E-Companion

EC.1. Overview

In this e-companion we supplement the main paper. In §EC.2 we provide the stationary test results for our call center data. In §EC.3 we supplement §3 and §5.3 by showing how idealized simulation models for the call center example were constructed. In §EC.4 we elaborate on the confidence intervals for the mean wait in the transient $M/M/1$ queue discussed in §6.1. We show that the distribution of the estimates are nor too far from being normal, and show ways to adjust the confidence interval halfwidths to get targeted coverage.

Additional material is in a report available online from the authors' web pages (Kim and Whitt 2012). Following an introduction in §1, in §2 we present additional information about the call center data used. In §3 we discuss an alternative way to construct the confidence intervals using the method of batch means and in §4 we elaborate on the bias discussed in §5, discussing the bias in the estimator $\bar{W}(t)$ in (1) for W in a stationary setting in §4.1. In §4.2 we discuss the bias in the alternative estimator $\bar{W}_{L,\lambda}(t)$ in (3) for $E[W(t)]$ in a nonstationary setting and W in a stationary setting. In §5 we introduce an alternative algorithm to estimate confidence intervals in approximately stationary intervals by batch means, exploiting Theorem 2. However, this algorithm did not improve the estimation for the call center example, so we do not emphasize it. The negative result itself is interesting, because it is natural to consider such alternatives. It remains to be seen if the new algorithm can be useful in other contexts. Finally, in §6 we present additional figures and tables.

EC.2. Results for the Stationarity Tests

In this section, we provide detailed results for the three tests for stationarity applied to the call center data in §3.2. The three test are the turning points test, the difference-sign test and the rank test for randomness, all discussed on p. 312 of Brockwell and Davis (1991). Given a sequence of observations $\{y_1, ..., y_n\}$, these tests basically check the hypothesis that $\{y_i\}$ is an i.i.d. sequence.

For *the turning points test*, we say that the data has a turning point at time i if $y_{i-1} < y_i$ and $y_i > y_{i+1}$ or $y_{i-1} > y_i$ and $y_i < y_{i+1}$. Let T be the number of turning points of the sequence $\{y_i\}$. If $\{y_i\}$ are observations of an i.i.d. sequence, the expected number of turning points is $\mu_T = 2(n-2)/3$ with variance $\sigma_T^2 = (16n-29)/90$, and T is approximately $N(\mu_T, \sigma_T^2)$. A large value of $T - \mu_T$ means the sequence is fluctuating more rapidly than expected for a random sequence. A large (in magnitude) negative value of $T - \mu_T$ suggests a positive correlation between neighboring observations. *The difference-sign test* counts the number (S) of values of i such that $y_i > y_{i-1}$,

$i = 2, ..., n$. Under the i.i.d. sequence assumption, $\mu_S = \frac{1}{2}(n-1)$ with variance $\sigma_S^2 = (n+1)/12$, and S is approximately $N(\mu_S, \sigma_S^2)$. *The rank test* counts the number (P) of pairs (i, j) such that $y_j > y_i$, $j > i$, $i = 1, ..., n-1$. If $\{y_i\}$ are observations of an i.i.d. sequence, then $\mu_P = \frac{1}{4}n(n-1)$ with variance $\sigma_P^2 = n(n-1)(2n+5/8)$, and P is approximately $N(\mu_P, \sigma_P^2)$. A large positive (negative) value of $S - \mu_S$ or $P - \mu_P$ indicates an increasing (decreasing) trend in $\{y_i\}$.

We apply the aforementioned three tests on arrival counts over successive subintervals of varying lengths, 1, 5 and 10 minutes. Table EC.1 provides the result. If we use significance level $\alpha = 0.95$, we have $\Phi_{1-\alpha/2} = 1.96$. The test results depend on the subinterval length, but they mainly accept the hypothesis that the arrival counts are i.i.d sequence in $[10, 16]$, but reject it outside of the interval $[10, 16]$. The results of the rank test, which is particularly useful for detecting a linear trend, is especially strong.

Table EC.1 Testing for stationarity: the assumption that $\{y_i\}$ are observations from an i.i.d sequence is rejected if $|\cdot - \mu_\cdot|/\sigma_\cdot > \Phi_{1-\alpha/2}$ where $\Phi_{1-\alpha/2}$ is the $1-\alpha/2$ percentage point of a standard normal distribution.

Subinterval	Interval	T	μ_t	$\lvert T-\mu_T\rvert/\sigma_T$	S	μ_S	$\lvert S-\mu_S\rvert/\sigma_S$	P	μ_P	$\lvert P-\mu_P\rvert/\sigma_P$
1	$[6,10]$	112	158.7	7.17	96	119.5	5.24	21517	14340.0	3.85
	$[10,16]$	207	238.7	3.97	159	179.5	3.74	29207	32310.0	0.91
	$[16,23]$	179	278.7	11.56	144	209.5	11.06	15294	43995.0	6.66
5	$[6,10]$	21	30.7	3.37	25	23.5	0.74	968	564.0	2.39
	$[10,16]$	49	46.7	0.66	31	35.5	1.82	1208	1278.0	0.23
	$[16,23]$	44	54.7	2.79	36	41.5	2.07	567	1743.0	3.03
10	$[6,10]$	12	14.7	1.34	14	11.5	1.73	244	138.0	1.75
	$[10,16]$	19	22.7	1.49	18	17.5	0.28	299	315.0	0.15
	$[16,23]$	24	26.7	1.00	16	20.5	2.38	109	430.5	2.32

EC.3. Details for the Simulation Experiments in Section 4

In §1.2.2, we suggest applying simulation to study how the estimation procedures proposed in this paper work for an idealized queueing model of the system. In this section, we describe in detail how we construct the idealized simulation models in §3.2 for the call center example.

