Simulation Experiments with $M/G/100$ Queues in the *Halfin-Whitt* (Q.E.D) Regime

Supervised By: Avishai Mandelbaum* Project By: Roy Schwartz*

‡ avim@ie.technion.ac.il
‡ schwartz@ie.technion.ac.il

Draft of July 9, 2002

Industrial Engineering and Management, Technion, Haifa 32000, Israel

Performance of $M/G/S$ vs. β

In this chapter we will show how different distributions of the service time cause changes in the performance of the $M/G/100$ system in the *Halfin-Whitt* regime. We will also show that by choosing different service time distributions with equal first two moments, we can obtain a difference in the performance. We will also show that these differences can be of much significance. This fact is in contradiction to the 'classical' heavy traffic approximations (*Khintchin-Polatchek* for example). Thus we can conclude that these classical heavy traffic approximations are not accurate in the *Halfin-Whitt* regime.

First we want to verify that the systems which we are going to deal with are indeed in the *Halfin-Whitt* regime. We will show that the loads of the system $M/G/100$ that we are going to experiment with, are sufficiently high for an amount of 100 servers. According to *Halfin-Whitt*: $\rho_m \approx 1 - \frac{\beta}{\sqrt{m}}$ and $P(Wait > 0) \approx [1 + \frac{\beta\phi(\beta)}{\varphi(\beta)}]^{-1}$. Thus we can conduct the following check. We can increase m and ρ_m such that we maintain the equality $\rho_m = 1 - \frac{\beta}{\sqrt{m}}$, we do this by keeping β constant. If $P(Wait > 0)$ does not change much during this change of values of m and ρ_m, we will conclude that we are indeed in the *Halfin-Whitt* regime. We have conducted this check for the $M/M/100$ and $M/D/100$ systems. The results are presented in Tables 1 and 2.

β	0.1	0.6	1.5
100 **Servers**	0.882768	0.433853	0.0748505
400 **Servers**	0.881546	0.433753	0.0799413
900 **Servers**	0.881141	0.433687	0.0815614
1600 **Servers**	0.880921	0.433648	0.0823574
2500 **Servers**	0.880795	0.433622	0.0828305

Table 1: $P(Wait > 0)$ of the *Halfin-Whitt* check for $M/M/100$ system

β	0.2	0.6	1.5
100 **Servers**	0.753871	0.402847	0.0705919
200 **Servers**	0.752231	0.412152	0.0687077
300 **Servers**	0.751502	0.40169	0.0699783
400 **Servers**	0.751067	0.401472	0.0705919

Table 2: $P(Wait > 0)$ of the *Halfin-Whitt* check for $M/D/100$ system

Note that for a high value of β in the $M/M/100$ system, the waiting probability increases as we increase m and ρ_m. On the other hand, for the smaller values of β the change in the waiting probability is very small, only about 0.224% (this is in the case where $\beta = 0.1$, and the change is even smaller in the case where $\beta = 0.6$). In the $M/D/100$ system the change in the value of the waiting probability is very small for all checked values of β. Notice that as β approaches 0, ρ approaches 1 (since $\rho_m \approx 1 - \frac{\beta}{\sqrt{m}}$) . Thus we will conclude that for β values that are smaller than 0.6 (i.e. ρ values that are larger than 0.94), the system $M/G/100$ is in the *Halfin-Whitt* regime. An intuitive way to understand this result, is to assert that a ρ of 0.94 is large enough so that a queue with a 100 servers is in the *Halfin-Whitt* regime.

$P(Wait > 0)$ vs. β

According to *Halfin-Whitt*:$P(Wait > 0) \approx [1 + \frac{\beta\phi(\beta)}{\varphi(\beta)}]^{-1}$. We will first would like to check whether we obtain this theoretical result in the case of the $M/M/100$ system. We will remind that all the results in this chapter were obtained by running 1000 simulations each for $10,000$ time units (a time unit is the average service time). Results for the $M/M/100$ and $M/D/100$ systems were obtained via *Tijm*'s software. The theoretical result and the empirical result appear in Figure 1. One can see that the empirical result is virtually identical to the theoretical result.

