-05	0.04367	0.04386	0.00113
0.225	5000	0.043885	0.0477	0.921	0.995	0.965	1.026	4.90E-05	0.04379	0.04398	0.00112
0.250	5000	0.044134	0.0477	0.926	1.000	0.970	1.031	4.91E-05	0.04404	0.04423	0.00111
0.275	5000	0.044318	0.0477	0.930	1.005	0.975	1.036	4.99 E- 05	0.04422	0.04442	0.00113
0.300	5000	0.044483	0.0477	0.933	1.009	0.979	1.041	4.99 E- 05	0.04439	0.04458	0.00112
0.325	5000	0.044729	0.0477	0.939	1.014	0.984	1.045	$4.97\mathrm{E}\text{-}05$	0.04463	0.04483	0.00111
0.350	5000	0.044932	0.0477	0.943	1.018	0.988	1.050	5.00E-05	0.04483	0.04503	0.00111
0.375	5000	0.045040	0.0477	0.945	1.022	0.991	1.053	4.98E-05	0.04494	0.04514	0.00110
0.400	5000	0.045175	0.0477	0.948	1.025	0.994	1.057	5.12E-05	0.04507	0.04528	0.00113
0.425	5000	0.045244	0.0477	0.949	1.027	0.997	1.059	5.07 E-05	0.04514	0.04534	0.00112
0.450	5000	0.045360	0.0477	0.952	1.029	0.999	1.061	5.19E-05	0.04526	0.04546	0.00114
0.475	5000	0.045519	0.0477	0.955	1.031	1.000	1.062	5.07 E-05	0.04542	0.04562	0.00111
0.500	5000	0.045536	0.0477	0.956	1.031	1.000	1.063	5.00E-05	0.04544	0.04563	0.00110
0.525	5000	0.045435	0.0477	0.953	1.031	1.000	1.062	5.13E-05	0.04533	0.04554	0.00113
0.550	5000	0.045563	0.0477	0.956	1.029	0.999	1.061	4.95E-05	0.04547	0.04566	0.00109
0.575	5000	0.045329	0.0477	0.951	1.027	0.997	1.059	5.08E-05	0.04523	0.04543	0.00112
0.600	5000	0.045185	0.0477	0.948	1.025	0.994	1.057	5.07 E-05	0.04509	0.04528	0.00112
0.625	5000	0.045032	0.0477	0.945	1.022	0.991	1.053	5.11E-05	0.04493	0.04513	0.00113
0.650	5000	0.044887	0.0477	0.942	1.018	0.988	1.050	5.12E-05	0.04479	0.04499	0.00114
0.675	5000	0.044731	0.0477	0.939	1.014	0.984	1.045	4.90E-05	0.04463	0.04483	0.00110
0.700	5000	0.044457	0.0477	0.933	1.009	0.979	1.041	5.14E-05	0.04436	0.04456	0.00116
0.725	5000	0.044321	0.0477	0.930	1.005	0.975	1.036	4.92E-05	0.04422	0.04442	0.00111
0.750	5000	0.044170	0.0477	0.927	1.000	0.970	1.031	4.93E-05	0.04407	0.04427	0.00112
0.775	5000	0.043813	0.0477	0.919	0.995	0.965	1.026	5.05E-05	0.04371	0.04391	0.00115
0.800	5000	0.043666	0.0477	0.916	0.991	0.961	1.021	4.94E-05	0.04357	0.04376	0.00113
0.825	5000	0.043504	0.0477	0.913	0.986	0.957	1.017	4.80E-05	0.04341	0.04360	0.00110
0.850	5000	0.043330	0.0477	0.909	0.982	0.953	1.013	4.91E-05	0.04323	0.04343	0.00113
0.875	5000	0.043244	0.0477	0.907	0.979	0.949	1.009	4.73E-05	0.04315	0.04334	0.00109
0.900	5000	0.043098	0.0477	0.904	0.976	0.946	1.006	4.82E-05	0.04300	0.04319	0.00112
0.925	5000	0.042836	0.0477	0.899	0.973	0.944	1.003	4.91E-05	0.04274	0.04293	0.00115
0.950	5000	0.042714	0.0477	0.896	0.971	0.942	1.001	4.87E-05	0.04262	0.04281	0.00114
0.975	5000	0.042777	0.0477	0.898	0.970	0.941	1.000	4.81E-05	0.04268	0.04287	0.00112

**Table A.31:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $(H_2)_t / M / 1$  model as a function of y based on 5,000 replications:  $\gamma = \frac{1}{16}, b = 80, \rho = 0.95, \bar{\lambda} = 0.95, \mu = 1, \beta = 0.05, \theta^* = 0.0356$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y/A$ approx	$A_y/A$ LB	$A_y/A$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.054303	0.0578	0.939	0.972	0.945	1.000	3.13E-05	0.05424	0.05436	5.77E-04
0.025	5000	0.054300	0.0578	0.939	0.972	0.945	1.000	3.15E-05	0.05424	0.05436	5.81E-04
0.050	5000	0.054360	0.0578	0.940	0.973	0.946	1.001	3.12E-05	0.05430	0.05442	5.74E-04
0.075	5000	0.054519	0.0578	0.943	0.975	0.948	1.003	3.09E-05	0.05446	0.05458	5.66E-04
0.100	5000	0.054550	0.0578	0.943	0.977	0.950	1.005	3.18E-05	0.05449	0.05461	5.83E-04
0.125	5000	0.054786	0.0578	0.947	0.980	0.953	1.008	3.16E-05	0.05472	0.05485	5.76E-04
0.150	5000	0.054931	0.0578	0.950	0.983	0.956	1.012	3.16E-05	0.05487	0.05499	5.76E-04
0.175	5000	0.055150	0.0578	0.954	0.987	0.959	1.016	3.15E-05	0.05509	0.05521	5.71E-04
0.200	5000	0.055383	0.0578	0.958	0.991	0.963	1.020	3.21E-05	0.05532	0.05545	5.79E-04
0.225	5000	0.055608	0.0578	0.961	0.996	0.968	1.024	3.18E-05	0.05555	0.05567	5.72E-04
0.250	5000	0.055865	0.0578	0.966	1.000	0.972	1.029	3.24E-05	0.05580	0.05593	5.80E-04
0.275	5000	0.056159	0.0578	0.971	1.004	0.976	1.034	3.27E-05	0.05610	0.05622	5.82E-04
0.300	5000	0.056380	0.0578	0.975	1.009	0.980	1.038	3.23E-05	0.05632	0.05644	5.73E-04
0.325	5000	0.056566	0.0578	0.978	1.013	0.985	1.042	3.30E-05	0.05650	0.05663	5.83E-04
0.350	5000	0.056841	0.0578	0.983	1.017	0.988	1.046	3.26E-05	0.05678	0.05691	5.74E-04
0.375	5000	0.056992	0.0578	0.985	1.020	0.992	1.050	3.39E-05	0.05693	0.05706	5.95E-04
0.400	5000	0.057139	0.0578	0.988	1.023	0.995	1.053	3.27E-05	0.05708	0.05720	5.72E-04
0.425	5000	0.057324	0.0578	0.991	1.026	0.997	1.055	3.36E-05	0.05726	0.05739	5.86E-04
0.450	5000	0.057391	0.0578	0.992	1.027	0.999	1.057	3.35E-05	0.05733	0.05746	5.83E-04
0.475	5000	0.057464	0.0578	0.994	1.029	1.000	1.058	3.32E-05	0.05740	0.05753	5.77E-04
0.500	5000	0.057466	0.0578	0.994	1.029	1.000	1.059	3.29E-05	0.05740	0.05753	5.73E-04
0.525	5000	0.057460	0.0578	0.993	1.029	1.000	1.058	3.34E-05	0.05739	0.05753	5.82E-04
0.550	5000	0.057412	0.0578	0.993	1.027	0.999	1.057	3.32E-05	0.05735	0.05748	5.79E-04
0.575	5000	0.057343	0.0578	0.991	1.026	0.997	1.055	3.30E-05	0.05728	0.05741	5.75E-04
0.600	5000	0.057165	0.0578	0.988	1.023	0.995	1.053	3.33E-05	0.05710	0.05723	5.83E-04
0.625	5000	0.057041	0.0578	0.986	1.020	0.992	1.050	3.35E-05	0.05697	0.05711	5.87E-04
0.650	5000	0.056788	0.0578	0.982	1.017	0.988	1.046	3.20E-05	0.05673	0.05685	5.63E-04
0.675	5000	0.056650	0.0578	0.979	1.013	0.985	1.042	3.20E-05	0.05659	0.05671	5.65E-04
0.700	5000	0.056391	0.0578	0.975	1.009	0.980	1.038	3.29E-05	0.05633	0.05646	5.84E-04
0.725	5000	0.056128	0.0578	0.970	1.004	0.976	1.034	3.23E-05	0.05606	0.05619	5.76E-04
0.750	5000	0.055929	0.0578	0.967	1.000	0.972	1.029	3.15E-05	0.05587	0.05599	5.62E-04
0.775	5000	0.055577	0.0578	0.961	0.996	0.968	1.024	3.30E-05	0.05551	0.05564	5.94E-04
0.800	5000	0.055409	0.0578	0.958	0.991	0.963	1.020	3.15E-05	0.05535	0.05547	5.68E-04
0.825	5000	0.055163	0.0578	0.954	0.987	0.959	1.016	3.18E-05	0.05510	0.05523	5.76E-04
0.850	5000	0.054896	0.0578	0.949	0.983	0.956	1.012	3.20E-05	0.05483	0.05496	5.84E-04
0.875	5000	0.054714	0.0578	0.946	0.980	0.953	1.008	3.15E-05	0.05465	0.05478	5.76E-04
0.900	5000	0.054613	0.0578	0.944	0.977	0.950	1.005	3.16E-05	0.05455	0.05467	5.79E-04
0.925	5000	0.054457	0.0578	0.942	0.975	0.948	1.003	3.25E-05	0.05439	0.05452	5.96E-04
0.950	5000	0.054428	0.0578	0.941	0.973	0.946	1.001	3.22E-05	0.05437	0.05449	5.91E-04
0.975	5000	0.054358	0.0578	0.940	0.972	0.945	1.000	3.12E-05	0.05430	0.05442	5.75E-04

**Table A.32:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $(H_2)_t / M / 1$  model as a function of y based on 5,000 replications:  $\gamma = 0.64, b = 25, \rho = 0.84, \bar{\lambda} = 0.84, \mu = 1, \beta = 0.16, \theta^* = 0.131$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.031500	0.0374	0.842	0.842	0.815	0.870	5.96E-05	0.03138	0.03162	0.00189
0.025	5000	0.031531	0.0374	0.843	0.843	0.815	0.871	6.05E-05	0.03141	0.03165	0.00192
0.050	5000	0.031537	0.0374	0.843	0.844	0.816	0.872	6.04E-05	0.03142	0.03166	0.00192
0.075	5000	0.031558	0.0374	0.844	0.845	0.818	0.873	5.96E-05	0.03144	0.03167	0.00189
0.100	5000	0.031603	0.0374	0.845	0.848	0.820	0.876	5.99 E- 05	0.03149	0.03172	0.00190
0.125	5000	0.031834	0.0374	0.851	0.850	0.823	0.879	6.01E-05	0.03172	0.03195	0.00189
0.150	5000	0.031906	0.0374	0.853	0.854	0.826	0.882	6.11E-05	0.03179	0.03203	0.00191
0.175	5000	0.032097	0.0374	0.858	0.857	0.830	0.886	6.08E-05	0.03198	0.03222	0.00189
0.200	5000	0.032215	0.0374	0.862	0.862	0.834	0.890	6.17E-05	0.03209	0.03234	0.00192
0.225	5000	0.032294	0.0374	0.864	0.866	0.838	0.895	6.19E-05	0.03217	0.03242	0.00192
0.250	5000	0.032581	0.0374	0.871	0.870	0.842	0.899	6.10E-05	0.03246	0.03270	0.00187
0.275	5000	0.032785	0.0374	0.877	0.875	0.847	0.904	6.30E-05	0.03266	0.03291	0.00192
0.300	5000	0.032967	0.0374	0.882	0.879	0.851	0.909	6.17E-05	0.03285	0.03309	0.00187
0.325	5000	0.033044	0.0374	0.884	0.883	0.855	0.913	6.23E-05	0.03292	0.03317	0.00189
0.350	5000	0.033178	0.0374	0.887	0.887	0.859	0.917	6.31E-05	0.03305	0.03330	0.00190
0.375	5000	0.033343	0.0374	0.892	0.891	0.862	0.921	6.36E-05	0.03322	0.03347	0.00191
0.400	5000	0.033383	0.0374	0.893	0.894	0.865	0.924	6.31E-05	0.03326	0.03351	0.00189
0.425	5000	0.033410	0.0374	0.894	0.896	0.867	0.926	6.42E-05	0.03328	0.03354	0.00192
0.450	5000	0.033533	0.0374	0.897	0.898	0.869	0.928	6.36E-05	0.03341	0.03366	0.00190
0.475	5000	0.033599	0.0374	0.899	0.899	0.870	0.929	6.42E-05	0.03347	0.03372	0.00191
0.500	5000	0.033561	0.0374	0.898	0.899	0.870	0.929	6.35E-05	0.03344	0.03369	0.00189
0.525	5000	0.033613	0.0374	0.899	0.899	0.870	0.929	6.43E-05	0.03349	0.03374	0.00191
0.550	5000	0.033546	0.0374	0.897	0.898	0.869	0.928	6.49E-05	0.03342	0.03367	0.00193
0.575	5000	0.033541	0.0374	0.897	0.896	0.867	0.926	6.26E-05	0.03342	0.03366	0.00187
0.600	5000	0.033365	0.0374	0.892	0.894	0.865	0.924	6.31E-05	0.03324	0.03349	0.00189
0.625	5000	0.033284	0.0374	0.890	0.891	0.862	0.921	6.47E-05	0.03316	0.03341	0.00194
0.650	5000	0.033196	0.0374	0.888	0.887	0.859	0.917	6.38E-05	0.03307	0.03332	0.00192
0.675	5000	0.033028	0.0374	0.883	0.883	0.855	0.913	6.36E-05	0.03290	0.03315	0.00193
0.700	5000	0.032913	0.0374	0.880	0.879	0.851	0.909	6.09E-05	0.03279	0.03303	0.00185
0.725	5000	0.032711	0.0374	0.875	0.875	0.847	0.904	6.35E-05	0.03259	0.03284	0.00194
0.750	5000	0.032600	0.0374	0.872	0.870	0.842	0.899	6.05E-05	0.03248	0.03272	0.00186
0.775	5000	0.032468	0.0374	0.868	0.866	0.838	0.895	6.07 E-05	0.03235	0.03259	0.00187
0.800	5000	0.032151	0.0374	0.860	0.862	0.834	0.890	6.25E-05	0.03203	0.03227	0.00194
0.825	5000	0.032071	0.0374	0.858	0.857	0.830	0.886	6.10E-05	0.03195	0.03219	0.00190
0.850	5000	0.031876	0.0374	0.852	0.854	0.826	0.882	6.17E-05	0.03175	0.03200	0.00193
0.875	5000	0.031610	0.0374	0.845	0.850	0.823	0.879	6.17E-05	0.03149	0.03173	0.00195
0.900	5000	0.031681	0.0374	0.847	0.848	0.820	0.876	5.95E-05	0.03156	0.03180	0.00188
0.925	5000	0.031634	0.0374	0.846	0.845	0.818	0.873	5.98E-05	0.03152	0.03175	0.00189
0.950	5000	0.031516	0.0374	0.843	0.844	0.816	0.872	6.01E-05	0.03140	0.03163	0.00191
0.975	5000	0.031469	0.0374	0.842	0.843	0.815	0.871	5.99E-05	0.03135	0.03159	0.00190

**Table A.33:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $(H_2)_t / M / 1$  model as a function of y based on 5,000 replications:  $\gamma = 0.16, b = 50, \rho = 0.92, \bar{\lambda} = 0.92, \mu = 1, \beta = 0.08, \theta^* = 0.0593$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
0.000	5000	0.047148	0.0515	0.915	0.913	0.887	0.941	4.20E-05	0.04707	0.04723	8.92E-04
0.025	5000	0.047196	0.0515	0.916	0.914	0.887	0.941	4.21E-05	0.04711	0.04728	8.91E-04
0.050	5000	0.047178	0.0515	0.915	0.915	0.888	0.942	4.21E-05	0.04710	0.04726	8.92E-04
0.075	5000	0.047238	0.0515	0.916	0.916	0.890	0.944	4.29E-05	0.04715	0.04732	9.07E-04
0.100	5000	0.047472	0.0515	0.921	0.919	0.892	0.946	9.77 E-05	0.04728	0.04766	2.06E-03
0.125	5000	0.047550	0.0515	0.922	0.921	0.894	0.949	4.30E-05	0.04747	0.04763	9.04E-04
0.150	5000	0.047720	0.0515	0.926	0.925	0.898	0.952	4.24E-05	0.04764	0.04780	8.89E-04
0.175	5000	0.048040	0.0515	0.932	0.928	0.901	0.956	8.37 E-05	0.04788	0.04820	1.74E-03
0.200	5000	0.048026	0.0515	0.932	0.932	0.905	0.960	4.39E-05	0.04794	0.04811	9.15E-04
0.225	5000	0.048282	0.0515	0.937	0.936	0.909	0.965	4.40E-05	0.04820	0.04837	9.11E-04
0.250	5000	0.048515	0.0515	0.941	0.941	0.913	0.969	4.47E-05	0.04843	0.04860	9.21E-04
0.275	5000	0.048691	0.0515	0.945	0.945	0.918	0.974	4.48E-05	0.04860	0.04878	9.20E-04
0.300	5000	0.048982	0.0515	0.950	0.950	0.922	0.978	4.54E-05	0.04889	0.04907	9.28E-04
0.325	5000	0.049243	0.0515	0.955	0.954	0.926	0.982	4.34E-05	0.04916	0.04933	8.81E-04
0.350	5000	0.049461	0.0515	0.960	0.957	0.929	0.986	4.57E-05	0.04937	0.04955	9.24E-04
0.375	5000	0.049662	0.0515	0.963	0.961	0.933	0.990	4.31E-05	0.04958	0.04975	8.68E-04
0.400	5000	0.049779	0.0515	0.966	0.964	0.936	0.993	4.39E-05	0.04969	0.04987	8.83E-04
0.425	5000	0.050002	0.0515	0.970	0.966	0.938	0.995	7.26E-05	0.04986	0.05014	1.45E-03
0.450	5000	0.049994	0.0515	0.970	0.968	0.939	0.997	4.41E-05	0.04991	0.05008	8.82E-04
0.475	5000	0.049949	0.0515	0.969	0.969	0.941	0.998	4.50E-05	0.04986	0.05004	9.01E-04
0.500	5000	0.050020	0.0515	0.970	0.969	0.941	0.998	4.44E-05	0.04993	0.05011	8.88E-04
0.525	5000	0.050090	0.0515	0.972	0.969	0.941	0.998	4.49E-05	0.05000	0.05018	8.97E-04
0.550	5000	0.050077	0.0515	0.971	0.968	0.939	0.997	4.31E-05	0.04999	0.05016	8.60E-04
0.575	5000	0.049931	0.0515	0.969	0.966	0.938	0.995	4.36E-05	0.04985	0.05002	8.72E-04
0.600	5000	0.049756	0.0515	0.965	0.964	0.936	0.993	4.45E-05	0.04967	0.04984	8.95E-04
0.625	5000	0.049611	0.0515	0.962	0.961	0.933	0.990	4.33E-05	0.04953	0.04970	8.72E-04
0.650	5000	0.049456	0.0515	0.959	0.957	0.929	0.986	4.43E-05	0.04937	0.04954	8.96E-04
0.675	5000	0.049202	0.0515	0.954	0.954	0.926	0.982	4.41E-05	0.04912	0.04929	8.95E-04
0.700	5000	0.048966	0.0515	0.950	0.950	0.922	0.978	4.47E-05	0.04888	0.04905	9.12E-04
0.725	5000	0.048780	0.0515	0.946	0.945	0.918	0.974	4.40E-05	0.04869	0.04887	9.02E-04
0.750	5000	0.048635	0.0515	0.944	0.941	0.913	0.969	4.28E-05	0.04855	0.04872	8.80E-04
0.775	5000	0.048339	0.0515	0.938	0.936	0.909	0.965	4.29E-05	0.04826	0.04842	8.88E-04
0.800	5000	0.048207	0.0515	0.935	0.932	0.905	0.960	4.21E-05	0.04812	0.04829	8.72E-04
0.825	5000	0.047963	0.0515	0.930	0.928	0.901	0.956	4.17E-05	0.04788	0.04804	8.69E-04
0.850	5000	0.047699	0.0515	0.925	0.925	0.898	0.952	4.32E-05	0.04761	0.04778	9.05E-04
0.875	5000	0.047584	0.0515	0.923	0.921	0.894	0.949	4.17E-05	0.04750	0.04767	8.77E-04
0.900	5000	0.047438	0.0515	0.920	0.919	0.892	0.946	4.14E-05	0.04736	0.04752	8.73E-04
0.925	5000	0.047343	0.0515	0.918	0.916	0.890	0.944	4.15E-05	0.04726	0.04742	8.76E-04
0.950	5000	0.047215	0.0515	0.916	0.915	0.888	0.942	4.18E-05	0.04713	0.04730	8.86E-04
0.975	5000	0.047194	0.0515	0.916	0.914	0.887	0.941	4.18E-05	0.04711	0.04728	8.87E-04