EC.3.1. Fitted Arrival Rate Function

In order to construct an idealized model to evaluate how the estimation procedure perform for the actual data, we need to mimic the behavior of the actual arrival process of the call center as much as possible. Given arrival rates measured in units of 10 minutes over the working day $[6, 23]$, we fit the arrival rate function to a continuous piecewise-linear function, with one increasing piece over $[6, 10]$, a constant piece over $[10, 16]$ and two decreasing linear pieces over $[16, 18]$ and $[18, 23]$. We

force two extra constraints: (i) the arrival rate starts and ends at 0, and hence the arrival rate is 0 at $t=6$ and $t=23$ and (ii) the arrival rate of successive pieces agree at endpoints. Figure EC.1 illustrates the result. The exact arrival rate function is given by

$$\lambda(t) = \begin{cases} 140(t-6) & \text{on } [6,10], \\ 560 & \text{on } [10,16], \\ 560-230(t-16) & \text{on } [16,18], \\ 100-20(t-18) & \text{on } [18,23]. \end{cases} \tag{EC.1}$$

We used the same 1000 arrival sample paths for all the infinite-server and finite-server models used in this paper. We generated these arrival processes by thinning a homogeneous arrival process with rate $\lambda^* = 560$. The homogeneous Poisson process generates potential arrivals. We then let a potential arrival at time t be an actual arrival in the nonhomogeneous arrival process with probability $\lambda(t)/\lambda^*$.

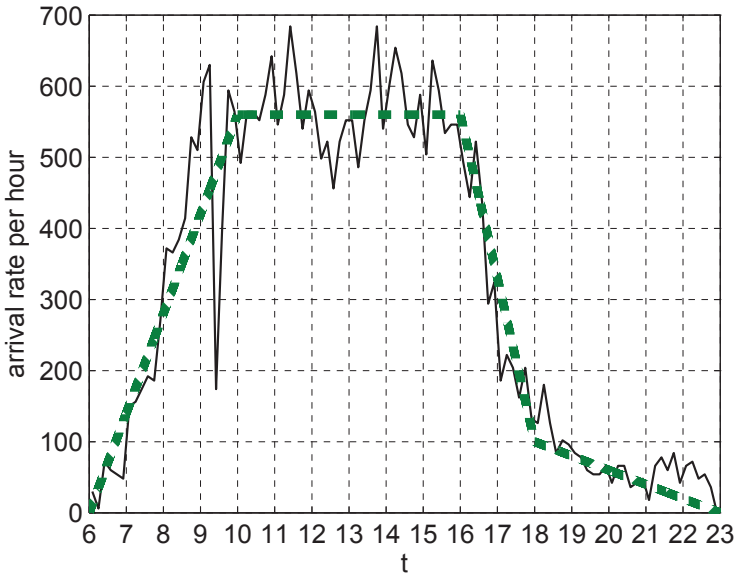

Figure EC.1 Fitted arrival rate function for the arrivals at the call center on May 25, 2001.

EC.3.2. Histogram of the Waiting Time Distribution in the Call Center

We assumed that all the service times were i.i.d. with a distribution obtained to match the observed waiting time distribution. In particular, we consider the waiting times (time spent in the system) over the interval $[10,16]$ on May 25, 2001. Figure EC.2 shows a histogram of the waiting times. An exponential approximation with mean 3.38 was found to be a good fit and was used. Thus, service times were generated according to an exponential distribution with mean 3.38 for $1,000$

replications. The same set of generated service times were used for all the infinite-server and finite-server models used in this paper. (Consistent with many call center empirical studies Brown et al. (2005), a lognormal distribution with mean 3.38 and squared coefficient of variation $c^2 = 1.017$ was also found to be a good fit, even better in the neighborhood of the origin, but we were not concerned that the mode falls to the right of 0. Simulation shows that the results are not significantly altered by using the fitted lognormal distribution. The exponential distribution makes the staffing easier for the $M_t/GI/s_t$ model; e.g., by applying Eick et al. (1993a).)