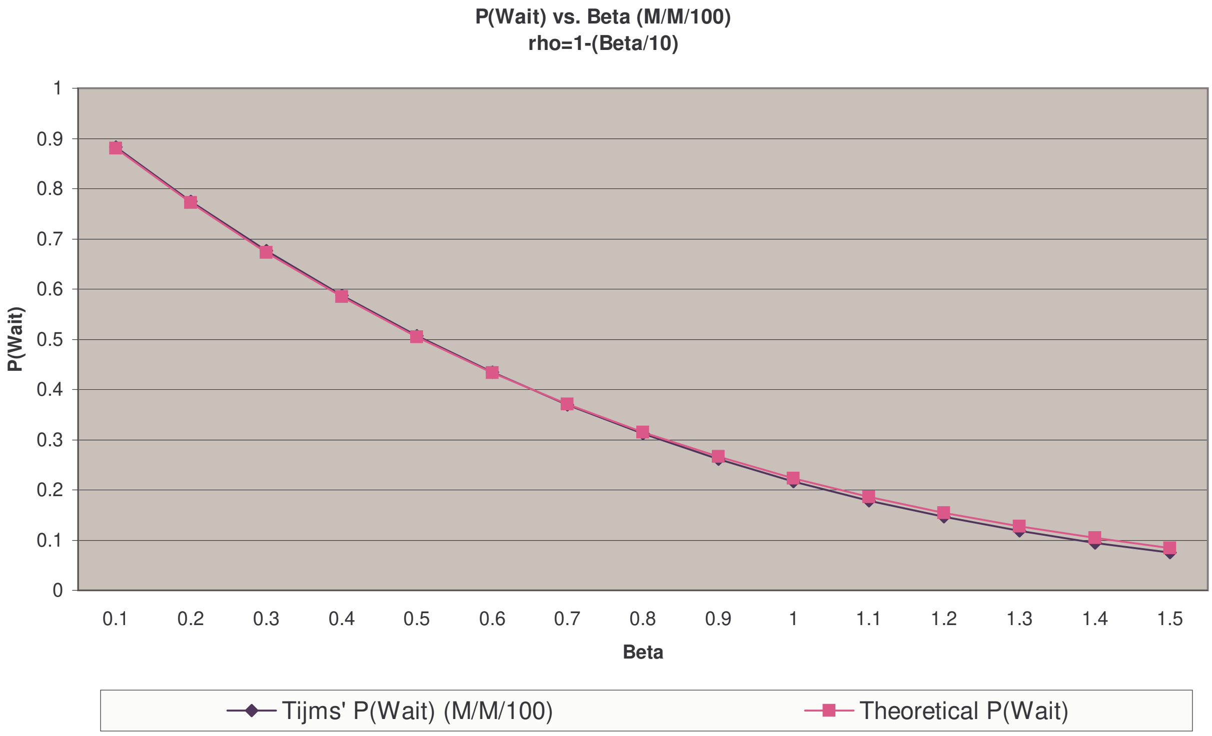

Figure 1: $P(Wait > 0)$ vs. β - empirical and theoretical results

Now we can turn our attention to the main part of this section, comparison of $P(Wait > 0)$ of systems which have different distributions of service time. We first check the deterministic service time distribution. Note that according to the traditional heavy traffic approximations, not only does $P(Wait > 0)$ depend on the first two moments of the service time distribution, but $P(Wait > 0)$ is an increasing function of these two moments (i.e. if at least one of these moments increases, $P(Wait > 0)$ also increases). Thus we would expect that the $M/D/100$ system in the *Halfin-Whitt* regime would have a smaller waiting probability in comparison to the $M/M/100$ system with the same β value (i.e. the same ρ). The results appear in Figure 2.

One can see that the results are indeed as predicted by the traditional approximations. The waiting probability is indeed smaller in the $M/D/100$ system in comparison with the $M/M/100$ system, for each value of β.