**Table A.34:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $(H_2)_t / M / 1$  model as a function of y based on 5,000 replications:  $\gamma = 0.01, b = 200, \rho = 0.98, \bar{\lambda} = 0.98, \mu = 1, \beta = 0.02, \theta^* = 0.0137$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y/A$ approx	$A_y/A$ LB	$A_y/A$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.062152	0.0647	0.961	0.973	0.947	1.000	2.09E-05	0.06211	0.06219	3.36E-04
0.025	5000	0.062156	0.0647	0.961	0.973	0.947	1.000	2.05E-05	0.06212	0.06220	3.30E-04
0.050	5000	0.062226	0.0647	0.962	0.974	0.948	1.001	2.09E-05	0.06218	0.06227	3.37E-04
0.075	5000	0.062304	0.0647	0.964	0.976	0.950	1.003	2.14E-05	0.06226	0.06235	3.43E-04
0.100	5000	0.062430	0.0647	0.966	0.978	0.952	1.005	2.10E-05	0.06239	0.06247	3.36E-04
0.125	5000	0.062626	0.0647	0.969	0.981	0.954	1.008	2.08E-05	0.06258	0.06267	3.32E-04
0.150	5000	0.062851	0.0647	0.972	0.984	0.957	1.011	2.08E-05	0.06281	0.06289	3.30E-04
0.175	5000	0.063082	0.0647	0.976	0.988	0.961	1.015	2.11E-05	0.06304	0.06312	3.34E-04
0.200	5000	0.063343	0.0647	0.980	0.992	0.965	1.019	2.11E-05	0.06330	0.06338	3.32E-04
0.225	5000	0.063591	0.0647	0.984	0.996	0.969	1.023	2.11E-05	0.06355	0.06363	3.32E-04
0.250	5000	0.063883	0.0647	0.988	1.000	0.973	1.028	2.14E-05	0.06384	0.06393	3.35E-04
0.275	5000	0.064143	0.0647	0.992	1.004	0.977	1.032	2.14E-05	0.06410	0.06418	3.34E-04
0.300	5000	0.064370	0.0647	0.996	1.008	0.981	1.037	2.15E-05	0.06433	0.06441	3.35E-04
0.325	5000	0.064690	0.0647	1.001	1.013	0.985	1.041	2.17E-05	0.06465	0.06473	3.35E-04
0.350	5000	0.064920	0.0647	1.004	1.016	0.989	1.044	2.18E-05	0.06488	0.06496	3.36E-04
0.375	5000	0.065129	0.0647	1.007	1.020	0.992	1.048	2.16E-05	0.06509	0.06517	3.32E-04
0.400	5000	0.065284	0.0647	1.010	1.022	0.995	1.051	2.21E-05	0.06524	0.06533	3.38E-04
0.425	5000	0.065469	0.0647	1.013	1.025	0.997	1.053	2.17E-05	0.06543	0.06551	3.31E-04
0.450	5000	0.065561	0.0647	1.014	1.026	0.999	1.055	2.19E-05	0.06552	0.06560	3.34E-04
0.475	5000	0.065605	0.0647	1.015	1.027	1.000	1.056	2.17E-05	0.06556	0.06565	3.31E-04
0.500	5000	0.065625	0.0647	1.015	1.028	1.000	1.056	2.18E-05	0.06558	0.06567	3.32E-04
0.525	5000	0.065597	0.0647	1.015	1.027	1.000	1.056	2.25E-05	0.06555	0.06564	3.42E-04
0.550	5000	0.065522	0.0647	1.013	1.026	0.999	1.055	2.17E-05	0.06548	0.06556	3.31E-04
0.575	5000	0.065497	0.0647	1.013	1.025	0.997	1.053	2.19E-05	0.06545	0.06554	3.35E-04
0.600	5000	0.065314	0.0647	1.010	1.022	0.995	1.051	2.23E-05	0.06527	0.06536	3.41E-04
0.625	5000	0.065144	0.0647	1.008	1.020	0.992	1.048	2.18E-05	0.06510	0.06519	3.35E-04
0.650	5000	0.064897	0.0647	1.004	1.016	0.989	1.044	2.20E-05	0.06485	0.06494	3.39E-04
0.675	5000	0.064678	0.0647	1.000	1.013	0.985	1.041	2.17E-05	0.06464	0.06472	3.36E-04
0.700	5000	0.064436	0.0647	0.997	1.008	0.981	1.037	2.13E-05	0.06439	0.06448	3.30E-04
0.725	5000	0.064149	0.0647	0.992	1.004	0.977	1.032	2.15E-05	0.06411	0.06419	3.36E-04
0.750	5000	0.063882	0.0647	0.988	1.000	0.973	1.028	2.12E-05	0.06384	0.06392	3.32E-04
0.775	5000	0.063605	0.0647	0.984	0.996	0.969	1.023	2.16E-05	0.06356	0.06365	3.39E-04
0.800	5000	0.063313	0.0647	0.979	0.992	0.965	1.019	2.10E-05	0.06327	0.06335	3.32E-04
0.825	5000	0.063053	0.0647	0.975	0.988	0.961	1.015	2.08E-05	0.06301	0.06309	3.30E-04
0.850	5000	0.062886	0.0647	0.973	0.984	0.957	1.011	2.09E-05	0.06285	0.06293	3.32E-04
0.875	5000	0.062639	0.0647	0.969	0.981	0.954	1.008	2.05E-05	0.06260	0.06268	3.27E-04
0.900	5000	0.062504	0.0647	0.967	0.978	0.952	1.005	2.07E-05	0.06246	0.06254	3.30E-04
0.925	5000	0.062342	0.0647	0.964	0.976	0.950	1.003	2.10E-05	0.06230	0.06238	3.36E-04
0.950	5000	0.062243	0.0647	0.963	0.974	0.948	1.001	2.05E-05	0.06220	0.06228	3.29E-04
0.975	5000	0.062175	0.0647	0.962	0.973	0.947	1.000	2.02E-05	0.06214	0.06221	3.25E-04

**Table A.35:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $(H_2)_t / M / 1$  model as a function of y based on 5,000 replications:  $\gamma = 0.0025, b = 400, \rho = 0.99, \bar{\lambda} = 0.99, \mu = 1, \beta = 0.01, \theta^* = 0.00676$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y/A$ approx	$A_y/A$ LB	$A_y/A$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.064912	0.0670	0.968	0.973	0.947	1.000	1.88E-05	0.06488	0.06495	2.90E-04
0.025	5000	0.064916	0.0670	0.968	0.974	0.948	1.000	1.86E-05	0.06488	0.06495	2.86E-04
0.050	5000	0.064997	0.0670	0.970	0.975	0.949	1.001	1.86E-05	0.06496	0.06503	2.87E-04
0.075	5000	0.065107	0.0670	0.971	0.976	0.950	1.003	1.86E-05	0.06507	0.06514	2.86E-04
0.100	5000	0.065259	0.0670	0.973	0.978	0.952	1.005	1.86E-05	0.06522	0.06529	2.85E-04
0.125	5000	0.065414	0.0670	0.976	0.981	0.955	1.008	1.85E-05	0.06538	0.06545	2.83E-04
0.150	5000	0.065616	0.0670	0.979	0.984	0.958	1.011	1.88E-05	0.06558	0.06565	2.86E-04
0.175	5000	0.065908	0.0670	0.983	0.988	0.961	1.015	1.88E-05	0.06587	0.06594	2.85 E-04
0.200	5000	0.066145	0.0670	0.987	0.992	0.965	1.019	1.89E-05	0.06611	0.06618	2.86E-04
0.225	5000	0.066400	0.0670	0.990	0.996	0.969	1.023	1.91E-05	0.06636	0.06644	2.88E-04
0.250	5000	0.066691	0.0670	0.995	1.000	0.973	1.027	1.89E-05	0.06665	0.06673	2.83E-04
0.275	5000	0.067005	0.0670	0.999	1.004	0.977	1.032	1.93E-05	0.06697	0.06704	2.88E-04
0.300	5000	0.067244	0.0670	1.003	1.008	0.981	1.036	1.91E-05	0.06721	0.06728	2.84E-04
0.325	5000	0.067532	0.0670	1.007	1.012	0.985	1.040	1.94E-05	0.06749	0.06757	2.87E-04
0.350	5000	0.067757	0.0670	1.011	1.016	0.989	1.044	1.93E-05	0.06772	0.06780	2.86E-04
0.375	5000	0.067990	0.0670	1.014	1.019	0.992	1.047	1.95E-05	0.06795	0.06803	2.86E-04
0.400	5000	0.068184	0.0670	1.017	1.022	0.995	1.050	1.93E-05	0.06815	0.06822	2.84E-04
0.425	5000	0.068312	0.0670	1.019	1.024	0.997	1.052	1.96E-05	0.06827	0.06835	2.86E-04
0.450	5000	0.068399	0.0670	1.020	1.026	0.999	1.054	1.95E-05	0.06836	0.06844	2.86E-04
0.475	5000	0.068526	0.0670	1.022	1.027	1.000	1.055	1.96E-05	0.06849	0.06856	2.86E-04
0.500	5000	0.068541	0.0670	1.022	1.027	1.000	1.056	1.96E-05	0.06850	0.06858	2.86E-04
0.525	5000	0.068500	0.0670	1.022	1.027	1.000	1.055	1.96E-05	0.06846	0.06854	2.86E-04
0.550	5000	0.068422	0.0670	1.021	1.026	0.999	1.054	1.95E-05	0.06838	0.06846	2.85E-04
0.575	5000	0.068339	0.0670	1.019	1.024	0.997	1.052	1.94E-05	0.06830	0.06838	2.84E-04
0.600	5000	0.068154	0.0670	1.017	1.022	0.995	1.050	1.95E-05	0.06812	0.06819	2.86E-04
0.625	5000	0.068002	0.0670	1.014	1.019	0.992	1.047	1.95E-05	0.06796	0.06804	2.87E-04
0.650	5000	0.067769	0.0670	1.011	1.016	0.989	1.044	1.93E-05	0.06773	0.06781	2.85E-04
0.675	5000	0.067519	0.0670	1.007	1.012	0.985	1.040	1.91E-05	0.06748	0.06756	2.83E-04
0.700	5000	0.067221	0.0670	1.003	1.008	0.981	1.036	1.94E-05	0.06718	0.06726	2.88E-04
0.725	5000	0.066975	0.0670	0.999	1.004	0.977	1.032	1.92E-05	0.06694	0.06701	2.86E-04
0.750	5000	0.066729	0.0670	0.995	1.000	0.973	1.027	1.92E-05	0.06669	0.06677	2.88E-04
0.775	5000	0.066409	0.0670	0.991	0.996	0.969	1.023	1.89E-05	0.06637	0.06645	2.84E-04
0.800	5000	0.066165	0.0670	0.987	0.992	0.965	1.019	1.86E-05	0.06613	0.06620	2.81E-04
0.825	5000	0.065873	0.0670	0.983	0.988	0.961	1.015	1.89E-05	0.06584	0.06591	2.87E-04
0.850	5000	0.065640	0.0670	0.979	0.984	0.958	1.011	1.87E-05	0.06560	0.06568	2.85E-04
0.875	5000	0.065434	0.0670	0.976	0.981	0.955	1.008	1.85E-05	0.06540	0.06547	2.83E-04
0.900	5000	0.065241	0.0670	0.973	0.978	0.952	1.005	1.85E-05	0.06521	0.06528	2.83E-04
0.925	5000	0.065099	0.0670	0.971	0.976	0.950	1.003	1.86E-05	0.06506	0.06514	2.85E-04
0.950	5000	0.064994	0.0670	0.969	0.975	0.949	1.001	1.87E-05	0.06496	0.06503	2.87E-04
0.975	5000	0.064916	0.0670	0.968	0.974	0.948	1.000	1.84E-05	0.06488	0.06495	2.84E-04

position	ho = 0.8	$\rho=0.9$	$\rho=0.95$	$\rho=0.98$	$\rho=0.99$
0.000	0.99885	0.99642	0.98833	0.98093	0.97804
0.025	0.99648	0.99938	0.98828	0.98099	0.97809
0.050	1.00011	0.99832	0.98936	0.98209	0.97932
0.075	1.00585	1.00064	0.99226	0.98332	0.98096
0.100	1.00635	1.00316	0.99283	0.98532	0.98325
0.125	1.01416	1.00532	0.99712	0.98840	0.98560
0.150	1.01717	1.00772	0.99975	0.99195	0.98864
0.175	1.01831	1.01505	1.00374	0.99561	0.99304
0.200	1.02699	1.02045	1.00798	0.99972	0.99660
0.225	1.03546	1.02326	1.01207	1.00363	1.00044
0.250	1.03895	1.02906	1.01676	1.00825	1.00483
0.275	1.04831	1.03335	1.02211	1.01234	1.00956
0.300	1.05001	1.03720	1.02614	1.01593	1.01317
0.325	1.05613	1.04294	1.02952	1.02099	1.01750
0.350	1.05483	1.04767	1.03452	1.02461	1.02090
0.375	1.06350	1.05019	1.03727	1.02791	1.02441
0.400	1.06624	1.05333	1.03995	1.03035	1.02733
0.425	1.06871	1.05493	1.04330	1.03327	1.02927
0.450	1.06943	1.05766	1.04453	1.03473	1.03056
0.475	1.06759	1.06136	1.04585	1.03542	1.03248
0.500	1.07590	1.06174	1.04589	1.03574	1.03270
0.525	1.06565	1.05938	1.04578	1.03530	1.03209
0.550	1.06860	1.06238	1.04491	1.03412	1.03092
0.575	1.06446	1.05693	1.04366	1.03372	1.02966
0.600	1.06795	1.05356	1.04041	1.03083	1.02688
0.625	1.05916	1.05001	1.03815	1.02815	1.02458
0.650	1.05741	1.04661	1.03355	1.02425	1.02107
0.675	1.05293	1.04298	1.03105	1.02080	1.01731
0.700	1.04973	1.03660	1.02633	1.01698	1.01282
0.725	1.04343	1.03342	1.02154	1.01245	1.00912
0.750	1.04025	1.02989	1.01792	1.00823	1.00540
0.775	1.03420	1.02157	1.01150	1.00387	1.00059
0.800	1.02393	1.01814	1.00845	0.99925	0.99691
0.825	1.01977	1.01437	1.00398	0.99515	0.99251
0.850	1.01453	1.01032	0.99913	0.99251	0.98901
0.875	1.01105	1.00830	0.99581	0.98861	0.98590
0.900	1.00724	1.00491	0.99396	0.98648	0.98299
0.925	1.00420	0.99879	0.99113	0.98393	0.98085
0.950	1.00273	0.99595	0.99061	0.98237	0.97926
0.975	1.00342	0.99743	0.98932	0.98129	0.97809
avg diff w.r.t. last column	0.03168	0.02345	0.01205	0.00318	0.00000
avg. abs. diff w.r.t. last column	0.03168	0.02345	0.01205	0.00318	0.00000
rmse w.r.t. last column	0.03234	0.02369	0.01213	0.00322	0.00000

**Table A.36:** Comparison of ratio  $P(W_y > b)/\rho$  in the  $(H_2)_t/M/1$  queue as a function of  $\rho$  with base parameter  $(\beta, \gamma, b) = (1, 25, 4)$  using the scaling in (3.39) of the main thesis.

**Table A.37:** Summary of simulation results for the  $(H_2)_t/M/1$  queue at y = 0 as a function of  $1 - \rho$  with base parameter  $(\beta, \gamma, b) = (1, 25, 4)$  using the scaling in (3.39) of the main thesis.

	$1 - \rho = 0.16$	$1-\rho=0.08$	$1-\rho=0.04$	$1-\rho=0.02$	$1-\rho=0.01$	$1 - \rho = 0.005$
$\theta^*$	0.113	0.0548	0.0270	0.0134	0.00669	0.00334
n	5000	5000	5000	5000	5000	5000
$\hat{p}$	0.051165	0.059299	0.063514	0.065607	0.066689	0.067212
$e^{-\theta^* b}$	0.0593	0.0645	0.0670	0.0682	0.0689	0.0692
$A_y$	0.862	0.920	0.948	0.961	0.968	0.972
$A_y$ approxi	0.861	0.919	0.947	0.960	0.967	0.970
$A_y$ LB	0.837	0.895	0.922	0.935	0.942	0.945
$A_y$ UB	0.885	0.945	0.973	0.987	0.993	0.997
s.e.	$8.57 \text{E}{-}05$	5.04E-05	2.97 E- 05	2.15E-05	1.89E-05	1.82E-05
95% CI (lb)	0.05100	0.05920	0.06346	0.06557	0.06665	0.06718
(ub)	0.05133	0.05940	0.06357	0.06565	0.06673	0.06725
r.e.	0.001675	0.000849	0.000467	0.000327	0.000284	0.000271
$P(W_y > b) / P(W > b)$						
ratio	0.97418	0.97338	0.97468	0.97445	0.97493	0.97491
diff w.r.t. last column	0.00074	0.00153	0.00023	0.00046	-0.00002	0.00000
abs diff w.r.t. last column	0.00074	0.00153	0.00023	0.00046	0.00002	0.00000
$A_y/ ho$						
ratio	1.02676	0.99988	0.98758	0.98100	0.97819	0.97652
diff w.r.t. last column	-0.05024	-0.02336	-0.01106	-0.00448	-0.00167	0.00000
abs diff w.r.t. last column	0.05024	0.02336	0.01106	0.00448	0.00167	0.00000

**Table A.38:** Summary of simulation results for the  $(H_2)_t/M/1$  queue at y = 0 and y = 0.5 as a function of  $1 - \rho$  with base parameter  $(\beta, \gamma, b) = (1, 2.5, 4)$  using the scaling in (3.39) of the main thesis.