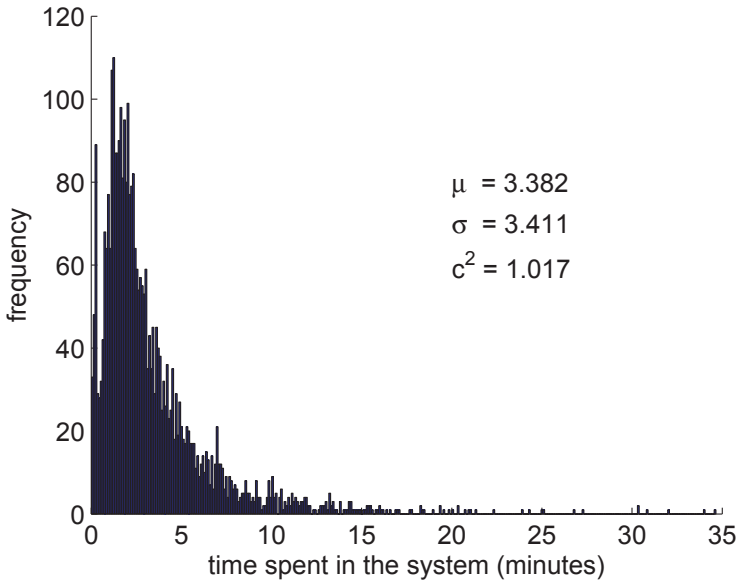

Figure EC.2 The histogram (empirical distribution) of the times spent in the system of all arrivals during the interval [10, 16] on May 25, 2001.

EC.3.3. Staffing for the $M_t/GI/s_t$ Models

For the call center example, we have data on the arrival times and waiting times as well as the number in system $L(s)$, $0 \le s \le t$, but we do not have data on the staffing and the complex call routing. Thus, in order to evaluate the estimation procedures, we simulate the single-class single-service-pool $M_t/GI/\infty$ IS model and associated $M_t/GI/s_t$ models with time-varying staffing levels chosen to yield good performance. Specifically, we use the fitted arrival rate function (from §EC.3.1) and assume service times are exponentially distributed with mean 3.38 (from §EC.3.2). We then compute $m(t)$, the *offered load* which is the time-varying mean number of busy servers in the IS model, using formulas (6) and (7) of Jennings et al. (1996). Finally, the staffing function $s(t)$ is determined by the SRS formula using a range of Quality-of-Service (QoS) parameters, $\beta = 2.5, 2.0, 1.5, 1.0$. Figure EC.3 illustrates the offered load as well as staffing levels for different values of β.

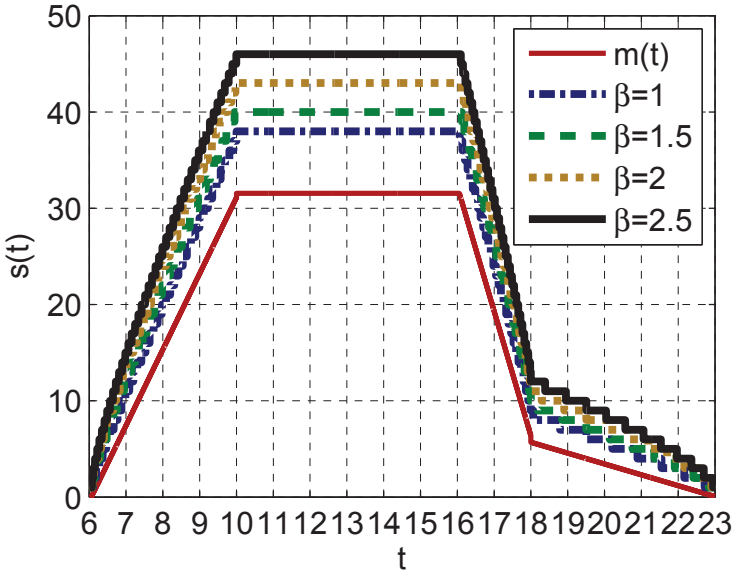

Figure EC.3 The offered load, $m(t)$, and staffing levels, $s(t)$, for different values of β.

EC.3.4. Sample Paths for Different Values of β

One way to diagnose whether simulated $M_t/GI/s_t$ models with different values of β are working as expected is to plot some sample paths of the number in the system, $L(t)$, along with the staffing levels, $s(t)$. In Figures EC.4 - EC.7, we provide one single sample path of each $M_t/GI/s_t$ model with different values of $\beta = 2.5, 2.0, 1.5, 1.0$ to illustrate that our simulation models work as expected.

EC.4. More on Confidence Intervals for the $M/M/1$ Example

In this section, we provide more details on confidence intervals for the mean wait in the transient $M/M/1$ queue, discussed in §6.1. Recall that we have 10 i.i.d. samples of the same $M/M/1$ model over the interval $[0, 10]$, starting empty at time 0. To see how the sample average approach in §6.3 can be applied to estimate CIs for the refined estimator in (41), CI coverage was studied by performing $1,000$ replications of the entire experiment.