The second step was checking *log-normal* service time distribution. Not like the exponential and deterministic service time distributions, the log-normal distribution has moments

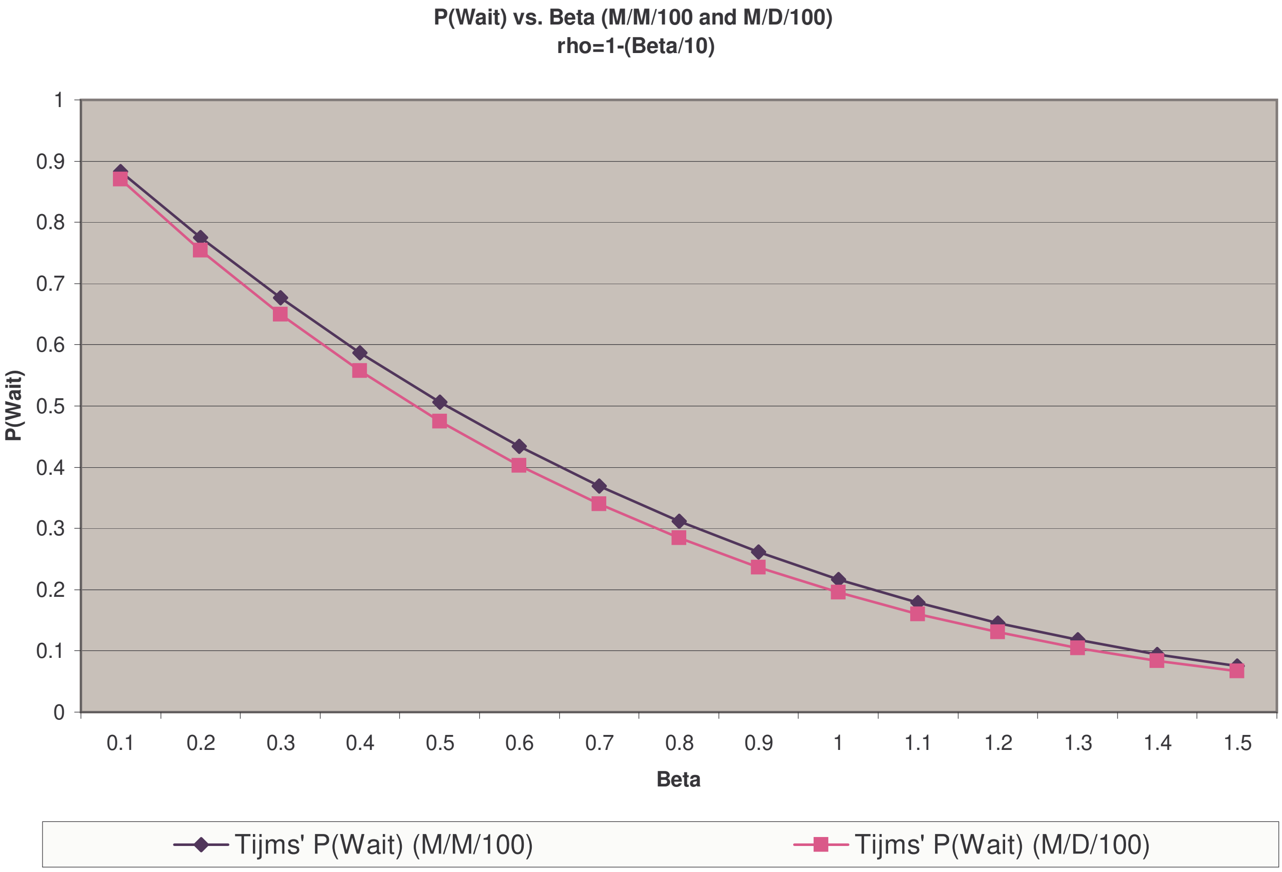

Figure 2: $P(Wait > 0)$ vs. β ($M/M/100$ and $M/D/100$)

the depend on the distribution parameters (μ and σ) such that we can set *both* moments to any size. To our purposes we would like to have the first moment equal to 1, as was in all the other service time distribution that we have checked and will be in the ones we would check. Regarding the second moment, we can choose a value for it such that we receive any coefficient of variation we want. This can be obtained by choosing: $\mu = -\frac{1}{2}\sigma^2$ and $\sigma = \sqrt{\ln(CV + 1)}$, where CV is the requested coefficient of variation. We have chosen the following values of the coefficient of variation: 1, 10 and approximately 1.3. We present results only for the first two values, since the results of the third value fall between the results of the first two. The results appear in Figure 3.