	$1 - \rho = 0.16$	$1-\rho=0.08$	$1-\rho=0.04$	$1-\rho=0.02$	$1-\rho=0.01$
$theta^*$	0.113	0.0548	0.0270	0.0134	0.00669
n	40000	40000	40000	40000	40000
y = 0					
$\hat{p}$	0.041099	0.047976	0.051467	0.053499	0.054240
$e^{-\theta^* b}$	0.0593	0.0645	0.0670	0.0682	0.0689
$A_y$	0.693	0.744	0.768	0.784	0.788
$A_y$ approxi	0.669	0.718	0.743	0.754	0.760
$A_y$ LB	0.504	0.546	0.567	0.577	0.582
$A_y$ UB	0.887	0.945	0.973	0.987	0.993
s.e.	4.62E-05	4.68E-05	4.82E-05	1.72E-04	4.96E-05
95% CI (lb)	0.04101	0.04788	0.05137	0.05316	0.05414
(ub)	0.04119	0.04807	0.05156	0.05384	0.05434
r.e.	0.001125	0.000975	0.000936	0.003208	0.000914
$P(W_y > b) / P(W > b)$					
ratio	0.78064	0.78762	0.78945	0.79463	0.79294
diff	0.01230	0.00532	0.00349	-0.00169	0.00000
abs diff	0.01230	0.00532	0.00349	0.00169	0.00000
$A_y/ ho$					
ratio	0.82476	0.80897	0.80027	0.79995	0.79559
diff	-0.02916	-0.01337	-0.00467	-0.00436	0.00000
abs diff	0.02916	0.01337	0.00467	0.00436	0.00000
y = 0.5					
$\hat{p}$	0.075260	0.086414	0.092196	0.095157	0.096491
$e^{- heta^*b}$	0.0593	0.0645	0.0670	0.0682	0.0689
$A_y$	1.269	1.341	1.376	1.394	1.401
$A_y$ approxi	1.177	1.243	1.275	1.290	1.298
$A_y$ LB	0.887	0.945	0.973	0.987	0.993
$A_y$ UB	1.561	1.635	1.671	1.688	1.696
s.e.	8.03E-05	7.92 E- 05	8.02E-05	1.83E-04	8.25E-05
95% CI (lb)	0.07510	0.08626	0.09204	0.09480	0.09633
(ub)	0.07542	0.08657	0.09235	0.09552	0.09665
r.e.	0.001067	0.000916	0.000870	0.001921	0.000855
$P(W_y > b) / P(W > b)$					
ratio	1.42950	1.41863	1.41419	1.41339	1.41060
diff	-0.01891	-0.00803	-0.00360	-0.00279	0.00000
abs diff	0.01891	0.00803	0.00360	0.00279	0.00000
$A_y/ ho$					
ratio	1.51029	1.45708	1.43357	1.42285	1.41532
diff	-0.09497	-0.04176	-0.01825	-0.00753	0.00000
abs diff	0.09497	0.04176	0.01825	0.00753	0.00000

### A.3.4 Tail Probabilities for the $M_t/H_2/1$ Periodic Queue

Tables A.39-A.49 present results for the  $M_t/H_2/1$  model paralleling the results for the  $M_t/M/1$  model in Tables A.18-A.28 and the results for the  $(H_2)_t/M/1$  model in Tables A.29-A.38.

position	n	$\hat{p}$	$\exp(-\theta^*b)$	$A_y$	$A_y/A$ approx	$A_y/A$ LB	$A_y/A$ UB	s.e.	95% CI (lb)	(ub)	r.e.
0.000	5000	0.061043	0.0839	0.728	0.976	0.952	1.000	2.46E-04	0.06056	0.06153	0.00403
0.025	5000	0.060935	0.0839	0.726	0.976	0.952	1.000	2.50E-04	0.06045	0.06142	0.00410
0.050	5000	0.060934	0.0839	0.726	0.977	0.953	1.001	2.47E-04	0.06045	0.06142	0.00406
0.075	5000	0.060531	0.0839	0.721	0.978	0.954	1.003	2.53E-04	0.06004	0.06103	0.00417
0.100	5000	0.061014	0.0839	0.727	0.980	0.956	1.005	2.50E-04	0.06052	0.06150	0.00410
0.125	5000	0.061186	0.0839	0.729	0.983	0.959	1.007	2.50E-04	0.06070	0.06168	0.00409
0.150	5000	0.061205	0.0839	0.729	0.986	0.961	1.010	2.56E-04	0.06070	0.06171	0.00418
0.175	5000	0.061299	0.0839	0.731	0.989	0.965	1.014	2.59E-04	0.06079	0.06181	0.00422
0.200	5000	0.062072	0.0839	0.740	0.992	0.968	1.017	2.50E-04	0.06158	0.06256	0.00402
0.225	5000	0.062331	0.0839	0.743	0.996	0.972	1.021	2.52E-04	0.06184	0.06283	0.00404
0.250	5000	0.062644	0.0839	0.747	1.000	0.976	1.025	2.50E-04	0.06215	0.06313	0.00399
0.275	5000	0.062369	0.0839	0.743	1.004	0.979	1.029	2.59E-04	0.06186	0.06288	0.00415
0.300	5000	0.063145	0.0839	0.753	1.008	0.983	1.033	2.57E-04	0.06264	0.06365	0.00407
0.325	5000	0.063175	0.0839	0.753	1.011	0.987	1.037	2.61E-04	0.06266	0.06369	0.00413
0.350	5000	0.063013	0.0839	0.751	1.015	0.990	1.040	2.61E-04	0.06250	0.06352	0.00414
0.375	5000	0.063367	0.0839	0.755	1.018	0.993	1.043	2.63E-04	0.06285	0.06388	0.00414
0.400	5000	0.063504	0.0839	0.757	1.020	0.995	1.046	2.64E-04	0.06299	0.06402	0.00415
0.425	5000	0.063472	0.0839	0.756	1.022	0.997	1.048	2.66E-04	0.06295	0.06399	0.00419
0.450	5000	0.063690	0.0839	0.759	1.024	0.999	1.050	2.65E-04	0.06317	0.06421	0.00415
0.475	5000	0.063951	0.0839	0.762	1.025	1.000	1.050	2.59E-04	0.06344	0.06446	0.00404
0.500	5000	0.063853	0.0839	0.761	1.025	1.000	1.051	2.65E-04	0.06333	0.06437	0.00415
0.525	5000	0.064030	0.0839	0.763	1.025	1.000	1.050	2.59E-04	0.06352	0.06454	0.00404
0.550	5000	0.063536	0.0839	0.757	1.024	0.999	1.050	2.63E-04	0.06302	0.06405	0.00415
0.575	5000	0.063183	0.0839	0.753	1.022	0.997	1.048	2.65E-04	0.06266	0.06370	0.00419
0.600	5000	0.063351	0.0839	0.755	1.020	0.995	1.046	2.68E-04	0.06283	0.06388	0.00423
0.625	5000	0.062683	0.0839	0.747	1.018	0.993	1.043	2.64E-04	0.06216	0.06320	0.00422
0.650	5000	0.063185	0.0839	0.753	1.015	0.990	1.040	2.57E-04	0.06268	0.06369	0.00407
0.675	5000	0.063070	0.0839	0.752	1.011	0.987	1.037	2.61E-04	0.06256	0.06358	0.00414
0.700	5000	0.062820	0.0839	0.749	1.008	0.983	1.033	2.59E-04	0.06231	0.06333	0.00412
0.725	5000	0.062393	0.0839	0.744	1.004	0.979	1.029	2.55E-04	0.06189	0.06289	0.00409
0.750	5000	0.062807	0.0839	0.749	1.000	0.976	1.025	2.52E-04	0.06231	0.06330	0.00401
0.775	5000	0.061698	0.0839	0.735	0.996	0.972	1.021	2.58E-04	0.06119	0.06220	0.00418
0.800	5000	0.061308	0.0839	0.731	0.992	0.968	1.017	2.57E-04	0.06080	0.06181	0.00419
0.825	5000	0.061566	0.0839	0.734	0.989	0.965	1.014	2.56E-04	0.06106	0.06207	0.00416
0.850	5000	0.060905	0.0839	0.726	0.986	0.961	1.010	2.57E-04	0.06040	0.06141	0.00423
0.875	5000	0.061046	0.0839	0.728	0.983	0.959	1.007	2.52E-04	0.06055	0.06154	0.00412
0.900	5000	0.060828	0.0839	0.725	0.980	0.956	1.005	2.51E-04	0.06034	0.06132	0.00412
0.925	5000	0.060998	0.0839	0.727	0.978	0.954	1.003	2.48E-04	0.06051	0.06148	0.00407
0.950	5000	0.060592	0.0839	0.722	0.977	0.953	1.001	2.51E-04	0.06010	0.06108	0.00414
0.975	5000	0.061300	0.0839	0.731	0.976	0.952	1.000	2.50E-04	0.06081	0.06179	0.00407

**Table A.39:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/H_2/1$  model as a function of y based on 5,000 replications:  $\gamma = 1, b = 20, \rho = 0.8, \bar{\lambda} = 0.8, \mu = 1, \beta = 0.2, \theta^* = 0.124$ 

**Table A.40:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/H_2/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.25, b = 40, \rho = 0.9, \bar{\lambda} = 0.9, \mu = 1, \beta = 0.1, \theta^* = 0.0644$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y/A$ approx	$A_y/A$ LB	$A_y/A$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.064535	0.0762	0.847	0.975	0.950	1.000	1.34E-04	0.06427	0.06480	0.00208
0.025	5000	0.064520	0.0762	0.847	0.975	0.950	1.000	1.32E-04	0.06426	0.06478	0.00204
0.050	5000	0.064459	0.0762	0.846	0.976	0.951	1.001	1.35E-04	0.06419	0.06472	0.00210
0.075	5000	0.064732	0.0762	0.850	0.977	0.952	1.003	1.35E-04	0.06447	0.06500	0.00208
0.100	5000	0.064849	0.0762	0.851	0.979	0.954	1.005	1.35E-04	0.06458	0.06511	0.00209
0.125	5000	0.064980	0.0762	0.853	0.982	0.957	1.008	1.39E-04	0.06471	0.06525	0.00214
0.150	5000	0.065339	0.0762	0.858	0.985	0.960	1.011	1.35E-04	0.06507	0.06560	0.00207
0.175	5000	0.065442	0.0762	0.859	0.988	0.963	1.014	1.37E-04	0.06517	0.06571	0.00210
0.200	5000	0.065886	0.0762	0.865	0.992	0.967	1.018	1.34E-04	0.06562	0.06615	0.00203
0.225	5000	0.065748	0.0762	0.863	0.996	0.971	1.022	1.39E-04	0.06548	0.06602	0.00212
0.250	5000	0.066450	0.0762	0.873	1.000	0.975	1.026	1.38E-04	0.06618	0.06672	0.00207
0.275	5000	0.066500	0.0762	0.873	1.004	0.979	1.030	1.40E-04	0.06622	0.06677	0.00211
0.300	5000	0.066845	0.0762	0.878	1.008	0.982	1.034	1.40E-04	0.06657	0.06712	0.00210
0.325	5000	0.066998	0.0762	0.880	1.012	0.986	1.038	1.37E-04	0.06673	0.06727	0.00205
0.350	5000	0.067298	0.0762	0.884	1.015	0.989	1.042	1.40E-04	0.06702	0.06757	0.00208
0.375	5000	0.067413	0.0762	0.885	1.018	0.992	1.045	1.40E-04	0.06714	0.06769	0.00207
0.400	5000	0.067933	0.0762	0.892	1.021	0.995	1.048	1.37E-04	0.06766	0.06820	0.00201
0.425	5000	0.067857	0.0762	0.891	1.023	0.997	1.050	1.40E-04	0.06758	0.06813	0.00207
0.450	5000	0.067828	0.0762	0.891	1.025	0.999	1.052	1.39E-04	0.06756	0.06810	0.00205
0.475	5000	0.067711	0.0762	0.889	1.026	1.000	1.053	1.43E-04	0.06743	0.06799	0.00211
0.500	5000	0.068304	0.0762	0.897	1.026	1.000	1.053	1.38E-04	0.06803	0.06857	0.00202
0.525	5000	0.068189	0.0762	0.895	1.026	1.000	1.053	1.39E-04	0.06792	0.06846	0.00204
0.550	5000	0.067899	0.0762	0.892	1.025	0.999	1.052	1.39E-04	0.06763	0.06817	0.00205
0.575	5000	0.067971	0.0762	0.892	1.023	0.997	1.050	1.42E-04	0.06769	0.06825	0.00209
0.600	5000	0.067537	0.0762	0.887	1.021	0.995	1.048	1.43E-04	0.06726	0.06782	0.00212
0.625	5000	0.067407	0.0762	0.885	1.018	0.992	1.045	1.39E-04	0.06713	0.06768	0.00207
0.650	5000	0.067242	0.0762	0.883	1.015	0.989	1.042	1.38E-04	0.06697	0.06751	0.00205
0.675	5000	0.067221	0.0762	0.883	1.012	0.986	1.038	1.37E-04	0.06695	0.06749	0.00203
0.700	5000	0.066949	0.0762	0.879	1.008	0.982	1.034	1.38E-04	0.06668	0.06722	0.00206
0.725	5000	0.066561	0.0762	0.874	1.004	0.979	1.030	1.40E-04	0.06629	0.06684	0.00210
0.750	5000	0.066419	0.0762	0.872	1.000	0.975	1.026	1.38E-04	0.06615	0.06669	0.00207
0.775	5000	0.065957	0.0762	0.866	0.996	0.971	1.022	1.39E-04	0.06569	0.06623	0.00210
0.800	5000	0.065570	0.0762	0.861	0.992	0.967	1.018	1.36E-04	0.06530	0.06584	0.00208
0.825	5000	0.065372	0.0762	0.858	0.988	0.963	1.014	1.38E-04	0.06510	0.06564	0.00211
0.850	5000	0.065319	0.0762	0.858	0.985	0.960	1.011	1.35E-04	0.06505	0.06558	0.00207
0.875	5000	0.065107	0.0762	0.855	0.982	0.957	1.008	1.34E-04	0.06484	0.06537	0.00206
0.900	5000	0.064955	0.0762	0.853	0.979	0.954	1.005	1.35E-04	0.06469	0.06522	0.00208
0.925	5000	0.064617	0.0762	0.848	0.977	0.952	1.003	1.36E-04	0.06435	0.06488	0.00211
0.950	5000	0.064682	0.0762	0.849	0.976	0.951	1.001	1.34E-04	0.06442	0.06495	0.00208
0.975	5000	0.064639	0.0762	0.849	0.975	0.950	1.000	1.35E-04	0.06437	0.06490	0.00209

**Table A.41:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/H_2/1$  model as a function of y based on 5,000 replications:  $\gamma = \frac{1}{16}, b = 80, \rho = 0.95, \bar{\lambda} = 0.95, \mu = 1, \beta = 0.05, \theta^* = 0.0328$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y/A$ approx	$A_y/A$ LB	$A_y/A$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.066280	0.0727	0.912	0.974	0.949	1.000	7.09E-05	0.06614	0.06642	0.00107
0.025	5000	0.066204	0.0727	0.911	0.974	0.949	1.000	7.10E-05	0.06607	0.06634	0.00107
0.050	5000	0.066377	0.0727	0.913	0.975	0.950	1.001	7.10E-05	0.06624	0.06652	0.00107
0.075	5000	0.066435	0.0727	0.914	0.977	0.952	1.003	6.96E-05	0.06630	0.06657	0.00105
0.100	5000	0.066556	0.0727	0.916	0.979	0.954	1.005	7.14E-05	0.06642	0.06670	0.00107
0.125	5000	0.066726	0.0727	0.918	0.982	0.956	1.008	7.11E-05	0.06659	0.06687	0.00106
0.150	5000	0.067002	0.0727	0.922	0.985	0.959	1.011	7.22E-05	0.06686	0.06714	0.00108
0.175	5000	0.067169	0.0727	0.924	0.988	0.963	1.014	7.11E-05	0.06703	0.06731	0.00106
0.200	5000	0.067501	0.0727	0.929	0.992	0.966	1.018	7.24E-05	0.06736	0.06764	0.00107
0.225	5000	0.067670	0.0727	0.931	0.996	0.970	1.022	7.21E-05	0.06753	0.06781	0.00107
0.250	5000	0.068039	0.0727	0.936	1.000	0.974	1.027	7.11E-05	0.06790	0.06818	0.00105
0.275	5000	0.068288	0.0727	0.939	1.004	0.978	1.031	7.34E-05	0.06814	0.06843	0.00107
0.300	5000	0.068467	0.0727	0.942	1.008	0.982	1.035	7.39E-05	0.06832	0.06861	0.00108
0.325	5000	0.068874	0.0727	0.947	1.012	0.986	1.039	7.23E-05	0.06873	0.06902	0.00105
0.350	5000	0.069024	0.0727	0.950	1.016	0.989	1.043	7.61 E- 05	0.06888	0.06917	0.00110
0.375	5000	0.069298	0.0727	0.953	1.019	0.992	1.046	$7.50\mathrm{E}\text{-}05$	0.06915	0.06945	0.00108
0.400	5000	0.069309	0.0727	0.953	1.021	0.995	1.049	7.57 E-05	0.06916	0.06946	0.00109
0.425	5000	0.069602	0.0727	0.957	1.024	0.997	1.051	7.45 E-05	0.06946	0.06975	0.00107
0.450	5000	0.069535	0.0727	0.957	1.025	0.999	1.052	7.69E-05	0.06938	0.06969	0.00111
0.475	5000	0.069728	0.0727	0.959	1.026	1.000	1.053	7.48E-05	0.06958	0.06987	0.00107
0.500	5000	0.069779	0.0727	0.960	1.027	1.000	1.054	7.44E-05	0.06963	0.06992	0.00107
0.525	5000	0.069683	0.0727	0.959	1.026	1.000	1.053	7.60 E- 05	0.06953	0.06983	0.00109
0.550	5000	0.069647	0.0727	0.958	1.025	0.999	1.052	$7.77\mathrm{E}\text{-}05$	0.06950	0.06980	0.00112
0.575	5000	0.069576	0.0727	0.957	1.024	0.997	1.051	7.56E-05	0.06943	0.06972	0.00109
0.600	5000	0.069369	0.0727	0.954	1.021	0.995	1.049	7.52E-05	0.06922	0.06952	0.00108
0.625	5000	0.069258	0.0727	0.953	1.019	0.992	1.046	7.44E-05	0.06911	0.06940	0.00107
0.650	5000	0.069145	0.0727	0.951	1.016	0.989	1.043	7.24E-05	0.06900	0.06929	0.00105
0.675	5000	0.068683	0.0727	0.945	1.012	0.986	1.039	7.49E-05	0.06854	0.06883	0.00109
0.700	5000	0.068628	0.0727	0.944	1.008	0.982	1.035	7.22E-05	0.06849	0.06877	0.00105
0.725	5000	0.068246	0.0727	0.939	1.004	0.978	1.031	7.47 E-05	0.06810	0.06839	0.00109
0.750	5000	0.067919	0.0727	0.934	1.000	0.974	1.027	7.29E-05	0.06778	0.06806	0.00107
0.775	5000	0.067731	0.0727	0.932	0.996	0.970	1.022	7.39E-05	0.06759	0.06788	0.00109
0.800	5000	0.067406	0.0727	0.927	0.992	0.966	1.018	7.36E-05	0.06726	0.06755	0.00109
0.825	5000	0.067147	0.0727	0.924	0.988	0.963	1.014	7.26E-05	0.06700	0.06729	0.00108
0.850	5000	0.066820	0.0727	0.919	0.985	0.959	1.011	7.29E-05	0.06668	0.06696	0.00109
0.875	5000	0.066765	0.0727	0.918	0.982	0.956	1.008	7.14E-05	0.06662	0.06690	0.00107
0.900	5000	0.066668	0.0727	0.917	0.979	0.954	1.005	7.14E-05	0.06653	0.06681	0.00107
0.925	5000	0.066418	0.0727	0.914	0.977	0.952	1.003	7.21E-05	0.06628	0.06656	0.00109
0.950	5000	0.066375	0.0727	0.913	0.975	0.950	1.001	7.11E-05	0.06624	0.06651	0.00107
0.975	5000	0.066309	0.0727	0.912	0.974	0.949	1.000	7.06E-05	0.06617	0.06645	0.00106