In Table 4, we observed that the unrefined estimator $\bar{W}_{L,\lambda}(t)$ in (3) does a very poor job in estimating the mean wait because of the bias. The performance of the refined estimator $\bar{W}_{L,\lambda,r}(t)$ in (28) and the direct estimator $\bar{W}(t)$ is not perfect, either, yielding coverage of about 90% instead of the targeted 95%. To see whether this coverage issue is due to the residual skewness of the estimates, we perform the Shapiro-Francia normality test, discussed on p. 314 of Brockwell and Davis (1991). The Shapiro-Francia normality test is specifically designed to address issues related with large sample sizes. Table EC.2 gives the test results; normality is rejected at 5% significance level in most cases. We also plot the histogram of the estimators as well as their normality plots,

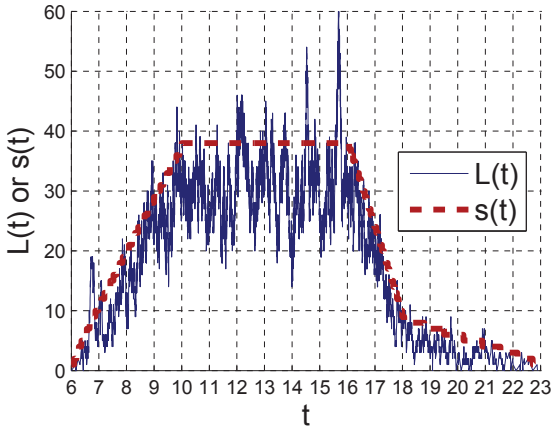

Figure EC.4 One sample path of $M_t/GI/s_t$ with $\beta = 1.0$.

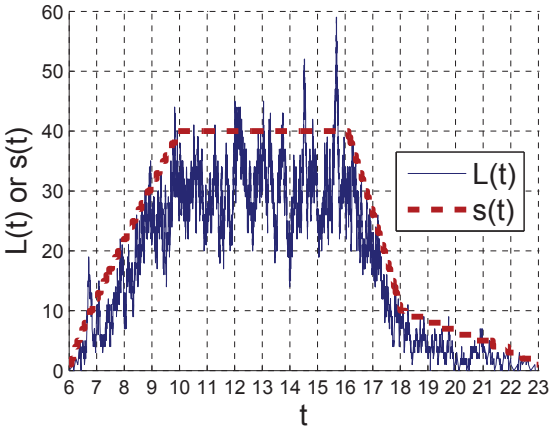

Figure EC.5 One sample path of $M_t/GI/s_t$ with $\beta = 1.5$.

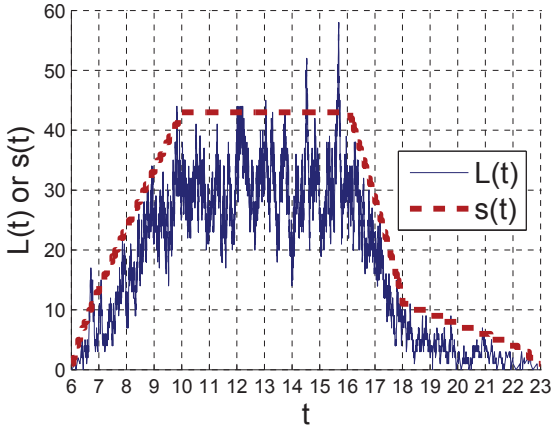

Figure EC.6 One sample path of $M_t/GI/s_t$ with $\beta = 2.0$.

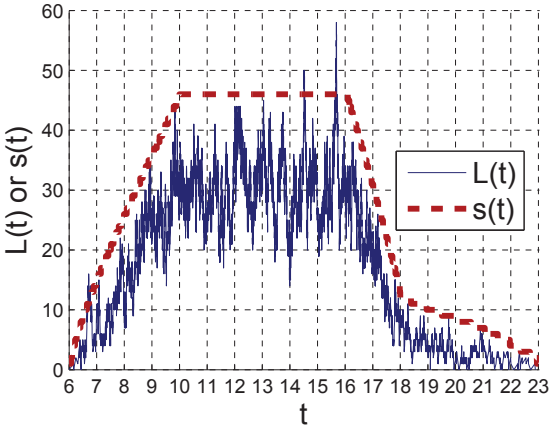

Figure EC.7 One sample path of $M_t/GI/s_t$ with $\beta = 2.5$.

from which we can see evidence of the heavy tails (skewness). However, we see that these are not extreme examples of non-normality and skewness.

Table EC.2 Shapiro-Francia normality test (discussed on p. 314 of Brockwell and Davis (1991)) of the direct and indirect mean waiting time estimator values over 1000 replications.

	$\bar{W}(t)$		$\bar{W}_{L,\lambda}(t)$		$\bar{W}_{L,\lambda,r}(t)$	
λ	R^2	p-value	R^2	p-value	R^2	p-value
0.7	0.9794	0.0000	0.9941	0.0011	0.9813	0.0000
1.0	0.9892	0.0000	0.9936	0.0006	0.9905	0.0000
2.0	0.9945	0.0019	0.9990	0.8284	0.9977	0.1562

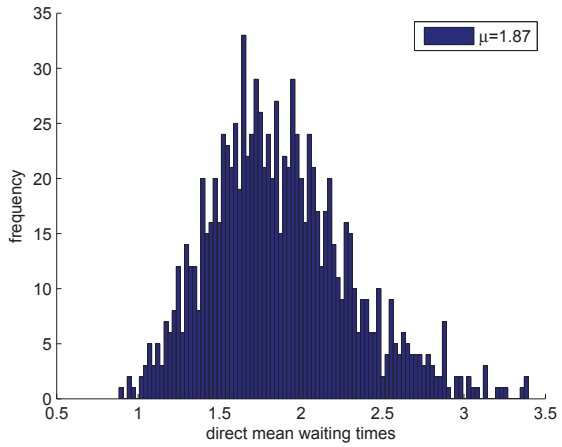

Figure EC.8 Histogram of $\bar{W}(t)$ for the $M/M/1$ model with $\lambda = 0.7$.