Notice that for the log-normal service time distribution the traditional approximations are correct in the sense that the waiting probability is higher in the $M/LN/100$ system with $CV = 10$ in comparison with the system $M/LN/100$ with $CV = 1$. However, according to these approximations the systems $M/M/100$ and $M/LN/100$ with $CV = 1$ are supposed to behave the same, since these systems have the exact same two first moments of service time distribution. However, one can easily see that these systems have *different* waiting probabilities. The waiting probability is smaller in the $M/LN/100$ system with $CV = 1$ in comparison with the $M/M/100$ system for each value of β. Thus we have obtained an example in which the waiting probability does not depend only on the first two moments of the service time distribution, thus we conclude that the traditional heavy traffic approximations

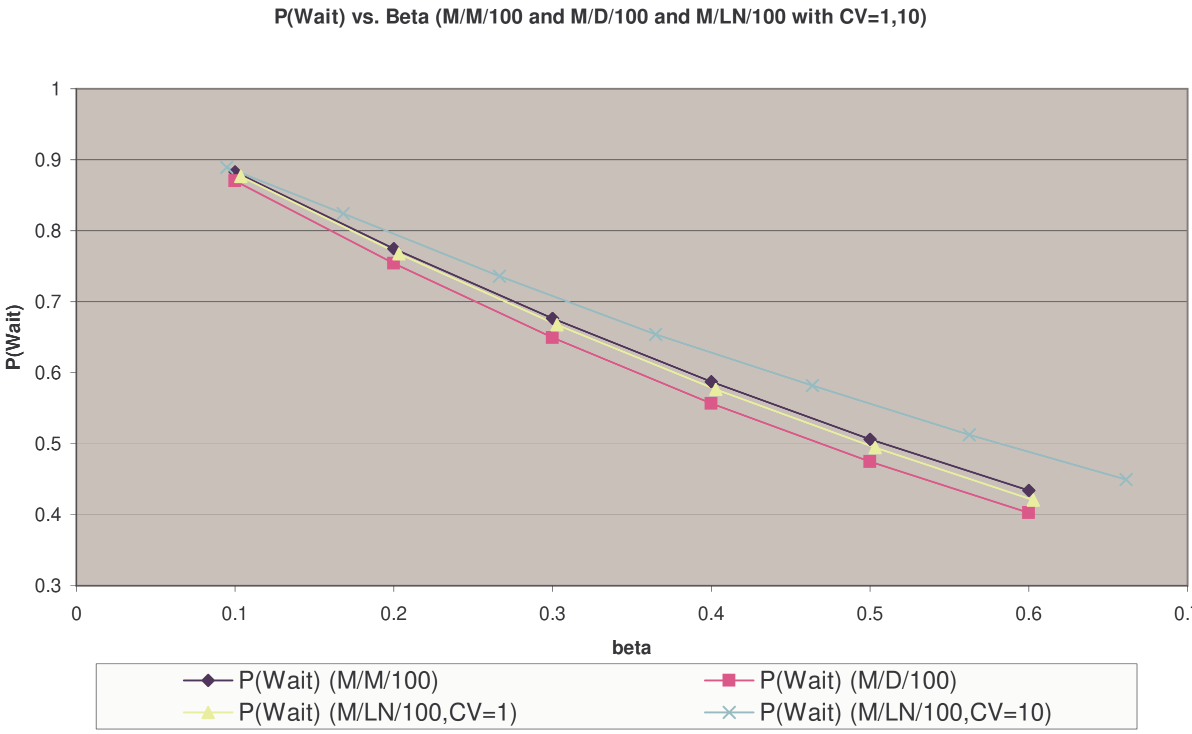

Figure 3: $P(Wait > 0)$ vs. β ($M/M/100$, $M/D/100$ and $M/LN/100$ with $CV = 1$ and $CV = 10$)

are not correct in the *Halfin-Whitt* regime. In chapter 6 we present a special service time distribution that achieved these results in a much clearer fashion.

$E(W_q)$ and $E(W_q|W_q > 0)$ vs. β

One can extend the *Khintchin-Polatchek* approximation for the $M/G/1$ system to more than one server in the following way (where m is the number of servers):

$$\begin{cases} E(W_q) \approx \frac{1}{m}\frac{1}{\mu}\frac{\rho}{1-\rho}\frac{1+C_s^2}{2} \\ E(W_q|W_q > 0) \approx \frac{1}{m}\frac{1}{\mu}\frac{1}{1-rho}\frac{1+C_s^2}{2} \end{cases}$$

We will check how the different service time distributions affect $E(W_q)$ and $E(W_q|W_q > 0)$ as a function of β. The results appear in Figures 4 and 5 (these results are for the same systems that were checked when dealing with waiting probability).