**Table A.42:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/H_2/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.64, b = 25, \rho = 0.84, \bar{\lambda} = 0.84, \mu = 1, \beta = 0.16, \theta^* = 0.101$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.062304	0.0807	0.772	0.771	0.752	0.791	2.07E-04	0.06190	0.06271	0.00332
0.025	5000	0.062294	0.0807	0.772	0.771	0.752	0.791	2.07E-04	0.06189	0.06270	0.00333
0.050	5000	0.062341	0.0807	0.773	0.772	0.753	0.792	2.08E-04	0.06193	0.06275	0.00333
0.075	5000	0.062355	0.0807	0.773	0.773	0.754	0.793	2.08E-04	0.06195	0.06276	0.00333
0.100	5000	0.062523	0.0807	0.775	0.775	0.756	0.795	2.08E-04	0.06212	0.06293	0.00333
0.125	5000	0.062689	0.0807	0.777	0.777	0.758	0.797	2.08E-04	0.06228	0.06310	0.00332
0.150	5000	0.062893	0.0807	0.780	0.779	0.760	0.799	2.09E-04	0.06248	0.06330	0.00332
0.175	5000	0.063137	0.0807	0.783	0.782	0.762	0.802	2.09E-04	0.06273	0.06355	0.00332
0.200	5000	0.063326	0.0807	0.785	0.785	0.765	0.805	2.10E-04	0.06291	0.06374	0.00332
0.225	5000	0.063531	0.0807	0.788	0.788	0.768	0.808	2.11E-04	0.06312	0.06394	0.00332
0.250	5000	0.063743	0.0807	0.790	0.791	0.771	0.811	2.12E-04	0.06333	0.06416	0.00333
0.275	5000	0.063997	0.0807	0.793	0.794	0.774	0.814	2.13E-04	0.06358	0.06441	0.00333
0.300	5000	0.064233	0.0807	0.796	0.797	0.777	0.817	2.14E-04	0.06381	0.06465	0.00333
0.325	5000	0.064497	0.0807	0.800	0.800	0.780	0.820	2.14E-04	0.06408	0.06492	0.00332
0.350	5000	0.064751	0.0807	0.803	0.803	0.783	0.823	2.14E-04	0.06433	0.06517	0.00331
0.375	5000	0.064921	0.0807	0.805	0.805	0.785	0.826	2.15E-04	0.06450	0.06534	0.00331
0.400	5000	0.065042	0.0807	0.806	0.807	0.787	0.828	2.16E-04	0.06462	0.06546	0.00332
0.425	5000	0.065179	0.0807	0.808	0.809	0.789	0.829	2.16E-04	0.06476	0.06560	0.00331
0.450	5000	0.065249	0.0807	0.809	0.810	0.790	0.831	2.16E-04	0.06482	0.06567	0.00332
0.475	5000	0.065345	0.0807	0.810	0.811	0.791	0.831	2.17E-04	0.06492	0.06577	0.00332
0.500	5000	0.065349	0.0807	0.810	0.811	0.791	0.832	2.17E-04	0.06492	0.06577	0.00332
0.525	5000	0.065313	0.0807	0.810	0.811	0.791	0.831	2.17E-04	0.06489	0.06574	0.00333
0.550	5000	0.065244	0.0807	0.809	0.810	0.790	0.831	2.17E-04	0.06482	0.06567	0.00333
0.575	5000	0.065193	0.0807	0.808	0.809	0.789	0.829	2.17E-04	0.06477	0.06562	0.00332
0.600	5000	0.065069	0.0807	0.807	0.807	0.787	0.828	2.16E-04	0.06465	0.06549	0.00332
0.625	5000	0.064912	0.0807	0.805	0.805	0.785	0.826	2.16E-04	0.06449	0.06534	0.00333
0.650	5000	0.064713	0.0807	0.802	0.803	0.783	0.823	2.16E-04	0.06429	0.06514	0.00333
0.675	5000	0.064523	0.0807	0.800	0.800	0.780	0.820	2.15E-04	0.06410	0.06494	0.00333
0.700	5000	0.064290	0.0807	0.797	0.797	0.777	0.817	2.14E-04	0.06387	0.06471	0.00333
0.725	5000	0.064135	0.0807	0.795	0.794	0.774	0.814	2.13E-04	0.06372	0.06455	0.00332
0.750	5000	0.063932	0.0807	0.792	0.791	0.771	0.811	2.12E-04	0.06352	0.06435	0.00332
0.775	5000	0.063708	0.0807	0.790	0.788	0.768	0.808	2.11E-04	0.06330	0.06412	0.00331
0.800	5000	0.063435	0.0807	0.786	0.785	0.765	0.805	2.10E-04	0.06302	0.06385	0.00331
0.825	5000	0.063174	0.0807	0.783	0.782	0.762	0.802	2.09E-04	0.06276	0.06358	0.00331
0.850	5000	0.062899	0.0807	0.780	0.779	0.760	0.799	2.09E-04	0.06249	0.06331	0.00332
0.875	5000	0.062675	0.0807	0.777	0.777	0.758	0.797	2.08E-04	0.06227	0.06308	0.00332
0.900	5000	0.062508	0.0807	0.775	0.775	0.756	0.795	2.08E-04	0.06210	0.06292	0.00333
0.925	5000	0.062427	0.0807	0.774	0.773	0.754	0.793	2.08E-04	0.06202	0.06283	0.00332
0.950	5000	0.062330	0.0807	0.773	0.772	0.753	0.792	2.07E-04	0.06192	0.06274	0.00333
0.975	5000	0.062298	0.0807	0.772	0.771	0.752	0.791	2.07E-04	0.06189	0.06270	0.00332

**Table A.43:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/H_2/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.16, b = 50, \rho = 0.92, \bar{\lambda} = 0.92, \mu = 1, \beta = 0.08, \theta^* = 0.0519$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.065304	0.0747	0.874	0.873	0.851	0.896	1.07E-04	0.06509	0.06551	0.00164
0.025	5000	0.065295	0.0747	0.874	0.873	0.851	0.896	1.07E-04	0.06508	0.06551	0.00165
0.050	5000	0.065349	0.0747	0.874	0.874	0.852	0.897	1.08E-04	0.06514	0.06556	0.00165
0.075	5000	0.065448	0.0747	0.876	0.875	0.853	0.898	1.08E-04	0.06524	0.06566	0.00165
0.100	5000	0.065563	0.0747	0.877	0.877	0.855	0.900	1.08E-04	0.06535	0.06577	0.00164
0.125	5000	0.065746	0.0747	0.880	0.880	0.857	0.903	1.08E-04	0.06553	0.06596	0.00164
0.150	5000	0.065973	0.0747	0.883	0.882	0.860	0.905	1.08E-04	0.06576	0.06618	0.00163
0.175	5000	0.066185	0.0747	0.886	0.885	0.863	0.909	1.08E-04	0.06597	0.06640	0.00164
0.200	5000	0.066471	0.0747	0.889	0.889	0.866	0.912	1.09E-04	0.06626	0.06668	0.00164
0.225	5000	0.066755	0.0747	0.893	0.892	0.869	0.916	1.09E-04	0.06654	0.06697	0.00163
0.250	5000	0.067060	0.0747	0.897	0.896	0.873	0.919	1.09E-04	0.06685	0.06727	0.00163
0.275	5000	0.067334	0.0747	0.901	0.899	0.876	0.923	1.10E-04	0.06712	0.06755	0.00163
0.300	5000	0.067595	0.0747	0.904	0.903	0.880	0.927	1.11E-04	0.06738	0.06781	0.00164
0.325	5000	0.067848	0.0747	0.908	0.906	0.883	0.930	1.11E-04	0.06763	0.06807	0.00164
0.350	5000	0.068096	0.0747	0.911	0.910	0.886	0.934	1.11E-04	0.06788	0.06831	0.00163
0.375	5000	0.068321	0.0747	0.914	0.912	0.889	0.936	1.11E-04	0.06810	0.06854	0.00163
0.400	5000	0.068484	0.0747	0.916	0.915	0.891	0.939	1.12E-04	0.06826	0.06870	0.00163
0.425	5000	0.068620	0.0747	0.918	0.917	0.893	0.941	1.12E-04	0.06840	0.06884	0.00163
0.450	5000	0.068664	0.0747	0.919	0.918	0.895	0.942	1.12E-04	0.06844	0.06888	0.00164
0.475	5000	0.068750	0.0747	0.920	0.919	0.896	0.943	1.12E-04	0.06853	0.06897	0.00164
0.500	5000	0.068803	0.0747	0.921	0.919	0.896	0.944	1.12E-04	0.06858	0.06902	0.00163
0.525	5000	0.068795	0.0747	0.920	0.919	0.896	0.943	1.12E-04	0.06858	0.06901	0.00163
0.550	5000	0.068749	0.0747	0.920	0.918	0.895	0.942	1.12E-04	0.06853	0.06897	0.00163
0.575	5000	0.068648	0.0747	0.918	0.917	0.893	0.941	1.12E-04	0.06843	0.06887	0.00163
0.600	5000	0.068517	0.0747	0.917	0.915	0.891	0.939	1.12E-04	0.06830	0.06874	0.00163
0.625	5000	0.068336	0.0747	0.914	0.912	0.889	0.936	1.11E-04	0.06812	0.06855	0.00163
0.650	5000	0.068135	0.0747	0.912	0.910	0.886	0.934	1.11E-04	0.06792	0.06835	0.00163
0.675	5000	0.067861	0.0747	0.908	0.906	0.883	0.930	1.11E-04	0.06764	0.06808	0.00163
0.700	5000	0.067609	0.0747	0.905	0.903	0.880	0.927	1.10E-04	0.06739	0.06783	0.00163
0.725	5000	0.067362	0.0747	0.901	0.899	0.876	0.923	1.10E-04	0.06715	0.06758	0.00163
0.750	5000	0.067105	0.0747	0.898	0.896	0.873	0.919	1.10E-04	0.06689	0.06732	0.00163
0.775	5000	0.066853	0.0747	0.894	0.892	0.869	0.916	1.09E-04	0.06664	0.06707	0.00163
0.800	5000	0.066597	0.0747	0.891	0.889	0.866	0.912	1.09E-04	0.06638	0.06681	0.00163
0.825	5000	0.066355	0.0747	0.888	0.885	0.863	0.909	1.08E-04	0.06614	0.06657	0.00163
0.850	5000	0.066106	0.0747	0.884	0.882	0.860	0.905	1.08E-04	0.06589	0.06632	0.00164
0.875	5000	0.065883	0.0747	0.881	0.880	0.857	0.903	1.08E-04	0.06567	0.06609	0.00164
0.900	5000	0.065680	0.0747	0.879	0.877	0.855	0.900	1.08E-04	0.06547	0.06589	0.00164
0.925	5000	0.065529	0.0747	0.877	0.875	0.853	0.898	1.07E-04	0.06532	0.06574	0.00164
0.950	5000	0.065427	0.0747	0.875	0.874	0.852	0.897	1.07E-04	0.06522	0.06564	0.00164
0.975	5000	0.065332	0.0747	0.874	0.873	0.851	0.896	1.07E-04	0.06512	0.06554	0.00164

**Table A.44:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/H_2/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.04, b = 100, \rho = 0.96, \bar{\lambda} = 0.96, \mu = 1, \beta = 0.04, \theta^* = 0.0263$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y$ approx	$A_y$ LB	$A_y$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.066528	0.0720	0.924	0.923	0.899	0.947	5.86E-05	0.06641	0.06664	8.80E-04
0.025	5000	0.066579	0.0720	0.924	0.923	0.899	0.947	5.86E-05	0.06646	0.06669	8.79E-04
0.050	5000	0.066657	0.0720	0.925	0.924	0.900	0.948	5.86E-05	0.06654	0.06677	8.79E-04
0.075	5000	0.066735	0.0720	0.926	0.925	0.901	0.950	5.91 E- 05	0.06662	0.06685	8.86E-04
0.100	5000	0.066889	0.0720	0.929	0.927	0.903	0.952	5.90 E- 05	0.06677	0.06700	8.82E-04
0.125	5000	0.067072	0.0720	0.931	0.930	0.906	0.954	5.89E-05	0.06696	0.06719	8.79E-04
0.150	5000	0.067302	0.0720	0.934	0.933	0.908	0.957	5.89E-05	0.06719	0.06742	8.76E-04
0.175	5000	0.067528	0.0720	0.937	0.936	0.912	0.961	5.88E-05	0.06741	0.06764	8.71E-04
0.200	5000	0.067773	0.0720	0.941	0.939	0.915	0.964	5.92E-05	0.06766	0.06789	8.73E-04
0.225	5000	0.068063	0.0720	0.945	0.943	0.919	0.968	5.96E-05	0.06795	0.06818	8.76E-04
0.250	5000	0.068348	0.0720	0.949	0.947	0.923	0.972	5.97 E- 05	0.06823	0.06846	8.73E-04
0.275	5000	0.068644	0.0720	0.953	0.951	0.926	0.976	5.97 E- 05	0.06853	0.06876	8.70E-04
0.300	5000	0.068937	0.0720	0.957	0.955	0.930	0.980	6.00E-05	0.06882	0.06906	8.70E-04
0.325	5000	0.069186	0.0720	0.960	0.958	0.934	0.984	6.02E-05	0.06907	0.06930	8.70E-04
0.350	5000	0.069408	0.0720	0.964	0.962	0.937	0.988	6.04E-05	0.06929	0.06953	8.70E-04
0.375	5000	0.069613	0.0720	0.966	0.965	0.940	0.991	6.06E-05	0.06949	0.06973	8.71E-04
0.400	5000	0.069791	0.0720	0.969	0.967	0.942	0.993	6.11E-05	0.06967	0.06991	8.75E-04
0.425	5000	0.069939	0.0720	0.971	0.970	0.944	0.995	6.15 E- 05	0.06982	0.07006	8.79E-04
0.450	5000	0.070041	0.0720	0.972	0.971	0.946	0.997	6.17 E-05	0.06992	0.07016	8.80E-04
0.475	5000	0.070110	0.0720	0.973	0.972	0.947	0.998	6.18E-05	0.06999	0.07023	8.82E-04
0.500	5000	0.070128	0.0720	0.974	0.972	0.947	0.998	6.20E-05	0.07001	0.07025	8.85E-04
0.525	5000	0.070100	0.0720	0.973	0.972	0.947	0.998	6.19E-05	0.06998	0.07022	8.82E-04
0.550	5000	0.070048	0.0720	0.972	0.971	0.946	0.997	6.19E-05	0.06993	0.07017	8.84E-04
0.575	5000	0.069921	0.0720	0.971	0.970	0.944	0.995	6.19E-05	0.06980	0.07004	8.86E-04
0.600	5000	0.069775	0.0720	0.969	0.967	0.942	0.993	6.19E-05	0.06965	0.06990	8.87E-04
0.625	5000	0.069596	0.0720	0.966	0.965	0.940	0.991	6.14E-05	0.06948	0.06972	8.82E-04
0.650	5000	0.069396	0.0720	0.963	0.962	0.937	0.988	6.11E-05	0.06928	0.06952	8.81E-04
0.675	5000	0.069136	0.0720	0.960	0.958	0.934	0.984	6.09E-05	0.06902	0.06926	8.80E-04
0.700	5000	0.068862	0.0720	0.956	0.955	0.930	0.980	6.05 E-05	0.06874	0.06898	8.78E-04
0.725	5000	0.068604	0.0720	0.952	0.951	0.926	0.976	5.98E-05	0.06849	0.06872	8.72E-04
0.750	5000	0.068341	0.0720	0.949	0.947	0.923	0.972	5.98E-05	0.06822	0.06846	8.75E-04
0.775	5000	0.068085	0.0720	0.945	0.943	0.919	0.968	5.95 E- 05	0.06797	0.06820	8.74E-04
0.800	5000	0.067825	0.0720	0.942	0.939	0.915	0.964	5.95 E- 05	0.06771	0.06794	8.78E-04
0.825	5000	0.067534	0.0720	0.938	0.936	0.912	0.961	5.94E-05	0.06742	0.06765	8.80E-04
0.850	5000	0.067285	0.0720	0.934	0.933	0.908	0.957	5.89E-05	0.06717	0.06740	8.75E-04
0.875	5000	0.067081	0.0720	0.931	0.930	0.906	0.954	5.89E-05	0.06697	0.06720	8.78E-04
0.900	5000	0.066872	0.0720	0.928	0.927	0.903	0.952	5.89E-05	0.06676	0.06699	8.81E-04
0.925	5000	0.066735	0.0720	0.926	0.925	0.901	0.950	5.86E-05	0.06662	0.06685	8.78E-04
0.950	5000	0.066627	0.0720	0.925	0.924	0.900	0.948	5.85E-05	0.06651	0.06674	8.78E-04
0.975	5000	0.066558	0.0720	0.924	0.923	0.899	0.947	5.88E-05	0.06644	0.06667	8.84E-04