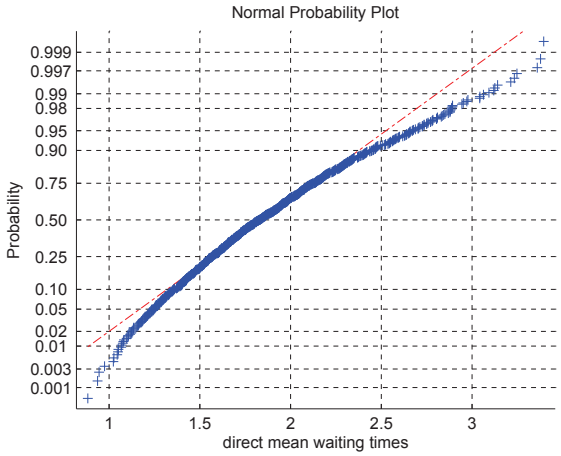

Figure EC.9 Normality plot of $\bar{W}(t)$ for the $M/M/1$ model with $\lambda = 0.7$.

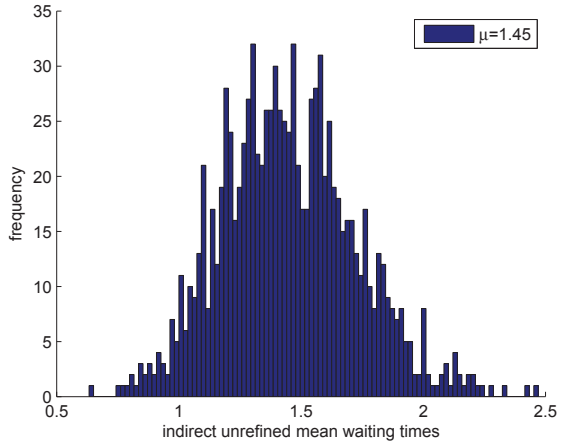

Figure EC.10 Histogram of $\bar{W}_{L,\lambda}(t)$ for the $M/M/1$ model with $\lambda = 0.7$.

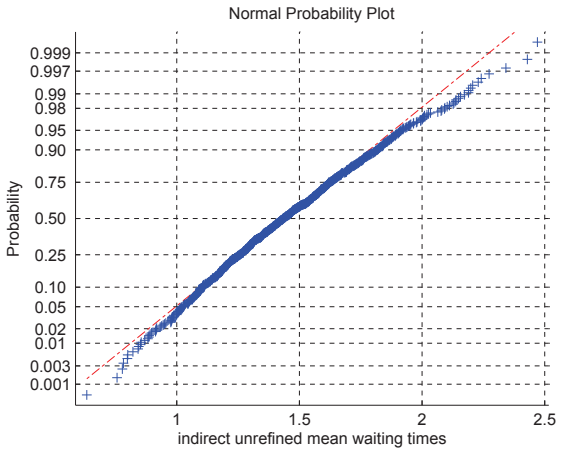

Figure EC.11 Normality plot of $\bar{W}_{L,\lambda}(t)$ for the $M/M/1$ model with $\lambda = 0.7$.

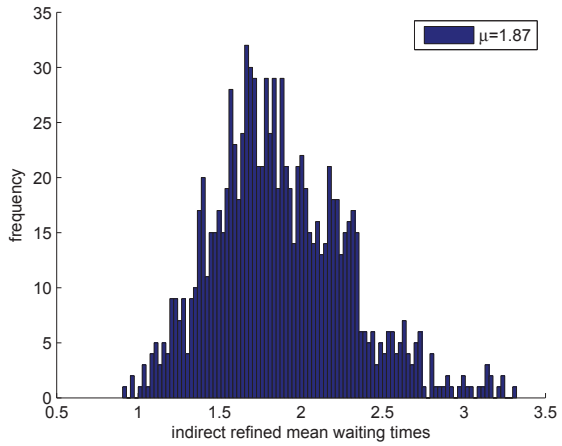

Figure EC.12 Histogram of $\bar{W}_{L,\lambda,r}(t)$ for the $M/M/1$ model with $\lambda = 0.7$.

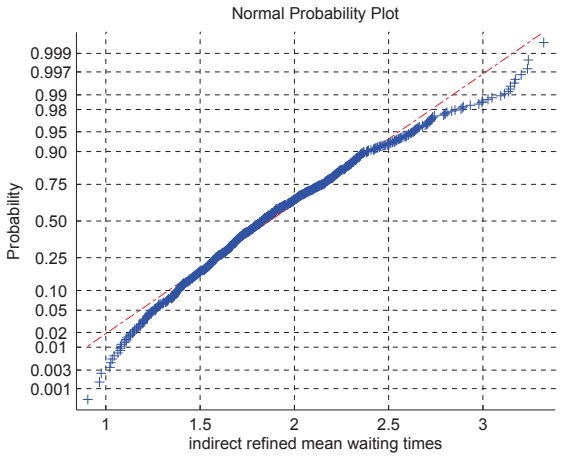

Figure EC.13 Normality plot of $\bar{W}_{L,\lambda,r}(t)$ for the $M/M/1$ model with $\lambda = 0.7$.