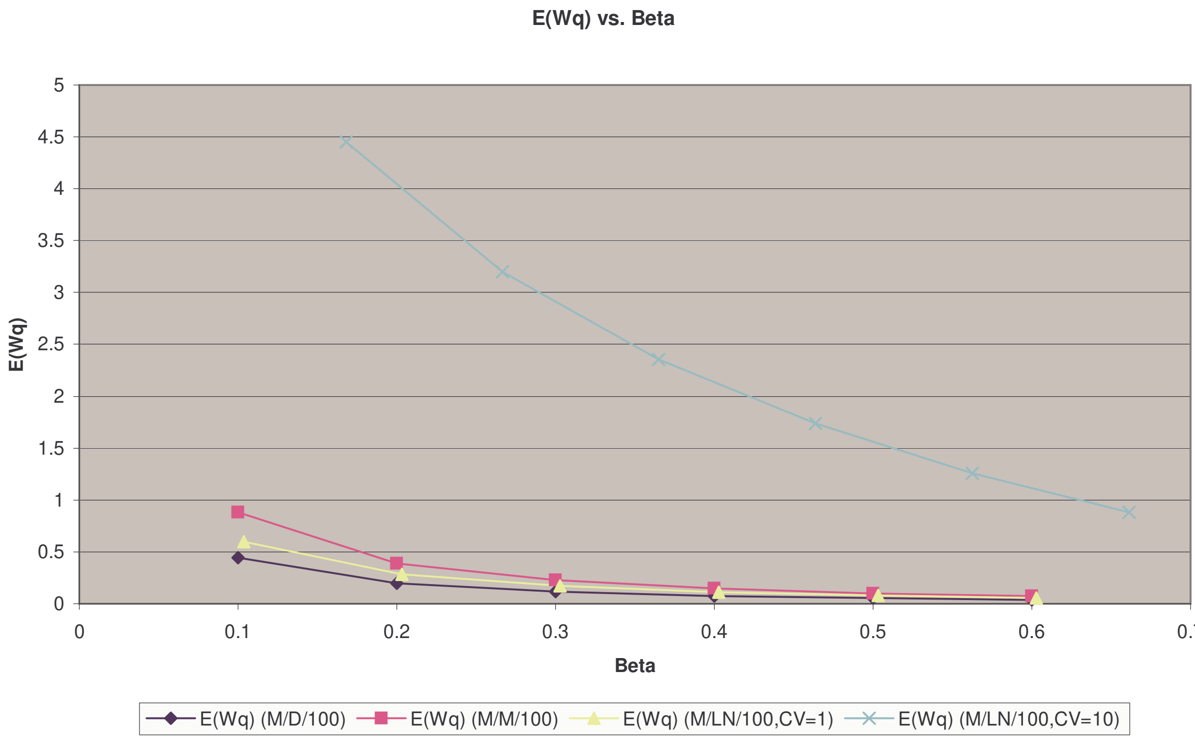

Figure 4: $E(W_q)$ vs. β ($M/M/100$, $M/D/100$ and $M/LN/100$ with $CV = 1$ and $CV = 10$)

Note that for the systems $M/LN/100$ with coefficient of variation 1 and 10, $E(W_q)$ and $E(W_q|W_q > 0)$ are larger when $CV = 10$ than in the case when $CV = 1$. However, $E(W_q)$ and $E(W_q|W_q > 0)$ are smaller in the $M/LN/10$ with $CV = 1$ than in the system $M/M/100$ for each value of β. This *gap* between the systems becomes more apparent as β approaches 0 (i.e. ρ approaches 1). Thus even where ρ approaches 1, the traditional approximations are not accurate since they claim that $E(W_q)$ and $E(W_q|W_q > 0)$ depend only on the first two moments of the service time distribution. That is clearly not the case in the example presented for systems that are in the *Halfin-Whitt* regime. Thus we can conclude that there is at least one additional factor that is missing in order to approximate $E(W_q)$ and $E(W_q|W_q > 0)$. Presented are graphs without the system $M/LN/100$ with $CV = 10$, so that these phenomena can be seen more easily. These results appear in Figures 6 and 7.