**Table A.45:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/H_2/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.01, b = 200, \rho = 0.98, \bar{\lambda} = 0.98, \mu = 1, \beta = 0.02, \theta^* = 0.0132$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y/A$ approx	$A_y/A$ LB	$A_y/A$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.067202	0.0707	0.950	0.974	0.948	1.000	3.33E-05	0.06714	0.06727	4.95E-04
0.025	5000	0.067195	0.0707	0.950	0.974	0.949	1.000	3.30E-05	0.06713	0.06726	4.91E-04
0.050	5000	0.067252	0.0707	0.951	0.975	0.950	1.001	3.35E-05	0.06719	0.06732	4.98E-04
0.075	5000	0.067346	0.0707	0.952	0.977	0.951	1.003	3.39E-05	0.06728	0.06741	5.04E-04
0.100	5000	0.067486	0.0707	0.954	0.979	0.953	1.005	3.39E-05	0.06742	0.06755	5.02E-04
0.125	5000	0.067691	0.0707	0.957	0.981	0.956	1.008	3.41E-05	0.06762	0.06776	5.04E-04
0.150	5000	0.067868	0.0707	0.959	0.985	0.959	1.011	3.40E-05	0.06780	0.06794	5.01E-04
0.175	5000	0.068178	0.0707	0.964	0.988	0.962	1.015	3.35E-05	0.06811	0.06824	4.92E-04
0.200	5000	0.068420	0.0707	0.967	0.992	0.966	1.018	3.40E-05	0.06835	0.06849	4.96E-04
0.225	5000	0.068639	0.0707	0.970	0.996	0.970	1.023	3.46E-05	0.06857	0.06871	5.04E-04
0.250	5000	0.068940	0.0707	0.975	1.000	0.974	1.027	3.42E-05	0.06887	0.06901	4.96E-04
0.275	5000	0.069295	0.0707	0.980	1.004	0.978	1.031	3.42E-05	0.06923	0.06936	4.94E-04
0.300	5000	0.069531	0.0707	0.983	1.008	0.982	1.035	3.45E-05	0.06946	0.06960	4.97E-04
0.325	5000	0.069780	0.0707	0.986	1.012	0.986	1.039	3.48E-05	0.06971	0.06985	4.99E-04
0.350	5000	0.069954	0.0707	0.989	1.016	0.989	1.043	3.57E-05	0.06988	0.07002	5.10E-04
0.375	5000	0.070282	0.0707	0.994	1.019	0.992	1.046	3.38E-05	0.07022	0.07035	4.81E-04
0.400	5000	0.070463	0.0707	0.996	1.022	0.995	1.049	3.51E-05	0.07039	0.07053	4.98E-04
0.425	5000	0.070639	0.0707	0.999	1.024	0.997	1.051	3.44E-05	0.07057	0.07071	4.87E-04
0.450	5000	0.070742	0.0707	1.000	1.026	0.999	1.053	3.43E-05	0.07068	0.07081	4.85E-04
0.475	5000	0.070811	0.0707	1.001	1.027	1.000	1.054	3.48E-05	0.07074	0.07088	4.91E-04
0.500	5000	0.070867	0.0707	1.002	1.027	1.000	1.054	3.46E-05	0.07080	0.07093	4.88E-04
0.525	5000	0.070781	0.0707	1.001	1.027	1.000	1.054	3.54E-05	0.07071	0.07085	5.00E-04
0.550	5000	0.070717	0.0707	1.000	1.026	0.999	1.053	3.55E-05	0.07065	0.07079	5.02E-04
0.575	5000	0.070637	0.0707	0.999	1.024	0.997	1.051	3.60E-05	0.07057	0.07071	5.09E-04
0.600	5000	0.070510	0.0707	0.997	1.022	0.995	1.049	3.47E-05	0.07044	0.07058	4.92E-04
0.625	5000	0.070218	0.0707	0.993	1.019	0.992	1.046	3.51E-05	0.07015	0.07029	5.00E-04
0.650	5000	0.070101	0.0707	0.991	1.016	0.989	1.043	3.44E-05	0.07003	0.07017	4.90E-04
0.675	5000	0.069818	0.0707	0.987	1.012	0.986	1.039	3.40E-05	0.06975	0.06988	4.87E-04
0.700	5000	0.069552	0.0707	0.983	1.008	0.982	1.035	3.40E-05	0.06949	0.06962	4.89E-04
0.725	5000	0.069181	0.0707	0.978	1.004	0.978	1.031	3.47E-05	0.06911	0.06925	5.02E-04
0.750	5000	0.068975	0.0707	0.975	1.000	0.974	1.027	3.45E-05	0.06891	0.06904	5.00E-04
0.775	5000	0.068746	0.0707	0.972	0.996	0.970	1.023	3.38E-05	0.06868	0.06881	4.92E-04
0.800	5000	0.068349	0.0707	0.966	0.992	0.966	1.018	3.44E-05	0.06828	0.06842	5.03E-04
0.825	5000	0.068149	0.0707	0.963	0.988	0.962	1.015	3.31E-05	0.06808	0.06821	4.86E-04
0.850	5000	0.067861	0.0707	0.959	0.985	0.959	1.011	3.41E-05	0.06779	0.06793	5.02E-04
0.875	5000	0.067708	0.0707	0.957	0.981	0.956	1.008	3.37E-05	0.06764	0.06777	4.98E-04
0.900	5000	0.067490	0.0707	0.954	0.979	0.953	1.005	3.31E-05	0.06742	0.06755	4.91E-04
0.925	5000	0.067377	0.0707	0.952	0.977	0.951	1.003	3.38E-05	0.06731	0.06744	5.02E-04
0.950	5000	0.067222	0.0707	0.950	0.975	0.950	1.001	3.39E-05	0.06716	0.06729	5.04E-04
0.975	5000	0.067210	0.0707	0.950	0.974	0.949	1.000	3.31E-05	0.06715	0.06727	4.92E-04

**Table A.46:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $M_t/H_2/1$  model as a function of y based on 5,000 replications:  $\gamma = 0.0025, b = 400, \rho = 0.99, \bar{\lambda} = 0.99, \mu = 1, \beta = 0.01, \theta^* = 0.00664$ 

 position	n	$\hat{p}$	$\exp(-\theta^* b)$	$A_y$	$A_y/A$ approx	$A_y/A$ LB	$A_y/A$ UB	s.e.	95% CI (lb)	(ub)	r.e.
 0.000	5000	0.067433	0.0701	0.962	0.974	0.948	1.000	2.31E-05	0.06739	0.06748	3.43E-04
0.025	5000	0.067502	0.0701	0.963	0.974	0.949	1.000	2.24E-05	0.06746	0.06755	3.32E-04
0.050	5000	0.067568	0.0701	0.964	0.975	0.949	1.001	2.27E-05	0.06752	0.06761	3.36E-04
0.075	5000	0.067632	0.0701	0.965	0.977	0.951	1.003	2.31E-05	0.06759	0.06768	3.42E-04
0.100	5000	0.067785	0.0701	0.967	0.979	0.953	1.005	2.30E-05	0.06774	0.06783	3.39E-04
0.125	5000	0.068011	0.0701	0.970	0.981	0.956	1.008	2.27E-05	0.06797	0.06806	3.34E-04
0.150	5000	0.068171	0.0701	0.972	0.984	0.959	1.011	2.32E-05	0.06813	0.06822	3.41E-04
0.175	5000	0.068453	0.0701	0.976	0.988	0.962	1.015	2.32E-05	0.06841	0.06850	3.39E-04
0.200	5000	0.068726	0.0701	0.980	0.992	0.966	1.019	2.33E-05	0.06868	0.06877	3.40E-04
0.225	5000	0.069005	0.0701	0.984	0.996	0.970	1.023	2.32E-05	0.06896	0.06905	3.37E-04
0.250	5000	0.069337	0.0701	0.989	1.000	0.974	1.027	2.31E-05	0.06929	0.06938	3.33E-04
0.275	5000	0.069615	0.0701	0.993	1.004	0.978	1.031	2.35E-05	0.06957	0.06966	3.37E-04
0.300	5000	0.069855	0.0701	0.996	1.008	0.982	1.035	2.36E-05	0.06981	0.06990	3.38E-04
0.325	5000	0.070150	0.0701	1.001	1.012	0.986	1.039	2.34E-05	0.07010	0.07020	3.34E-04
0.350	5000	0.070378	0.0701	1.004	1.016	0.989	1.043	2.38E-05	0.07033	0.07043	3.39E-04
0.375	5000	0.070605	0.0701	1.007	1.019	0.992	1.046	2.40E-05	0.07056	0.07065	3.40E-04
0.400	5000	0.070764	0.0701	1.009	1.022	0.995	1.049	2.39E-05	0.07072	0.07081	3.38E-04
0.425	5000	0.070956	0.0701	1.012	1.024	0.997	1.052	2.33E-05	0.07091	0.07100	3.29E-04
0.450	5000	0.071038	0.0701	1.013	1.026	0.999	1.053	2.38E-05	0.07099	0.07108	3.36E-04
0.475	5000	0.071140	0.0701	1.015	1.027	1.000	1.054	2.39E-05	0.07109	0.07119	3.36E-04
0.500	5000	0.071167	0.0701	1.015	1.027	1.000	1.055	2.43E-05	0.07112	0.07121	3.41E-04
0.525	5000	0.071103	0.0701	1.014	1.027	1.000	1.054	2.43E-05	0.07106	0.07115	3.42E-04
0.550	5000	0.071106	0.0701	1.014	1.026	0.999	1.053	2.42E-05	0.07106	0.07115	3.40E-04
0.575	5000	0.070928	0.0701	1.012	1.024	0.997	1.052	2.39E-05	0.07088	0.07097	3.38E-04
0.600	5000	0.070775	0.0701	1.010	1.022	0.995	1.049	2.44E-05	0.07073	0.07082	3.45E-04
0.625	5000	0.070609	0.0701	1.007	1.019	0.992	1.046	2.36E-05	0.07056	0.07066	3.34E-04
0.650	5000	0.070368	0.0701	1.004	1.016	0.989	1.043	2.37E-05	0.07032	0.07041	3.36E-04
0.675	5000	0.070112	0.0701	1.000	1.012	0.986	1.039	2.39E-05	0.07007	0.07016	3.41E-04
0.700	5000	0.069854	0.0701	0.996	1.008	0.982	1.035	2.36E-05	0.06981	0.06990	3.38E-04
0.725	5000	0.069574	0.0701	0.992	1.004	0.978	1.031	2.32E-05	0.06953	0.06962	3.33E-04
0.750	5000	0.069314	0.0701	0.989	1.000	0.974	1.027	2.34E-05	0.06927	0.06936	3.37E-04
0.775	5000	0.069002	0.0701	0.984	0.996	0.970	1.023	2.32E-05	0.06896	0.06905	3.36E-04
0.800	5000	0.068719	0.0701	0.980	0.992	0.966	1.019	2.31E-05	0.06867	0.06876	3.36E-04
0.825	5000	0.068468	0.0701	0.977	0.988	0.962	1.015	2.29E-05	0.06842	0.06851	3.35E-04
0.850	5000	0.068245	0.0701	0.973	0.984	0.959	1.011	2.32E-05	0.06820	0.06829	3.40E-04
0.875	5000	0.067991	0.0701	0.970	0.981	0.956	1.008	2.28E-05	0.06795	0.06804	3.35E-04
0.900	5000	0.067803	0.0701	0.967	0.979	0.953	1.005	2.33E-05	0.06776	0.06785	3.43E-04
0.925	5000	0.067694	0.0701	0.966	0.977	0.951	1.003	2.22E-05	0.06765	0.06774	3.27E-04
0.950	5000	0.067575	0.0701	0.964	0.975	0.949	1.001	2.23E-05	0.06753	0.06762	3.30E-04
0.975	5000	0.067501	0.0701	0.963	0.974	0.949	1.000	2.33E-05	0.06746	0.06755	3.45E-04

position	$\rho = 0.8$	$\rho = 0.9$	$\rho=0.95$	$\rho=0.98$	$\rho=0.99$
0.000	0.90943	0.94152	0.95975	0.96940	0.97159
0.025	0.90782	0.94131	0.95866	0.96930	0.97258
0.050	0.90781	0.94042	0.96116	0.97012	0.97353
0.075	0.90180	0.94440	0.96200	0.97147	0.97445
0.100	0.90900	0.94611	0.96375	0.97349	0.97667
0.125	0.91156	0.94802	0.96621	0.97645	0.97992
0.150	0.91184	0.95325	0.97021	0.97901	0.98223
0.175	0.91325	0.95476	0.97262	0.98348	0.98629
0.200	0.92476	0.96123	0.97744	0.98696	0.99022
0.225	0.92863	0.95922	0.97988	0.99013	0.99423
0.250	0.93328	0.96945	0.98522	0.99447	0.99903
0.275	0.92918	0.97018	0.98883	0.99960	1.00303
0.300	0.94075	0.97523	0.99142	1.00300	1.00649
0.325	0.94119	0.97746	0.99731	1.00659	1.01073
0.350	0.93878	0.98184	0.99949	1.00910	1.01403
0.375	0.94405	0.98350	1.00346	1.01383	1.01730
0.400	0.94610	0.99109	1.00361	1.01644	1.01958
0.425	0.94562	0.98998	1.00786	1.01898	1.02235
0.450	0.94886	0.98957	1.00688	1.02047	1.02353
0.475	0.95275	0.98786	1.00967	1.02146	1.02500
0.500	0.95129	0.99651	1.01042	1.02227	1.02538
0.525	0.95393	0.99483	1.00902	1.02103	1.02447
0.550	0.94658	0.99060	1.00851	1.02011	1.02451
0.575	0.94131	0.99165	1.00748	1.01895	1.02194
0.600	0.94382	0.98532	1.00448	1.01711	1.01974
0.625	0.93386	0.98342	1.00288	1.01291	1.01735
0.650	0.94133	0.98102	1.00125	1.01122	1.01388
0.675	0.93963	0.98072	0.99455	1.00713	1.01019
0.700	0.93590	0.97674	0.99375	1.00330	1.00647
0.725	0.92954	0.97109	0.98823	0.99795	1.00244
0.750	0.93571	0.96901	0.98348	0.99497	0.99869
0.775	0.91919	0.96227	0.98076	0.99167	0.99420
0.800	0.91338	0.95662	0.97606	0.98594	0.99012
0.825	0.91722	0.95373	0.97230	0.98306	0.98650
0.850	0.90737	0.95296	0.96758	0.97891	0.98328
0.875	0.90948	0.94987	0.96677	0.97670	0.97963
0.900	0.90623	0.94765	0.96537	0.97355	0.97691
0.925	0.90875	0.94272	0.96176	0.97193	0.97535
0.950	0.90272	0.94366	0.96113	0.96969	0.97363
0.975	0.91325	0.94305	0.96017	0.96952	0.97257
avg diff w.r.t. last column	-0.07108	-0.03151	-0.01397	-0.00346	0.00000
avg. abs. diff w.r.t. last column	0.07108	0.03151	0.01397	0.00346	0.00000
rmse w.r.t. last column	0.07126	0.03157	0.01402	0.00351	0.00000

**Table A.47:** Comparison of ratio  $P(W_y > b)/\rho$  as a function of  $\rho$  in  $M_t/H_2/1$  queue with base parameter  $(\beta, \gamma, b) = (1, 25, 4)$  using the scaling in (3.39) of the main thesis.

**Table A.48:** Summary of simulation results for  $M_t/H_2/1$  queue at y = 0 as a function of  $1 - \rho$  with base parameter  $(\beta, \gamma, b) = (1, 25, 4)$  using the scaling in (3.39) of the main thesis

	$1 - \rho = 0.16$	$1-\rho=0.08$	$1-\rho=0.04$	$1-\rho=0.02$	$1-\rho=0.01$	$1 - \rho = 0.005$
$ heta^*$	0.101	0.0519	0.0263	0.0132	0.00664	0.00333
n	5000	5000	5000	5000	5000	5000
$\hat{p}$	0.061910	0.065213	0.066492	0.067148	0.067429	0.067641
$e^{-\theta^* b}$	0.0807	0.0747	0.0720	0.0707	0.0701	0.0698
$A_y$	0.767	0.873	0.923	0.949	0.962	0.969
$A_y$ approxi	0.766	0.873	0.921	0.948	0.961	0.967
$A_y$ LB	0.747	0.851	0.897	0.923	0.936	0.942
$A_y$ UB	0.786	0.896	0.945	0.973	0.987	0.993
s.e.	2.06E-04	1.09E-04	5.88E-05	3.28E-05	2.27E-05	1.92E-05
95% CI (lb)	0.06151	0.06500	0.06638	0.06708	0.06738	0.06760
(ub)	0.06231	0.06543	0.06661	0.06721	0.06747	0.06768
r.e.	0.003327	0.001665	0.000885	0.000489	0.000337	0.000283
$P(W_y > b) / P(W > b)$						
ratio	0.97659	0.97396	0.97632	0.97526	0.97480	0.97562
diff w.r.t. last column	-0.00097	0.00166	-0.00070	0.00036	0.00082	0.00000
abs diff w.r.t. last column	0.00097	0.00166	0.00070	0.00036	0.00082	0.00000
$A_y/ ho$						
ratio	0.91361	0.94838	0.96155	0.96861	0.97153	0.97403
diff w.r.t. last column	0.06042	0.02564	0.01248	0.00541	0.00250	0.00000
abs diff w.r.t. last column	0.06042	0.02564	0.01248	0.00541	0.00250	0.00000

	$1 - \rho = 0.16$	$1-\rho=0.08$	$1-\rho=0.04$	$1-\rho=0.02$	$1 - \rho = 0.01$
$ heta^*$	0.101	0.0519	0.0263	0.0132	0.00664
n	40000	40000	40000	40000	40000
y = 0					
$\hat{p}$	0.050594	0.052946	0.054024	0.054544	0.054904
$e^{-\theta^*b}$	0.0807	0.0747	0.0720	0.0707	0.0701
$A_y$	0.627	0.708	0.750	0.771	0.783
$A_y$ approxi	0.613	0.690	0.728	0.747	0.756
$A_y$ LB	0.477	0.532	0.560	0.573	0.580
$A_y$ UB	0.789	0.894	0.947	0.974	0.987
s.e.	7.49E-05	5.64E-05	5.13E-05	5.03E-05	5.01E-05
95% CI (lb)	0.05045	0.05284	0.05392	0.05445	0.05481
(ub)	0.05074	0.05306	0.05412	0.05464	0.05500
r.e.	0.001480	0.001065	0.000950	0.000923	0.000913
$P(W_y > b) / P(W > b)$					
ratio	0.79534	0.79246	0.79200	0.79200	0.79377
diff w.r.t. last column	-0.00158	0.00131	0.00177	0.00177	0.00000
abs diff	0.00158	0.00131	0.00177	0.00177	0.00000
$A_y/ ho$					
ratio	0.74662	0.76999	0.78125	0.78680	0.79107
diff w.r.t. last column	0.04445	0.02108	0.00982	0.00427	0.00000
abs diff	0.04445	0.02108	0.00982	0.00427	0.00000
y = 0.5					
$\hat{p}$	0.086646	0.092721	0.095707	0.096711	0.097186
$e^{-\theta^* b}$	0.0807	0.0747	0.0720	0.0707	0.0701
$A_y$	1.074	1.241	1.329	1.367	1.386
$A_y$ approxi	1.014	1.159	1.232	1.269	1.287
$A_y$ LB	0.789	0.894	0.947	0.974	0.987
$A_y$ UB	1.305	1.502	1.603	1.654	1.679
s.e.	1.25E-04	9.42E-05	8.49E-05	8.28E-05	8.28E-05
95% CI (lb)	0.08640	0.09254	0.09554	0.09655	0.09702
(ub)	0.08689	0.09291	0.09587	0.09687	0.09735
r.e.	0.001442	0.001016	0.000887	0.000856	0.000852
$P(W_y > b) / P(W > b)$					
ratio	1.36208	1.38777	1.40307	1.40428	1.40505
diff w.r.t. last column	0.04297	0.01728	0.00198	0.00077	0.00000
abs diff	0.04297	0.01728	0.00198	0.00077	0.00000
$A_y/ ho$	1				
ratio	1.27865	1.34842	1.38403	1.39507	1.40028
diff w.r.t. last column	0.12163	0.05186	0.01625	0.00521	0.00000
abs diff	0.12163	0.05186	0.01625	0.00521	0.00000

**Table A.49:** Summary of simulation results for  $M_t/H_2/1$  queue at y = 0 and y = 0.5 as a function of  $1 - \rho$  with base parameter  $(\beta, \gamma, b) = (1, 25, 4)$  using the scaling in (3.39) of the main thesis

### A.3.5 Estimates of the Mean and Standard Deviation

In §A.3.5 we report additional results on experiments to estimate the mean  $E[W_y]$  and standard deviation  $SD[W_y]$  using §3.4.5 of the main thesis. Tables A.50-A.52 report results for the  $M_t/M/1$ model, while Tables A.53 and A.54 report results for the  $(H_2)_t/M/1$  and  $M_t/H_2/1$  models, respectively. The parameters  $n_s$  and  $\delta$  are the parameters for the discrete sum approximations of the integrals;  $n_s$  is the number of terms after truncation and  $\delta$  is the time increment.