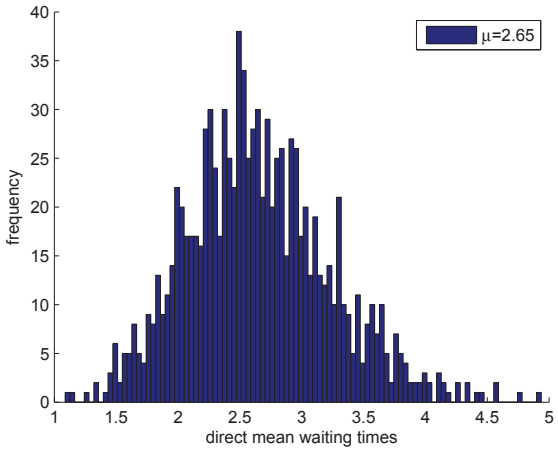

Figure EC.14 Histogram of $\bar{W}(t)$ for the $M/M/1$ model with $\lambda = 1.0$.

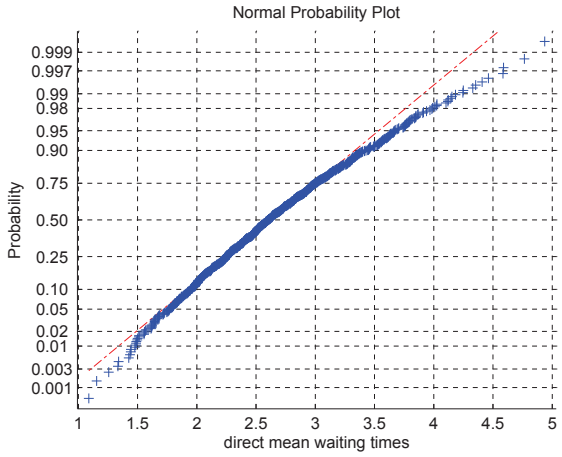

Figure EC.15 Normality plot of $\bar{W}(t)$ for the $M/M/1$ model with $\lambda = 1.0$.

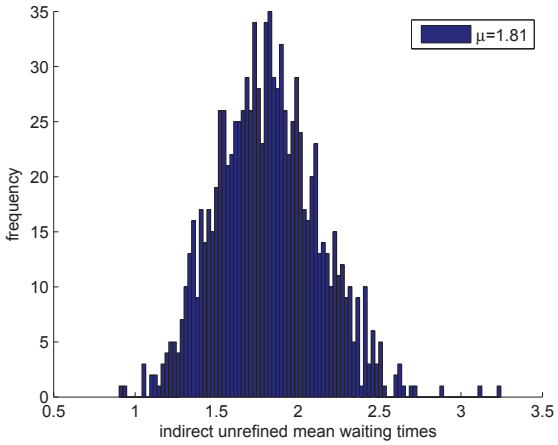

Figure EC.16 Histogram of $\bar{W}_{L,\lambda}(t)$ for the $M/M/1$ model with $\lambda = 1.0$.

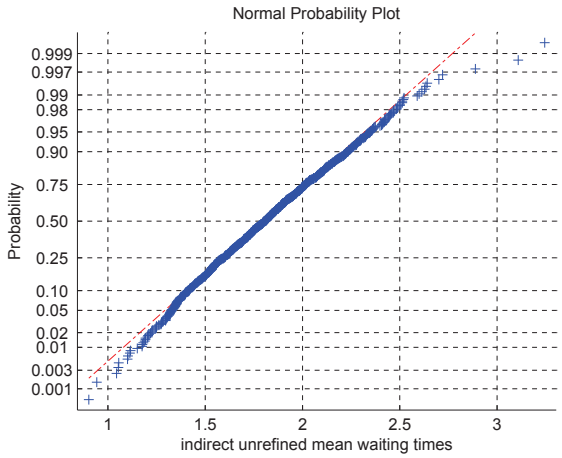

Figure EC.17 Normality plot of $\bar{W}_{L,\lambda}(t)$ for the $M/M/1$ model with $\lambda = 1.0$.

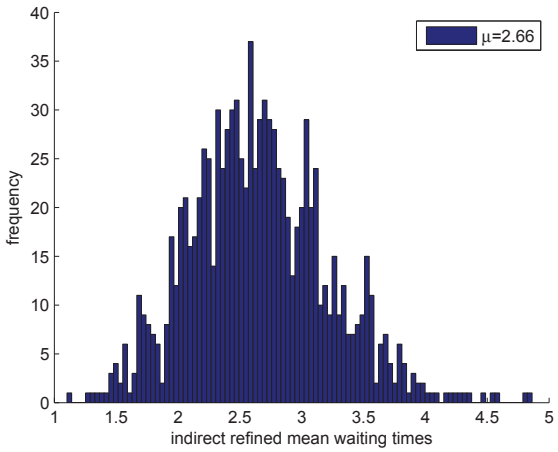

Figure EC.18 Histogram of $\bar{W}_{L,\lambda,r}(t)$ for the $M/M/1$ model with $\lambda = 1.0$.