**Table A.50:** Estimated mean and standard deviation of the steady-state waiting time in M/M/1 queue as a function of  $1 - \rho$ :  $\mu = 1, \bar{\lambda} = \rho$ 

$1-\rho$	0.16	0.08	0.04	0.02	0.01
n	40,000	40,000	40,000	40,000	40,000
δ	0.001	0.001	0.001	0.001	0.001
b	41	86	173	345	691
$P(W_y > 0)$	0.8396	0.9201	0.9601	0.9799	0.9900
s.e. of $P(W_y > 0)$	6.86E-04	3.71E-04	1.93E-04	9.73E-05	4.98E-05
%95 CI of $P(W_y > 0)$	[0.8383, 0.8410]	[0.9194, 0.9209]	[0.9598, 0.9605]	[0.9797, 0.9801]	[0.9899, 0.9901]
$E[W_y]$	5.249	11.499	23.999	49.000	99.000
s.e. of $E[W_y]$	1.59E-03	1.27E-03	9.51E-04	6.93E-04	4.94E-04
%95 CI of $E[W_y]$	[5.246, 5.252]	[11.497, 11.502]	[23.997, 24.001]	$[48.999, \ 49.001]$	[98.999, 99.001]
$E[W_y W_y > 0]$	6.251	12.497	24.995	50.003	100.005
%95 CI of $E[W_y W_y>0]$	[6.238, 6.265]	[12.485, 12.510]	[24.983, 25.007]	[49.992, 50.014]	[99.994, 100.015]
$E[W_y^2]$	65.624	287.494	1199.982	4899.957	19800.030
s.e. of $E[W_y^2]$	1.50E-02	2.33E-02	3.40E-02	4.92E-02	7.04E-02
%95 CI of $E[W_y^2]$	[65.595, 65.654]	[287.449, 287.540]	[1199.916, 1200.049]	[4899.860, 4900.053]	[19799.892, 19800.168]
$SD[W_y]$	6.170	12.460	24.981	49.990	99.995
$P(W_y > 0)/\rho$	0.9995	1.0002	1.0001	0.9999	1.0000
$(1-\rho)E[W_y]$	0.8398	0.9200	0.9600	0.9800	0.9900
$(1-\rho)SD[W_y]$	0.9873	0.9968	0.9992	0.9998	0.9999
$(1-\rho)E[W_y]/\rho$	0.9998	0.9999	0.9999	1.0000	1.0000
$(1-\rho)SD[W_y]/\rho$	0.8293	0.9171	0.9593	0.9798	0.9899
$(1-\rho)E[W_y W_y>0]$	1.0002	0.9998	0.9998	1.0001	1.0000
$(1-\rho)SD[W_y W_y>0]$	1.0002	1.0000	1.0000	1.0000	1.0000

**Table A.51:** Estimated mean  $E[W_y]$  and standard deviation  $SD[W_y]$  as a function of  $1 - \rho$  for five cases of the  $M_t/M/1$  queue at y = 0.0 and y = 0.5:  $\mu = 1, \bar{\lambda} = \rho$ , base parameter pair  $(\beta, \gamma) = (1, 2.5)$  using the scaling in (3.39). n = 40,000 and  $\delta = 0.001$  for all  $\rho$ 's.

$1-\rho$	0.16	0.08	0.04	0.02	0.01
b	41	86	173	345	691
$\mathbf{y} = 0$					
$P(W_y > 0)$	0.8028	0.9013	0.9507	0.9751	0.9874
s.e. of $P(W_y > 0)$	8.22E-04	5.22E-04	3.36E-04	2.23E-04	1.61E-04
%95 CI of $P(W_y > 0)$	[0.8012, 0.8044]	[0.9003, 0.9024]	[0.9501, 0.9514]	[0.9747, 0.9755]	[0.9870, 0.9877]
$E[W_y]$	4.249	9.416	19.714	40.309	81.624
std of $E[W_y]$	3.07E-03	5.93E-03	1.19E-02	2.38E-02	4.72E-02
%95 CI of $E[W_y]$	[4.243, 4.255]	[9.404, 9.427]	[19.691, 19.737]	$[40.262, \ 40.355]$	[81.531, 81.716]
$E[W_y W_y > 0]$	5.293	10.446	20.736	41.337	82.669
%95 CI of $E[W_y W_y>0]$	[5.275, 5.311]	[10.422, 10.471]	[20.697, 20.775]	$[41.271, \ 41.404]$	[82.549, 82.789]
$E[W_y^2]$	48.677	213.860	892.838	3644.475	14740.585
std of $E[W_y^2]$	3.50E-02	1.40E-01	5.66E-01	2.279	9.123
%95 CI of $E[W_y^2]$	$[48.608, \ 48.745]$	[213.585, 214.135]	[891.729, 893.948]	[3640.009, 3648.942]	[14722.703, 14758.466]
$SD[W_y]$	5.534	11.190	22.454	44.941	89.878
$P(W_y > 0)/\rho$	0.9557	0.9797	0.9903	0.9950	0.9973
$(1-\rho)E[W_y]$	0.6798	0.7532	0.7886	0.8062	0.8162
$(1-\rho)SD[W_y]$	0.8854	0.8952	0.8982	0.8988	0.8988
$(1-\rho)E[W_y]/\rho$	0.8093	0.8187	0.8214	0.8226	0.8245
$(1-\rho)SD[W_y]/\rho$	0.7437	0.8236	0.8622	0.8808	0.8898
$(1-\rho)E[W_y W_y>0]$	0.8469	0.8357	0.8294	0.8267	0.8267
$(1-\rho)SD[W_y W_y>0]$	0.9138	0.9056	0.9026	0.9008	0.8997
$\mathbf{y} = 0.5$					
$P(W_y > 0)$	0.8801	0.9411	0.9714	0.9851	0.9930
s.e. of $P(W_y > 0)$	9.85E-04	6.54E-04	4.51E-04	2.92E-04	2.19E-04
%95 CI of $P(W_y > 0)$	[0.8782, 0.8820]	[0.9399, 0.9424]	[0.9705, 0.9723]	[0.9845, 0.9856]	[0.9926, 0.9934]
$E[W_y]$	6.839	14.927	31.194	63.667	128.411
std of $E[W_y]$	6.42E-03	1.20E-02	2.36E-02	4.69E-02	9.30E-02
%95 CI of $E[W_y]$	$[6.827, \ \ 6.852]$	[14.903, 14.950]	[31.147, 31.240]	$[63.575, \ \ 63.759]$	[128.228, 128.593]
$E[W_y W_y > 0]$	7.771	15.860	32.113	64.632	129.315
%95 CI of $E[W_y W_y>0]$	[7.740, 7.803]	[15.814, 15.907]	[32.036, 32.189]	$[64.501, \ \ 64.763]$	[129.075, 129.554]
$E[W_y^2]$	97.057	427.685	1795.344	7344.665	29673.770
std of $E[W_y^2]$	7.81E-02	0.302	1.207	4.829	19.314
%95 CI of $E[W_y^2]$	[96.904, 97.210]	[427.092, 428.277]	[1792.979, 1797.709]	[7335.201, 7354.129]	[29635.915, 29711.625]
$SD[W_y]$	7.091	14.314	28.676	57.369	114.824
$P(W_y > 0)/\rho$	1.0478	1.0230	1.0119	1.0052	1.0030
$(1-\rho)E[W_y]$	1.0943	1.1941	1.2477	1.2733	1.2841
$(1-\rho)SD[W_y]$	1.1345	1.1451	1.1470	1.1474	1.1482
$(1-\rho)E[W_y]/\rho$	1.3028	1.2980	1.2997	1.2993	1.2971
$(1-\rho)SD[W_y]/\rho$	0.9530	1.0535	1.1011	1.1244	1.1368
$(1-\rho)E[W_y W_y>0]$	1.2434	1.2688	1.2845	1.2926	1.2931
$(1-\rho)SD[W_{u} W_{u}>0]$	1.1301	1.1395	1.1433	1.1452	1.1472

**Table A.52:** Estimated  $E[W_y]$  and  $SD[W_y]$  as a function of  $1 - \rho$  for  $M_t/M/1$  queue at y = 0.0and 0.5:  $\mu = 1, \bar{\lambda} = \rho$ , base parameter pair  $(\beta, \gamma) = (4, 2.5)$  (with longer cycles than in Table A.51) using the scaling in (3.39). n = 40,000 and  $\delta = 0.001$  for all  $\rho$ 's.

$1-\rho$	0.16	0.08	0.04	0.02	0.01
b	41	86	173	345	691
$\mathbf{y} = 0$					
$P(W_y > 0)$	0.7346	0.8679	0.9349	0.9665	0.9828
s.e. of $P(W_y > 0)$	1.28E-03	9.20E-04	6.45E-04	4.75E-04	3.46E-04
%95 CI of $P(W_y > 0)$	[0.7321, 0.7371]	[0.8661, 0.8697]	[0.9336, 0.9361]	[0.9656, 0.9675]	[0.9821, 0.9835]
$E[W_y]$	3.115	7.091	15.097	31.129	63.073
std of $E[W_y]$	5.46E-03	1.10E-02	2.21E-02	4.36E-02	8.71E-02
%95 CI of $E[W_y]$	[3.104, 3.126]	[7.091, 7.134]	[15.054, 15.141]	[31.043, 31.214]	[62.902,  63.243]
$E[W_y W_y > 0]$	4.240	8.171	16.149	32.206	64.178
%95 CI $E[W_y W_y>0]$	[4.211, 4.269]	[8.154, 8.237]	[16.081, 16.218]	[32.087, 32.326]	[63.960, 64.396]
$E[W_y^2]$	33.071	147.266	619.769	2547.465	10295.922
std of $E[W_y^2]$	5.99E-02	2.50E-01	1.028	4.144	0.733
%95 CI of $E[W_y^2]$	[32.954, 33.189]	[146.775, 147.756]	[617.754, 621.784]	[2539, 2555]	[10263, 10328]
$SD[W_y]$	4.834	9.832	19.795	39.730	79.484
$P(W_y > 0)/\rho$	0.8745	0.9433	0.9738	0.9863	0.9927
$(1-\rho)E[W_y]$	0.4984	0.5673	0.6039	0.6226	0.6307
$(1-\rho)SD[W_y]$	0.7735	0.7866	0.7918	0.7946	0.7948
$(1-\rho)E[W_y]/\rho$	0.5933	0.6166	0.6291	0.6353	0.6371
$(1-\rho)SD[W_y]/\rho$	0.6497	0.7237	0.7601	0.7787	0.7869
$(1-\rho)E[W_y W_y>0]$	0.6784	0.6537	0.6460	0.6441	0.6418
$(1-\rho)SD[W_y W_y>0]$	0.8320	0.8116	0.8022	0.7996	0.7973
y = 0.5					
$P(W_y > 0)$	0.9728	0.9883	0.9967	0.9965	0.9993
s.e. of $P(W_y > 0)$	3.61E-03	2.69E-03	2.05E-03	1.16E-03	8.52E-04
%95 CI of $P(W_y > 0)$	[0.9657, 0.9799]	[0.9831, 0.9936]	[0.9927, 1.0000]	[0.9943, 0.9988]	[0.9976, 1.0000]
$E[W_y]$	15.148	33.583	70.677	145.183	294.222
std of $E[W_y]$	5.58E-02	1.13E-01	2.27E-01	4.59E-01	9.15E-01
%95 CI $E[W_y]$	[15.039, 15.258]	[33.362, 33.805]	[70.232, 71.121]	[144.284, 146.081]	[292.428, 296.016]
$E[W_y W_y > 0]$	15.572	33.980	70.909	145.690	294.437
%95 CI of $E[W_y W_y>0]$	[15.348, 15.799]	[33.576, 34.387]	[70.232, 71.643]	[144.458, 146.926]	[292.428, 296.728]
$E[W_y^2]$	331.868	1528.127	6547.951	27092.166	110239.942
std of $E[W_y^2]$	1.023	4.263	17.227	69.632	0.785
%95 CI of $E[W_y^2]$	[329.864, 333.873]	[1519.773, 1536.481]	[6514.187, 6581.716]	[26955, 27228]	[109691, 110787]
$SD[W_y]$	10.119	20.007	39.405	77.551	153.861
$P(W_y > 0)/\rho$	1.1581	1.0743	1.0383	1.0169	1.0094
$(1-\rho)E[W_y]$	2.4237	2.6867	2.8271	2.9037	2.9422
$(1-\rho)SD[W_y]$	1.6190	1.6006	1.5762	1.5510	1.5386
$(1-\rho)E[W_y]/\rho$	2.8854	2.9203	2.9449	2.9629	2.9719
$(1-\rho)SD[W_y]/\rho$	1.3600	1.4725	1.5132	1.5200	1.5232
$(1-\rho)E[W_y W_y>0]$	2.4915	2.7184	2.8364	2.9138	2.9444
$(1-\rho)SD[W_y W_y>0]$	1.5892	1.5830	1.5704	1.5442	1.5371

**Table A.53:** Estimated  $E[W_y]$  and  $SD[W_y]$  as a function of  $1 - \rho$  for  $(H_2)_t/M/1$  at y = 0.0 and 0.5:  $\mu = 1, \bar{\lambda} = \rho$ , base  $(\beta, \gamma) = (1, 2.5)$  with scaling in (3.39). n = 40,000 for all  $\rho$ 's.

$1-\rho$	0.16	0.08	0.04	0.02	0.01
$ heta^*$	0.113	0.0548	0.0270	0.0134	0.00669
δ	0.001	0.002	0.004	0.008	0.016
b	41	86	173	345	691
$\mathbf{y} = 0$					
$P(W_y > 0)$	0.8617	0.9333	0.9668	0.9837	0.9918
s.e. of $P(W_y > 0)$	6.16E-04	3.69E-04	2.39E-04	1.50E-04	1.05E-04
%95 CI of $P(W_y > 0)$	[0.8605, 0.8629]	[0.9326, 0.9340]	[0.9663, 0.9673]	[0.9834, 0.9840]	[0.9916, 0.9920]
$E[W_y]$	6.636	14.715	30.874	63.199	127.735
std of $E[W_y]$	3.25E-03	6.41E-03	1.27E-02	2.53E-02	5.05E-02
%95 CI of $E[W_y]$	[6.629, 6.642]	[14.703, 14.728]	[30.849, 30.899]	$[63.149, \ \ 63.248]$	[127.636, 127.834]
$E[W_y W_y > 0]$	7.701	15.767	31.934	64.246	128.786
%95 CI of $E[W_y W_y > 0]$	[7.683, 7.719]	[15.742, 15.793]	[31.893, 31.976]	$[64.176, \ \ 64.315]$	[128.659, 128.912]
$E[W_y^2]$	110.805	504.944	2148.048	8845.680	35881.950
std of $E[W_y^2]$	5.24E-02	2.14E-01	8.74E-01	3.506	14.028
%95 CI of $E[W_y^2]$	[110.702, 110.908]	[504.524, 505.365]	[2146.336, 2149.760]	[8838.808, 8852.552]	[35854.456, 35909.445]
$SD[W_y]$	8.171	16.983	34.566	69.654	139.878
$P(W_y > 0)/\rho$	1.0258	1.0144	1.0071	1.0038	1.0019
$(1-\rho)E[W_y]$	1.0617	1.1772	1.2350	1.2640	1.2773
$(1-\rho)SD[W_y]$	1.3074	1.3586	1.3827	1.3931	1.3988
$(1-\rho)E[W_y]/\rho$	1.2640	1.2796	1.2864	1.2898	1.2903
$(1-\rho)SD[W_y]/\rho$	1.0982	1.2499	1.3274	1.3652	1.3848
$(1-\rho)E[W_y W_y>0]$	1.2322	1.2614	1.2774	1.2849	1.2879
$(1-\rho)SD[W_y W_y>0]$	1.3318	1.3681	1.3868	1.3950	1.3997
y = 0.5					
$P(W_y > 0)$	0.9123	0.9576	0.9802	0.9897	0.9950
s.e. of $P(W_y > 0)$	6.97E-04	4.26E-04	2.89E-04	1.75E-04	1.31E-04
%95 CI of $P(W_y > 0)$	[0.9109, 0.9136]	[0.9568, 0.9584]	[0.9796, 0.9807]	[0.9894, 0.9901]	[0.9948, 0.9953]
$E[W_y]$	9.615	20.988	43.720	89.079	180.034
std of $E[W_y]$	5.76E-03	1.07E-02	2.07E-02	4.07E-02	8.15E-02
%95 CI of $E[W_y]$	[9.604, 9.626]	[20.967, 21.009]	$[43.679, \ \ 43.760]$	[88.999, 89.159]	[179.874, 180.194]
$E[W_y W_y > 0]$	10.540	21.917	44.603	90.005	180.934
%95 CI of $E[W_y W_y>0]$	[10.512, 10.568]	[21.876, 21.958]	[44.536, 44.671]	[89.893, 90.117]	[180.726, 181.141]
$E[W_y^2]$	185.574	836.287	3534.258	14511.739	58834.208
std of $E[W_y^2]$	9.24E-02	0.362	1.441	5.761	23.019
%95 CI of $E[W_y^2]$	[185.392, 185.755]	[835.578, 836.997]	[3531.433, 3537.082]	[14500.447, 14523.030]	[58789.091, 58879.324]
$SD[W_y]$	9.650	19.895	40.285	81.097	162.548
$P(W_y > 0)/\rho$	1.0860	1.0409	1.0210	1.0099	1.0051
$(1-\rho)E[W_y]$	1.5384	1.6790	1.7488	1.7816	1.8003
$(1-\rho)SD[W_y]$	1.5440	1.5916	1.6114	1.6219	1.6255
$(1-\rho)E[W_y]/\rho$	1.8314	1.8250	1.8216	1.8179	1.8185
$(1-\rho)SD[W_y]/\rho$	1.2970	1.4643	1.5469	1.5895	1.6092
$(1-\rho)E[W_y W_y>0]$	1.6864	1.7533	1.7841	1.8001	1.8093
$(1-\rho)SD[W_y W_y>0]$	1.5375	1.5859	1.6081	1.6201	1.6245

**Table A.54:** Estimated  $E[W_y]$  and  $SD[W_y]$  as a function of  $1 - \rho$  for  $M_t/H_2/1$  at y = 0.0 and 0.5:  $\mu = 1, \bar{\lambda} = \rho$ , base  $(\beta, \gamma) = (1, 2.5)$  with scaling in (3.39). n = 40,000 for all  $\rho$ 's.