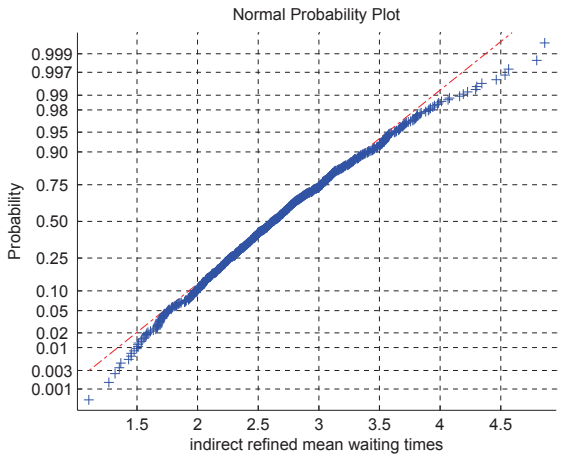

Figure EC.19 Normality plot of $\bar{W}_{L,\lambda,r}(t)$ for the $M/M/1$ model with $\lambda = 1.0$.

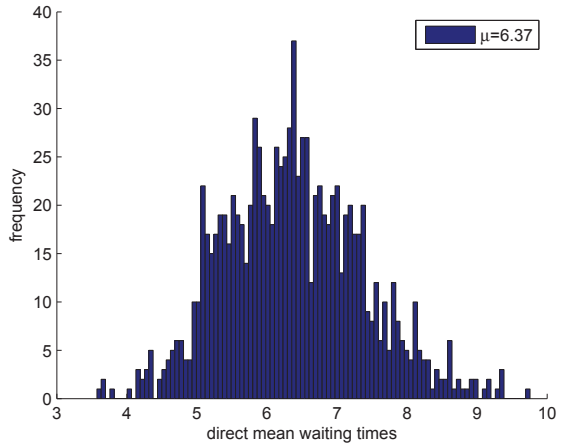

Figure EC.20 Histogram of $\bar{W}(t)$ for the $M/M/1$ model with $\lambda = 2.0$.

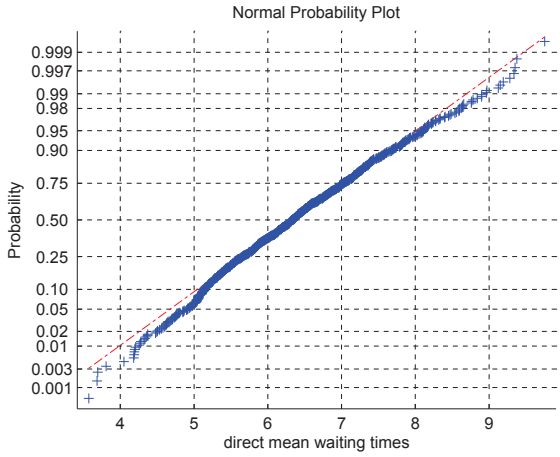

Figure EC.21 Normality plot of $\bar{W}(t)$ for the $M/M/1$ model with $\lambda = 2.0$.

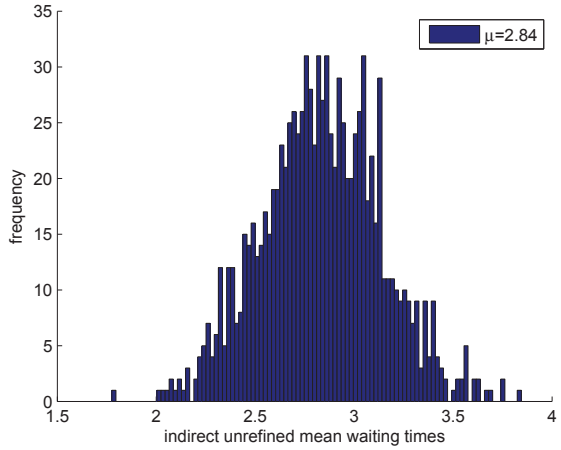

Figure EC.22 Histogram of $\bar{W}_{L,\lambda}(t)$ for the $M/M/1$ model with $\lambda = 2.0$.

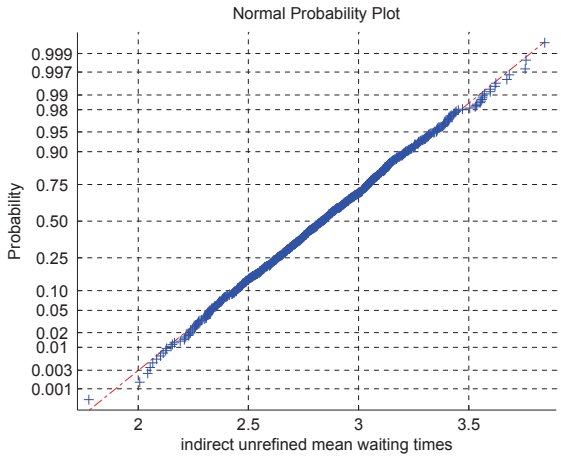

Figure EC.23 Normality plot of $\bar{W}_{L,\lambda}(t)$ for the $M/M/1$ model with $\lambda = 2.0$.