$1-\rho$	0.16	0.08	0.04	0.02	0.01
$ heta^*$	0.101	0.0519	0.0263	0.0132	0.00664
δ	0.001	0.002	0.004	0.008	0.016
b	41	86	173	345	691
$\mathbf{y} = 0$					
$P(W_y > 0)$	0.8071	0.9028	0.9511	0.9762	0.9878
s.e. of $P(W_y > 0)$	9.33E-04	5.64E-04	3.41E-04	2.03E-04	1.35E-04
%95 CI of $P(W_y > 0)$	[0.8052, 0.8089]	[0.9017, 0.9039]	[0.9505, 0.9518]	[0.9758, 0.9766]	[0.9876, 0.9881]
$E[W_y]$	6.698	14.779	30.943	63.250	127.753
std of $E[W_y]$	4.38E-03	6.75E-03	1.27E-02	2.53E-02	5.05E-02
%95 CI of $E[W_y]$	[6.689, 6.707]	[14.766, 14.792]	[30.918, 30.968]	$[63.201, \ \ 63.300]$	[127.654, 127.852]
$E[W_y W_y > 0]$	8.299	16.369	32.532	64.794	129.328
%95 CI of $E[W_y W_y > 0]$	[8.270, 8.329]	[16.335, 16.404]	[32.483, 32.581]	$[64.717, \ \ 64.871]$	[129.193, 129.463]
$E[W_y^2]$	126.556	539.343	2217.805	8990.031	36149.733
std of $E[W_y^2]$	7.55E-02	2.36E-01	8.95E-01	3.548	14.131
%95 CI of $E[W_y^2]$	[126.408, 126.704]	[538.880, 539.806]	[2216.051, 2219.559]	[8983.078, 8996.985]	[36122.036, 36177.429]
$SD[W_y]$	9.038	17.914	35.502	70.636	140.815
$P(W_y > 0)/\rho$	0.9608	0.9813	0.9908	0.9961	0.9978
$(1-\rho)E[W_y]$	1.0717	1.1823	1.2377	1.2650	1.2775
$(1-\rho)SD[W_y]$	1.4461	1.4332	1.4201	1.4127	1.4082
$(1-\rho)E[W_y]/\rho$	1.2758	1.2851	1.2893	1.2908	1.2904
$(1-\rho)SD[W_y]/\rho$	1.2148	1.3185	1.3633	1.3845	1.3941
$(1-\rho)E[W_y W_y>0]$	1.3279	1.3096	1.3013	1.2959	1.2933
$(1-\rho)SD[W_y W_y>0]$	1.5004	1.4520	1.4274	1.4158	1.4096
$\mathbf{y} = 0.5$					
$P(W_y > 0)$	0.8771	0.9399	0.9699	0.9847	0.9924
s.e. of $P(W_y > 0)$	9.68E-04	5.87E-04	3.76E-04	2.34E-04	1.64E-04
%95 CI of $P(W_y > 0)$	[0.8752, 0.8790]	[0.9387, 0.9410]	[0.9691, 0.9706]	[0.9842, 0.9851]	[0.9921, 0.9928]
$E[W_y]$	9.558	20.905	43.593	88.977	179.983
std of $E[W_y]$	7.53E-03	1.16E-02	2.11E-02	4.12E-02	8.15E-02
%95 CI of $E[W_y]$	[9.543, 9.573]	[20.882, 20.927]	$[43.552, \ \ 43.635]$	[88.896, 89.058]	[179.823, 180.142]
$E[W_y W_y > 0]$	10.897	22.241	44.948	90.364	181.352
%95 CI of $E[W_y W_y>0]$	[10.857, 10.938]	[22.190, 22.293]	$[44.871, \ 45.025]$	[90.240,  90.488]	[181.133, 181.572]
$E[W_y^2]$	201.796	870.147	3603.439	14652.678	59167.620
std of $E[W_y^2]$	1.30E-01	0.397	1.478	5.833	23.190
%95 CI of $E[W_y^2]$	[201.540, 202.051]	[869.368, 870.926]	[3600.542, 3606.336]	[14641.246, 14664.110]	[59122.168, 59213.072]
$SD[W_y]$	10.509	20.812	41.268	82.072	163.627
$P(W_y>0)/\rho$	1.0442	1.0216	1.0103	1.0047	1.0025
$(1-\rho)E[W_y]$	1.5293	1.6724	1.7437	1.7795	1.7998
$(1-\rho)SD[W_y]$	1.6815	1.6650	1.6507	1.6414	1.6363
$(1-\rho)E[W_y]/\rho$	1.8206	1.8178	1.8164	1.8159	1.8180
$(1-\rho)SD[W_y]/\rho$	1.4124	1.5318	1.5847	1.6086	1.6199
$(1-\rho)E[W_y W_y>0]$	1.7435	1.7793	1.7979	1.8073	1.8135
$(1-\rho)SD[W_{\eta} W_{\eta}>0]$	1.6882	1.6611	1.6469	1.6390	1.6349

### A.3.6 The Impact of the Adjustment in §3.5.3 and §3.5.4

Tables 3.6 and 3.10 of the main thesis for the  $(H_2)_t/M/1$  model would be different if we ignored the adjustment for the exceptional first interarrival time in the rare-event algorithm that were introduced in §3.5.3 and §3.5.4 there. We now show the corresponding tables without this refinement. Consistent with intuition and the fact that the two processes have identical steady-state limits, we see that the difference disappears as  $\rho$  increases. Nevertheless, the difference is noticeable in all cases.

First, Table A.55 shows analog of the results in Table 3.6 of the main thesis for the  $(H_2)_t/M/1$  model.

Second, Table A.56 shows results related to Table 3.10 of the main thesis.

**Table A.55:** Simulation estimates of  $\hat{p} \equiv P(W_y > b) \equiv A_y e^{-\theta^* b}$  in the  $(H_2)_t / M / 1$  model without the factor  $m_{X_1}(\theta^*)$  in (3.28) of the main thesis for y = 0.0 and y = 0.5 as a function of  $1 - \rho$  with base parameter triple  $(\beta, \gamma, b) = (1, 2.5, 4)$  in (3.39) based on 40,000 replications.

$1-\rho$	0.16	0.08	0.04	0.02	0.01
$ heta^*( ho)$	0.113	0.0548	0.0270	0.0134	0.00669
$\hat{p}$ for $y = 0.0$	0.041099	0.047976	0.051467	0.053499	0.054240
$e^{-\theta^*b}$	0.0593	0.0645	0.0670	0.0682	0.0689
$A_y$	0.693	0.744	0.768	0.784	0.788
$A_y^-$	0.504	0.546	0.567	0.577	0.582
$A_y^+$	0.887	0.945	0.973	0.987	0.993
s.e.	4.62 E- 05	4.68E-05	4.82E-05	1.72E-04	4.96E-05
95% CI (lb)	0.04101	0.04788	0.05137	0.05316	0.05414
(ub)	0.04119	0.04807	0.05156	0.05384	0.05434
r.e.	0.001125	0.000975	0.000936	0.003208	0.000914
$P(W_y > b) / P(W > b)$	0.78064	0.78762	0.78945	0.79463	0.79294
diff	0.01230	0.00532	0.00349	-0.00169	0.00000
abs diff	0.01230	0.00532	0.00349	0.00169	0.00000
$A_y/ ho$	0.82476	0.80897	0.80027	0.79995	0.79559
diff	-0.02916	-0.01337	-0.00467	-0.00436	0.00000
abs diff	0.02916	0.01337	0.00467	0.00436	0.00000
$\hat{p}$ for $y = 0.5$	0.075260	0.086414	0.092196	0.095157	0.096491
$e^{- heta^*b}$	0.0593	0.0645	0.0670	0.0682	0.0689
$A_y$	1.269	1.341	1.376	1.394	1.401
$A_y^-$ LB	0.887	0.945	0.973	0.987	0.993
$A_y^+$ UB	1.561	1.635	1.671	1.688	1.696
s.e.	8.03E-05	7.92E-05	8.02E-05	1.83E-04	8.25E-05
95% CI (lb)	0.07510	0.08626	0.09204	0.09480	0.09633
(ub)	0.07542	0.08657	0.09235	0.09552	0.09665
r.e.	0.001067	0.000916	0.000870	0.001921	0.000855
$P(W_y > b) / P(W > b)$	1.42950	1.41863	1.41419	1.41339	1.41060
diff	-0.01891	-0.00803	-0.00360	-0.00279	0.00000
abs diff	0.01891	0.00803	0.00360	0.00279	0.00000
$A_y/ ho$	1 51020	1.45708	1.43357	1.42285	1.41532
	1.01023				
diff	-0.09497	-0.04176	-0.01825	-0.00753	0.00000

**Table A.56:** Estimated mean  $E[W_y]$  and standard deviation  $SD[W_y]$  as a function of  $1 - \rho$  for five cases of the  $(H_2)_t/M/1$  queue without the factor  $m_{X_1}(\theta^*)$  in (3.28) of the main thesis at y = 0.5:  $\mu = 1, \bar{\lambda} = \rho$  and base parameter pair  $(\beta, \gamma) = (1, 2.5)$ .

$1 - \rho$	0.16	0.08	0.04	0.02	0.01
$ heta^*( ho)$	0.113	0.0548	0.0270	0.0134	0.00669
n	40,000	40,000	40,000	40,000	40,000
δ	0.001	0.002	0.004	0.008	0.016
largest $b$	41	86	173	345	691
$P(W_y > 0)$	0.9123	0.9576	0.9802	0.9897	0.9950
s.e. of $P(W_y > 0)$	6.97E-04	4.26E-04	2.89E-04	1.75E-04	1.31E-04
%95 CI of $P(W_y > 0)$	[0.9109, 0.9136]	[0.9568, 0.9584]	[0.9796, 0.9807]	[0.9894, 0.9901]	[0.9948, 0.9953]
$E[W_y]$	9.615	20.988	43.720	89.079	180.034
std of $E[W_y]$	5.76E-03	1.07E-02	2.07 E-02	4.07 E-02	8.15E-02
%95 CI of $E[W_y]$	[9.604,  9.626]	[20.97,  21.01]	[43.68, 43.76]	[80.00, 89.16]	[179.87,  180.19]
$E[W_y W_y > 0]$	10.540	21.917	44.603	90.005	180.934
%95 CI of $E[W_y W_y > 0]$	[10.512, 10.568]	[21.876,  21.958]	[44.54,  44.67]	[89.89,  90.12]	[180.73, 181.14]
$E[W_y^2]$	185.574	836.287	3534.26	14,511.7	58,834.2
std of $E[W_y^2]$	9.24E-02	0.362	1.441	5.761	23.019
%95 CI of $E[W_y^2]$	[185.39, 185.76]	[835.58, 837.00]	[3531.4, 3537.1]	[14,500, 14,523]	$[58,789,\ 58,879]$
$SD[W_y]$	9.650	19.90	40.29	81.10	162.55
$P(W_y > 0)/\rho$	1.0860	1.0409	1.0210	1.0099	1.0051
$(1-\rho)E[W_y]$	1.5384	1.6790	1.7488	1.7816	1.8003
$(1-\rho)SD[W_y]$	1.5440	1.5916	1.6114	1.6219	1.6255
$(1-\rho)E[W_y]/\rho$	1.8314	1.8250	1.8216	1.8179	1.8185
$(1-\rho)SD[W_y]/\rho$	1.2970	1.4643	1.5469	1.5895	1.6092
$(1-\rho)E[W_y W_y>0]$	1.6864	1.7533	1.7841	1.8001	1.8093
$(1-\rho)SD[W_y W_y > 0]$	1.5375	1.5859	1.6081	1.6201	1.6245

## Appendix B

# Matlab Code for Simulation

The Matlab code presented in this section is also available in my github repository github.com/nmresearch/simulation\_code.

### B.1 Chapter 2 Matlab Code

In 2, we developed an efficient simulation algorithm to generate a  $G_t/G_t/1$  queue with a timevarying arrival rate and a time-varying service rate. We display the Matlab code for this algorithm below in Listing B.1. In the following code, we first set parameters for the sinusoidal arrival rate function and for multiple service-rate functions, and then construct the table for the inverse functions for cumulative arrival rate function and cumulative service rate function,  $\Lambda^{-1}$  and  $M^{-1}$ . For each independent simulation experiment, we generate arrival times using the outside table, and generate service times using the recursion and the formula 2.27 in §2.4. Therefore we can calculate waiting times, arrival times and departure times for each customer and finally compute the virtual waiting time and queue length at each time point within our considered time horizon. Mean values across all the simulation experiments are used for value estimation and confidence intervals are also constructed.

**Listing B.1:** Efficient simulation for  $G_t/G_t/1$  queue

#### 1 %parameters

 $_{2}$  time=20000;

```
nogenerate = 40000; % number of iid variables we first generate to
3
      simulate arrivals
  beta = 0.2;
4
  gamma = 0.001;
\mathbf{5}
  rho=0.8; %rho for rate matching control
6
  nu1=0.2; % constant for the first square root service rate control
\overline{7}
  nu2=1; %constant for PSA-based square root service rate control
8
   cl=2*pi/gamma; %length of a cycle
9
  dt = cl / 1000;
10
  l=floor(time/dt); %length of time frame
11
  run=10000;
12
  delta=0.95; % confidence interval
13
  defy=10^{(-6)}; % accuracy for tabling functions
14
   defx=10^{(-7)}; % accuracy for tabling functions
15
16
  %performance measures
17
  W1t_a{1} = zeros(run, l);
18
  W2t_a{1} = zeros(run, 1);
19
  W3t_a{1} = zeros(run, l);
20
  Q1t_a{1} = zeros(run, l);
^{21}
  Q2t_a{1} = zeros(run, l);
22
  Q3t_a{1} = zeros(run, l);
23
24
  \frac{1}{1} for one cycle for generating arrivals and for
25
      generating service times under rate matching control
  xarray1 = (0:defx:cl);
26
```

```
_{27} yarray1=xarray1-beta/gamma*(cos(gamma*xarray1)-1);
```

```
_{28} yvec1 = (0:defy:cl);
```

```
29 xvec1=zeros(1, length(yvec1));
```

```
<sub>30</sub> i=1;
```

```
while j<length(xarray1)+1 && i<length(yvec1)+1
32
  y=yvec1(i);
33
  x=xarray1(j);
34
  if y>yarray1(j)
35
  j = j + 1;
36
  else
37
  xvec1(i)=x;
38
  i = i + 1;
39
  end
40
  end
41
42
43 % table fn. M and M^{-1} for one cycle for generating service times
      under first square root control (M(t) is the integral of (mu(s))
  x array 2 = (0: defx: cl);
44
  yarray2=nu1*sqrt(1+beta*sin(gamma*xarray2));
45
  yarray2=cumsum(defx*yarray2)+yarray1;
46
  yvec2 = (0:defy:yarray2(end));
47
  xvec2=zeros(1, length(yvec2));
^{48}
  i = 1;
49
  j = 1;
50
  while j<length(xarray2)+1 && i<length(yvec2)+1
51
  y=yvec2(i);
52
53 x=xarray2(j);
  if y>yarray2(j)
54
_{55} j=j+1;
  else
56
57 xvec2(i)=x;
  i=i+1;
58
```

```
59 end
```

j = 1;

31

```
60 end
```

```
61
```

```
_{62} %table fn. M and M^{-1} for one cycle for generating service times under second square root control
```

```
63 xarray3 = (0:defx:cl);
```

```
lambda3=1+beta*sin(gamma*xarray3);
```

```
_{65} yarray3=lambda3/2.*(sqrt(1+nu2./lambda3)+1);
```

```
66 yarray3=cumsum(defx*yarray3);
```

```
67 yvec3 = (0:defy:yarray3(end));
```

```
68 xvec3=zeros(1, length(yvec3));
```

69 i=1;

 $_{70}$  j=1;

```
<sup>71</sup> while j<length(xarray3)+1 && i<length(yvec3)+1
```

```
<sup>72</sup> y=yvec3(i);
```

```
73 x=xarray3(j);
```

```
<sup>74</sup> if y>yarray3(j)
```

- $_{75}$  j=j+1;
- 76 else

```
77 xvec3(i)=x;
```

```
78 i=i+1;
```

```
79 end
```

```
80 end
```

```
81
```

```
82 %run independent replications
```

```
83 for s=1:run
```

- $_{84}$  %simulate the customer arrivals  $A\_k$
- <sup>85</sup> U1=rand (1, nogenerate);
- 86 X=-log(U1); %exponential distribution

```
87 T=cumsum(X);
```

```
ss q=floor (T/cl);
```

```
^{89} \quad r = T - c \, l \ast q \, ;
```

```
90 A{1}=xvec1(1+floor(r/defy))+cl*q;
```

91  $A{1}=A{1}(A{1}=ime)$ ; %we only want those arrivals that occur within the time interval we consider

92 num=length(A{1}); %number of arrivals we consider

93

```
94 %generate service requirements S_k
```

```
95 U2=rand(num, 1);
```

96 S= $-\log(U2)$ ; %exponential distribution

97

98 %<br/>simulate begin service time  $B_{-}k\,,$  departure time<br/>  $D_{-}k\,,$  service time  $V_{-}k\,,$  waiting time<br/>  $W_{-}k$ 

```
99 D1{1}=zeros(1,num+1);
```

```
100 D2\{1\} = z eros (1, num+1);
```

- 101  $D3{1} = z e r o s (1, num+1);$
- 102  $B1{1} = z eros(1, num);$

```
103 B2{1} = z e r o s (1, num);
```

```
104 B3{1} = zeros(1, num);
```

```
105 V1{1}=zeros(1,num);
```

```
106 V2{1}=zeros(1,num);
```

```
107 V3{1}=zeros(1,num);
```

```
108 W1{1}=zeros(1, num);
```

```
109 W2{1} = z e r o s (1, num);
```

```
110 W3{1}=zeros(1,num);
```

```
111 for i=1:num
```

- <sup>112</sup> B1{1}(i)=max(D1{1}(i),A{1}(i));
- <sup>113</sup> B2{1}(i)=max(D2{1}(i),A{1}(i));
- <sup>114</sup> B3{1}(i)=max(D3{1}(i),A{1}(i));
- sum1=rho\*S(i)+B1{1}(i)-beta/gamma\*(cos(gamma\*B1{1}(i))-1);

```
V1{1}(i) = xvec1(1 + floor(rem(sum1, cl)/defy)) + sum1 - rem(sum1, cl) - B1{1}(i);
116
       %rate matching control
   sum2=S(i)+yarray2(1+floor(rem(B2{1}(i),c1)/defx))+floor(B2{1}(i)/c1)*
117
       varray2(end);
   V2{1}(i) = xvec2(1 + floor(rem(sum2, yvec2(end)))/defy)) + floor(sum2/yvec2(end)))
118
       ))*cl-B2{1}(i); %first square root service rate control
   sum3=S(i)+yarray3(1+floor(rem(B3{1}(i),c1)/defx))+floor(B3{1}(i)/c1)*
119
       yarray3(end);
   V3{1}(i) = xvec3(1 + floor(rem(sum3, yvec3(end)))/defy)) + floor(sum3/yvec3(end)))
120
       ))*cl-B3{1}(i); %second square root service rate control
  D1\{1\}(i+1)=B1\{1\}(i)+V1\{1\}(i);
121
<sup>122</sup> D2{1}(i+1)=B2{1}(i)+V2{1}(i);
<sup>123</sup> D3{1}(i+1)=B3{1}(i)+V3{1}(i);
<sup>124</sup> W1{1}(i)=B1{1}(i)-A{1}(i);
   W_{1}(i) = B_{1}(i) - A_{1}(i);
125
   W3\{1\}(i)=B3\{1\}(i)-A\{1\}(i);
126
   end
127
128
   % convert to At, Dt, Qt, Wt
129
   At\{1\}=num*ones(1,1);
130
   i = 1;
131
   k = 0;
132
   while k<num
133
   if i * dt < A\{1\}(k+1)
134
   At \{1\}(i) = k;
135
   i = i + 1;
136
   else
137
   k=k+1;
138
   end
139
140
   end
```

```
D1t\{1\}=num*ones(1,1);
141
    i = 1;
142
143 k=1;
    while i < l+1 \&\& k < num+1
144
   if i * dt < D1{1}{1}(k+1)
145
   D1t\{1\}(i)=k-1;
146
   i=i+1;
147
    else
148
_{149} k=k+1;
    end
150
    end
151
   D2t{1}=num*ones(1,l);
152
   i = 1;
153
   k = 1;
154
    while i < l+1 \&\& k < num +1
155
   if i * dt < D2 \{1\}(k+1)
156
   D2t\{1\}(i)=k-1;
157
   i=i+1;
158
    else
159
   k=k+1;
160
    end
161
    end
162
    D3t\{1\}=num*ones(1, 1);
163
   i = 1;
164
   k = 1;
165
    while i < l+1 \&\& k < num +1
166
   if i * dt < D3 \{1\}(k+1)
167
   D3t\{1\}(i)=k-1;
168
    i=i+1;
169
   else
170
```
```
k = k + 1;
171
            end
172
            end
173
            Q1t{1} = At{1} - D1t{1};
174
            Q2t{1} = At{1} - D2t{1};
175
          Q3t{1}=At{1}-D3t{1};
176
           i=floor(A\{1\}(1)/dt); %time before the first arrival
177
          W1t\{1\} = zeros(1, 1);
178
179 W2t\{1\} = z eros(1, 1);
           W3t\{1\} = zeros(1, 1);
180
           i=i+1;
181
           W_{1}_{1}(i:1) = max(W_{1}_{1}(A_{1}_{1}(i:1)) + V_{1}_{1}(A_{1}_{1}(i:1)) - ((i:1)*dt - A_{1}(A_{1}_{1}(A_{1}_{1}(i:1))) - ((i:1)*dt - A_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}_{1}(A_{1}(A_{1}(A_{1}(A_{1}_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1}(A_{1
182
                         \{1\}(i:1)),0);
            183
                         \{1\}(i:1)), 0);
            W3t\{1\}(i:1) = max(W3\{1\}(At\{1\}(i:1)) + V3\{1\}(At\{1\}(i:1)) - ((i:1)*dt - A\{1\}(At\{1\}(i:1)))
184
                         \{1\}(i:1)),0);
185
           %performance measures
186
            W1t_a \{1\}(s, :) = W1t\{1\};
187
            Q1t_a \{1\}(s,:) = Q1t\{1\};
188
           W2t_a \{1\}(s_{,:}) = W2t\{1\};
189
            Q2t_a \{1\}(s,:) = Q2t\{1\};
190
            W3t_a \{1\}(s, :) = W3t\{1\};
191
            Q3t_a \{1\}(s, :) = Q3t\{1\};
192
            end %independent replication loop ends
193
194
```