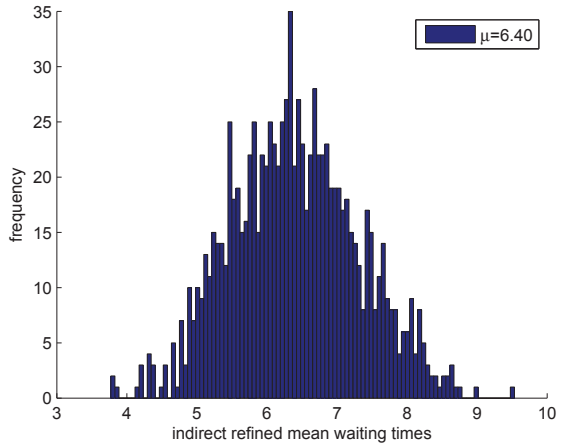

Figure EC.24 Histogram of $\bar{W}_{L,\lambda,r}(t)$ for the $M/M/1$ model with $\lambda = 2.0$.

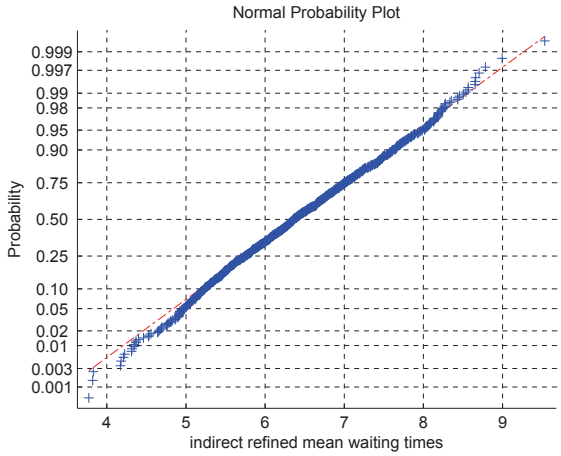

Figure EC.25 Normality plot of $\bar{W}_{L,\lambda,r}(t)$ for the $M/M/1$ model with $\lambda = 2.0$.

One way to try to obtain a better estimate of confidence interval is to use appropriate confidence interval inflation factor. The idea is to estimate the inflation factor x such that mean $\pm$ x confidence interval halfwidth has targeted (e.g., 95%) coverage. For our transient $M/M/1$ model, we test inflation factors ranging from 1.00 to 2.00, with increments of size 0.05. Table EC.3 provides the results for different values of the inflation factor. Given this result, we would estimate it to be about 1.55, 1.45 and 1.05 for $\lambda = 0.7$, 1.0 and 2.0, respectively.

Table EC.3 Estimating confidence interval inflation factor.

	$\lambda = 0.7$		$\lambda = 1.0$		$\lambda = 2.0$	
x	Cov. of $\bar{W}(t)$	Cov. of $\bar{W}_{L,\lambda,r}(t)$	Cov. of $\bar{W}(t)$	Cov. of $\bar{W}_{L,\lambda,r}(t)$	Cov. of $\bar{W}(t)$	Cov. of $\bar{W}_{L,\lambda,r}(t)$
1.00	87.3%	87.8%	90.2%	90.7%	94.0%	95.1%
1.05	88.5%	89.3%	91.4%	91.9%	95.4%	95.8%
1.10	89.2%	89.9%	92.3%	92.8%	96.4%	96.9%
1.15	90.2%	90.8%	93.2%	93.3%	96.9%	97.2%
1.20	91.3%	91.6%	93.6%	94.1%	97.2%	97.7%
1.25	91.8%	92.6%	93.9%	95.0%	97.8%	98.1%
1.30	92.7%	93.6%	94.2%	95.2%	98.1%	98.5%
1.35	93.4%	93.9%	94.4%	95.4%	98.5%	98.7%
1.40	93.8%	94.2%	94.7%	95.6%	98.6%	98.9%
1.45	94.2%	94.4%	95.2%	95.6%	98.7%	99.5%
1.50	94.5%	95.1%	95.7%	95.7%	98.7%	99.5%
1.55	94.9%	95.3%	95.9%	96.0%	99.0%	99.7%
1.60	95.2%	95.4%	96.2%	96.3%	99.1%	99.7%
1.65	95.6%	95.7%	96.5%	96.3%	99.3%	99.7%
1.70	95.9%	95.8%	96.7%	96.6%	99.5%	99.7%
1.75	96.3%	96.0%	97.0%	96.6%	99.5%	99.7%
1.80	96.4%	96.1%	97.3%	97.1%	99.6%	99.7%
1.85	96.6%	96.4%	97.3%	97.4%	99.6%	99.7%
1.90	96.9%	96.7%	97.3%	97.5%	99.6%	99.7%
1.95	96.9%	96.9%	97.4%	97.8%	99.6%	99.7%
2.00	97.1%	97.2%	97.6%	97.8%	99.6%	99.7%