- 195 %calculate mean and construct confidence intervals from independent experiments
- <sup>196</sup> W1t\_m{1}=mean(W1t\_a{1});

```
<sup>197</sup> Q1t_m{1}=mean(Q1t_a{1});
```

```
<sup>198</sup> W2t_m{1}=mean(W2t_a{1});
```

- <sup>199</sup> Q2t\_m{1}=mean(Q2t\_a{1});
- 200 W3t\_m{1}=mean(W3t\_a{1});
- $_{201} \quad Q3t\_m\{1\} {=} \underline{mean} \left( \, Q3t\_a\left\{1\right\} \right) \, ;$
- 202  $d1_m{1}=sum(W1t_a{1}>0)/run;$
- 203  $d2_m{1}=sum(W2t_a{1}>0)/run;$
- 204  $d3_m{1}=sum(W3t_a{1}>0)/run;$
- 205 W1t\_hw{1}=  $tinv(1-(1-delta)/2, run-1)*std(W1t_a{1})/sqrt(run); %half-width of the confidence interval$
- 206  $Q1t_hw{1} = tinv(1-(1-delta)/2, run-1)*std(Q1t_a{1})/sqrt(run);$
- 207 W2t\_hw{1}= tinv(1-(1-delta)/2,run-1)\*std(W2t\_a{1})/sqrt(run);
- 208  $Q2t_hw{1} = tinv(1-(1-delta)/2, run-1)*std(Q2t_a{1})/sqrt(run);$
- 209 W3t\_hw{1}= tinv(1-(1-delta)/2,run-1)\*std(W3t\_a{1})/sqrt(run);
- 210 Q3t\_hw{1}= tinv(1-(1-delta)/2,run-1)\*std(Q3t\_a{1})/sqrt(run);
- 211 W1t\_CIu{1}=W1t\_m{1}+W1t\_hw{1}; %upper bound of confidence interval
- 212 W1t\_CIl{1}=W1t\_m{1}-W1t\_hw{1}; %lower bound of confidence interval
- $_{213}$  Q1t\_CIu{1}=Q1t\_m{1}+Q1t\_hw{1};
- $_{^{214}} \quad Q1t\_CII\{1\} {=} Q1t\_m\{1\} {-} Q1t\_hw\{1\};$
- $_{^{215}} W2t\_CIu\{1\}=W2t\_m\{1\}+W2t\_hw\{1\};$
- 216 W2t\_CIl{1}=W2t\_m{1}-W2t\_hw{1};
- $_{217} \quad Q2t\_CIu\{1\} = Q2t\_m\{1\} + Q2t\_hw\{1\};$
- $_{218}$  Q2t\_CIl{1}=Q2t\_m{1}-Q2t\_hw{1};
- $_{219}$  W3t\_CIu{1}=W3t\_m{1}+W3t\_hw{1};
- $W3t_CII{1}=W3t_m{1}-W3t_hw{1};$
- $_{221}$  Q3t\_CIu{1}=Q3t\_m{1}+Q3t\_hw{1};
- 222  $Q3t_CII{1}=Q3t_m{1}-Q3t_hw{1};$

## B.2 Chapter 3 Matlab Code

In 3, we developed the rare-event simulation algorithm to estimate tail probabilities of steady-state workload  $W_y$  in  $GI_t/GI/1$  queues, which can be applied to estimate moments of  $W_y$ . In Listing B.2 below, we show the Matlab code for estimating tail probabilities of  $W_y$  at 41 equally spaced positions within a cycle for  $GI_t/GI/1$  queues. We first construct the inverse function  $\tilde{\Lambda}_y^{-1}$  for each position y and then for each y, we generate random variables U and V until the process  $R_n$  in 3.27 reaches level b so that we can use formula 3.28 to estimate tail probabilities. In Listing B.3 below, we display the Matlab code for estimating the first and second moments of  $W_y$  for  $GI_t/GI/1$ queues. We make use of the formula in 3.36 and estimate all the tail probabilities in the equation in a computing time linear in the length of the vector b.

**Listing B.2:** Rare event simulation for tail probabilities of  $W_y$  in  $GI_t/GI/1$  queues

- <sup>1</sup> %parameters
- $_{2}$  beta = 0.2;
- 3 gamma=1;
- $_{4}$  rho=0.8;
- 5 lambda=rho;
- 6 mu=1;

```
7 cl=2*pi/gamma; %length of a cycle
```

- <sup>8</sup> run=5000;
- 9 b=20;

```
10 defy=10^{(-4)}; % accuracy for tabling functions
```

<sup>11</sup> defx= $10^{(-5)}$ ; % accuracy for tabling functions

```
_{12} delta = 0.95;
```

- pos = (0: cl / 40: cl); % y's in a cycle
- <sup>14</sup> P=zeros(length(pos),run);
- <sup>15</sup> number=floor (10\*b/(1/lambda-1));
- <sup>16</sup> X= $-1/\text{lambda} \cdot \log(\text{rand}(\text{number}, \text{run}));$
- $Y = -\log(rand(number, run));$

```
18
```

```
% for each position y in a cycle
19
   for r=1:length(pos)
20
^{21}
  %table fn. \Lambda_y^{-1} (normalized) for one cycle
22
   xarray1 = (0:defx:cl);
23
  yarray1=xarray1+beta/gamma*(cos(gamma*(xarray1-pos(r)))-cos(gamma*pos(r
^{24}
      )));
  yvec1 = (0:defy:cl);
25
  xvec1=zeros(1,length(yvec1));
26
  i = 1;
27
  j = 1;
28
  while j<length(xarray1)+1 && i<length(yvec1)+1
29
  y=yvec1(i);
30
  x=xarray1(j);
31
  if y>yarray1(j)
32
  j = j + 1;
33
  else
34
  xvec1(i)=x;
35
  i = i + 1;
36
  end
37
  end
38
39
  for s=1:run
40
  S1 = 0;
41
_{42} S2=0;
_{43} S3=0;
_{44} j=0;
  while S1-S3<b
45
  j=j+1;
46
  if j>number
47
```

```
Z = rand(1, 2);
48
  Z(1) = -1/lambda \cdot log(Z(1));
49
  Z(2) = -\log(Z(2));
50
  S1=S1+Z(1);
51
  S2=S2+Z(2);
52
   else
53
  S1=S1+X(j,s);
54
  S2=S2+Y(j,s);
55
  end
56
  S3=xvec1(1+floor(rem(S2,cl)/defy))+S2-rem(S2,cl);
57
  end
58
  P(r, s) = exp(-(mu-lambda)*(S1-S2));
59
  end
60
  end
61
62
  P_{-}mean=mean(P');
63
   P_{std} = std(P') / sqrt(run);
64
  P_hw=norminv(1-(1-delta)/2)*P_std;
65
  P_re=P_std./P_mean;
66
  P_approx = exp(-(mu-lambda)*b)*ones(1, length(pos));
67
  A=P_mean./P_approx;
68
  A_approx=rho*exp(-(mu-lambda)*beta/gamma*cos(gamma*pos));
69
  ALB=rho*exp(-(mu-lambda)*beta/gamma*(cos(gamma*pos)+1));
70
  A_UB=rho * exp(-(mu-lambda) * beta/gamma*(cos(gamma*pos)-1));
71
```

**Listing B.3:** Rare event simulation for moments of  $W_y$  in  $GI_t/GI/1$  queues

```
1 %parameters
```

- $_{2}$  beta = 0.2;
- <sup>3</sup> gamma=0.1;
- $_4~rho\!=\!0.8;~\% rho$  for rate matching control

```
lambda=rho;
5
  mu=1;
6
   cl=2*pi/gamma; %length of a cycle
7
   run = 40000;
8
9
   defy=10^{(-4)}; % accuracy for tabling functions
10
   defx = 10^{(-5)}; % accuracy for tabling functions
11
  Delta = 0.95;
12
  delta = 0.001;
13
  b=[0:35000]*delta; %35 rho=0.8; 43 rho0.84; 86 rho0.92; 173 rho0.96;
14
      345 rho0.98; 691 rho0.99
  pos = (0: cl / 40: cl);
15
16
   P_{positive=zeros}(length(pos),run);
17
   P_{\text{-estimate}=\text{zeros}(2, \text{length}(b));
18
<sup>19</sup> EW=zeros(length(pos),run);
  EW2=zeros(length(pos),run);
20
  number=floor (10*b(end)/(1/lambda-1));
21
22
  %table fn. \Lambda_y^{-1} (normalized) for one cycle
23
   xarray1 = (0:defx:cl);
24
   yarray1=xarray1+beta/gamma*(cos(gamma*(xarray1-pos(1)))-cos(gamma*pos
25
      (1)));
  yarray2=xarray1+beta/gamma*(cos(gamma*(xarray1-pos(2)))-cos(gamma*pos
26
      (2)));
  yvec1 = (0:defy:c1);
27
  xvec1=zeros(1,length(yvec1));
28
  i = 1;
29
  j = 1;
30
  while j<length(xarray1)+1 && i<length(yvec1)+1
31
```

```
y=yvec1(i);
32
   x=xarray1(j);
33
   if y>yarray1(j)
34
  j=j+1;
35
   else
36
  xvec1(i)=x;
37
  i=i+1;
38
   end
39
   end
40
^{41}
   xvec2=zeros(1,length(yvec1));
42
  i = 1;
43
_{44} j=1;
   while j<length(xarray1)+1 && i<length(yvec1)+1
45
  y=yvec1(i);
46
  x=xarray1(j);
47
  if y>yarray2(j)
^{48}
  j=j+1;
49
   else
50
  xvec2(i)=x;
51
   i=i+1;
52
   end
53
   end
54
55
   for s=1:run
56
57 P=zeros(length(b),2);
58 X=-1/\text{lambda} \cdot \log(\text{rand}(\text{number}, 1));
  Y = -\log(rand(number, 1));
59
   S1 = cumsum(X);
60
   S2=cumsum(Y);
61
```

```
ind=1+floor(rem(S2, cl)/defy);
62
   S3_1=xvec1(ind)'+S2-rem(S2,c1);
63
   S3_2=xvec2(ind)'+S2-rem(S2,cl);
64
65
   i = 1;
66
   j = 1;
67
   while j < length(b) + 1;
68
   if S1(i)-S3_1(i)<b(j)
69
  i = i + 1;
70
   else
71
<sup>72</sup> P(j, 1) = \exp(-(mu-lambda) * (S1(i)-S2(i)));
  j = j + 1;
73
   end
74
   end
75
76
   i = 1;
77
   j = 1;
78
   while j < length(b) + 1;
79
   if S1(i) - S3_2(i) < b(j)
80
   i=i+1;
^{81}
   else
82
<sup>83</sup> P(j,2) = \exp(-(mu-lambda) * (S1(i)-S2(i)));
  j=j+1;
^{84}
   end
85
   end
86
87
   P_estimate=P_estimate+P';
88
   P_{-}positive (:, s)=P(1,:)';
89
  EW(1,s) = sum(delta * P(:,1)) + P(length(b),1) / (mu-lambda);
90
<sup>91</sup> EW(2,s)=sum(delta*P(:,2))+P(length(b),2)/(mu-lambda);
```

- $_{92} EW2(1,s) = sum(2*P(:,1)`.*b*delta) + 2*(b(end)/(mu-lambda) + 1/(mu-lambda)^2) \\ *P(length(b),1);$

94 end

- $P_{est} = P_{est} = P_{est} = P_{est}$
- 96 Ppos\_final=mean(P\_positive ')
- 97 Ppos\_std=std(P\_positive')/sqrt(run)
- 98  $Ppos_hw=norminv(1-(1-Delta)/2)*Ppos_std;$
- 99 Ppos\_CI=[Ppos\_final-Ppos\_hw, Ppos\_final+Ppos\_hw]
- 100  $EW_{final}=mean(EW')$
- 101  $EW_std=std(EW')/sqrt(run)$
- 102 EW\_hw=norminv $(1-(1-\text{Delta})/2)*\text{EW}_std;$
- <sup>103</sup> EW\_CI=[EW\_final-EW\_hw, EW\_final+EW\_hw]
- 104  $EW2_final=mean(EW2')$
- 105  $EW2\_std=std(EW2')/sqrt(run)$
- 106 EW2\_hw=norminv $(1-(1-\text{Delta})/2)*\text{EW2\_std};$
- <sup>107</sup> EW2\_CI=[EW2\_final-EW2\_hw, EW2\_final+EW2\_hw]
- 108 SDW=(EW2\_final-EW\_final.^2).^0.5

## B.3 Chapter 5 Matlab Code

In 5, we extended the rare-event simulation algorithm to estimate tail probabilities and moments of  $W_y$  in  $GI_t/GI_t/1$  queues. In Listing B.4 below, we show the code for estimating moments of  $W_y$  for  $GI_t/GI_t/1$  queues. We estimate tail probabilities using formula 5.24 and the algorithm in §5.5.2 and then calculate moments using equation 3.36.

**Listing B.4:** Rare event simulation for moments of  $W_y$  in  $GI_t/GI_t/1$  queues

- 1 %parameters
- $_{2}$  beta = 0.2;
- 3 gamma = 0.1;

```
4 rho=0.8; %rho for rate matching control
  lambda=rho;
\mathbf{5}
6 \text{ mu}=1;
  cl=2*pi/gamma; %length of a cycle
7
  run = 40000;
8
  defy=10^{(-4)}; % accuracy for tabling functions
9
  defx = 10^{(-5)}; % accuracy for tabling functions
10
  Delta = 0.95;
11
  delta = 0.001;
12
  b=[0:35000]*delta; %35 rho=0.8; 43 rho0.84; 86 rho0.92; 173 rho0.96;
13
     345 rho0.98; 691 rho0.99
14
  pos = [0: cl / 40: cl];
15
16
  P_positive=zeros(length(pos),run);
17
  EW=zeros(length(pos),run);
18
  EW2=zeros(length(pos),run);
19
  number=floor (10*b(end)/(1/lambda-1));
20
21
22
  23
  xarray1 = (0:defx:cl);
24
  yarray=zeros(length(pos),length(xarray1));
25
  yvec1 = (0:defy:cl);
26
  xvec=zeros(length(pos),length(yvec1));
27
28
  for k=1:length(pos)
29
  yarray(k,:)=xarray1+beta/gamma*(cos(gamma*(xarray1-pos(k)))-cos(gamma*
30
     pos(k)));
_{31} i=1;
```

```
_{32} j=1;
```

```
while j<length(xarray1)+1 && i<length(yvec1)+1
```

```
34 y=yvec1(i);
```

```
35 x=xarray1(j);
```

```
_{36} if y>yarray(k,j)
```

- 37 j=j+1;
- зв else

```
39 xvec(k, i)=x;
```

- $_{40}$  i=i+1;
- 41 end
- 42 end
- 43 end

```
44 clearvars yarray
```

```
45
```

```
46 for s=1:run
```

```
47 P=zeros(length(b), length(pos));
```

```
_{48} \ X=-1/lambda*log(rand(number,1));
```

```
_{49} \hspace{0.1 in} Y\hspace{-1mm}=\hspace{-1mm} -\hspace{-1mm} \log\left( \hspace{0.1 mm} \text{number} \hspace{0.1 mm}, 1 \hspace{0.1 mm} \right) \hspace{0.1 mm} ) \hspace{0.1 mm} ;
```

```
50 S1=cumsum(X);
```

- 51 S2=cumsum(Y);
- ind=1+floor(rem(S2,cl)/defy);

```
S3 = zeros(number, length(pos));
```

```
54
```

```
55 for k=1:length(pos)
```

```
56 tic
```

```
57 S3(:,k)=xvec(k,ind)'+S2-rem(S2,cl);
```

```
S3(:,k)=S3(:,k)-beta/gamma*(cos(gamma*pos(k))-cos(gamma*(S3(:,k)-pos(k))));
```

```
{}_{59} \ \ Mb=b-beta/gamma*(cos(gamma*(pos(k)+b))-cos(gamma*pos(k)));
```

```
_{60} i=1;
```

```
i = 1;
61
   while i < length(b) + 1;
62
   if S1(i)-S3(i,k)<Mb(j)
63
  i=i+1;
64
   else
65
  P(j,k) = \exp(-(mu-lambda) * (S1(i)-S2(i)));
66
  j = j + 1;
67
  end
68
  end
69
  EW(k, s) = sum(delta * P(:, k)) + P(length(b), k) / (mu-lambda);
70
  EW2(k, s) = sum(2*P(:, k)' \cdot *b*delta) + 2*(b(end)/(mu-lambda) + 1/(mu-lambda)^2)
71
      *P(length(b),k);
  end
72
   P_{-}positive (:, s)=P(1,:)';
73
  end
74
75
   Ppos_final=mean(P_positive ')
76
   Ppos_std=std (P_positive ') / sqrt (run)
77
  Ppos_hw=norminv(1-(1-Delta)/2)*Ppos_std;
78
  Ppos_CI=[Ppos_final-Ppos_hw, Ppos_final+Ppos_hw]
79
  EW_{final}=mean(EW')
80
  EW_std=std (EW') / sqrt (run)
81
  EW_hw=norminv(1-(1-Delta)/2)*EW_std;
82
  EW_CI=[EW_final-EW_hw, EW_final+EW_hw]
83
  EW2_final=mean(EW2')
84
  EW2_std=std (EW2') / sqrt (run)
85
  EW2\_hw=norminv(1-(1-Delta)/2)*EW2\_std;
86
  EW2_CI=[EW2_final-EW2_hw, EW2_final+EW2_hw]
87
  SDW = (EW2_final - EW_final^2) \cdot 0.5
88
```