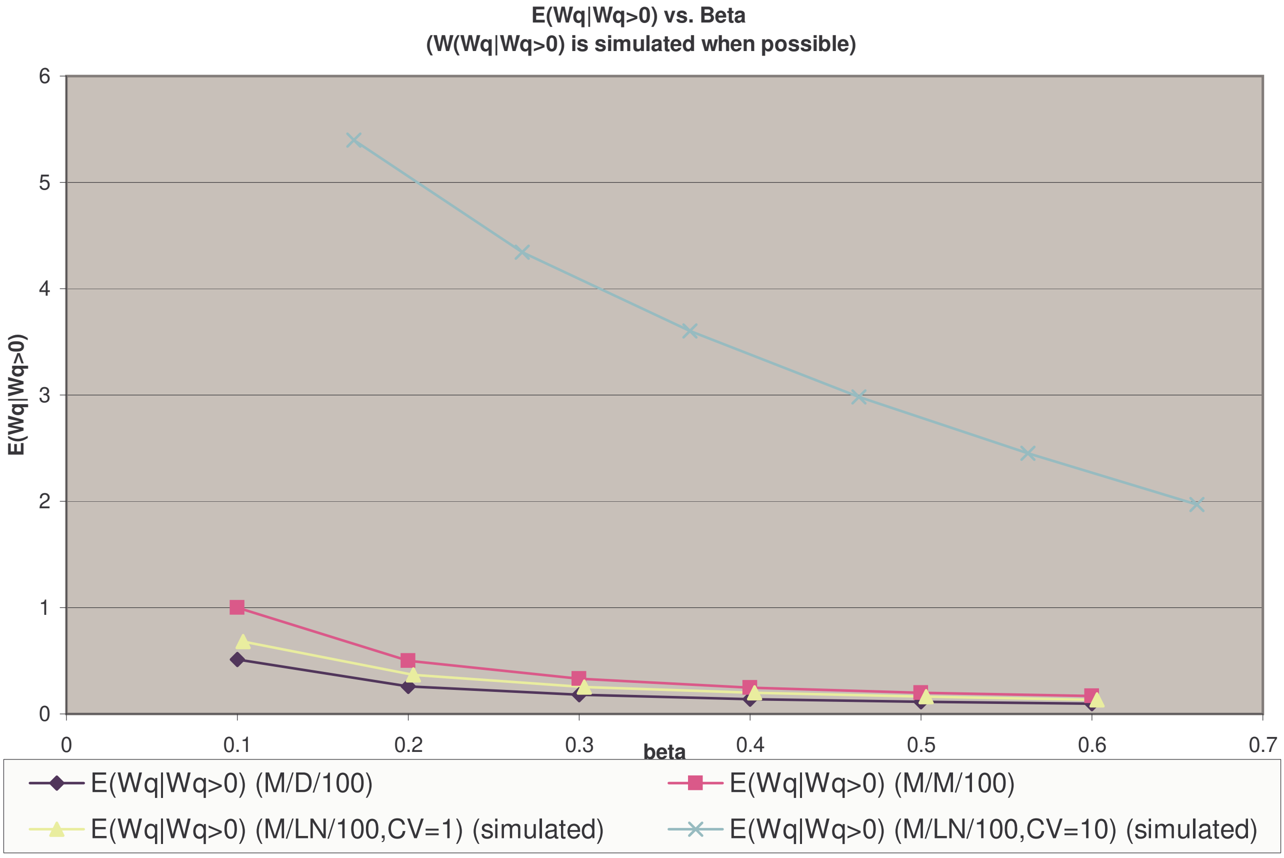

Figure 5: $E(W_q|W_q > 0)$ vs. β ($M/M/100$, $M/D/100$ and $M/LN/100$ with $CV = 1$ and $CV = 10$)

One can see from these two Figures that the traditional approximation is correct for the two systems $M/M/100$ and $M/D/100$. Note that according to the traditional heavy traffic approximation, $E(W_q)$ and $E(W_q|W_q > 0)$ is supposed to be half in the $M/D/100$ system in comparison to the $M/M/100$ system. The reason for this is that the coefficient of variation of a deterministic service time is 0 and the coefficient of variation of an exponential service time is 1. Thus we obtain a factor of 2 as a result from the term $\frac{1+C_s^2}{2}$. Indeed for $E(W_q)$ and $\beta = 0.1$ we receive a value of 0.88277 for the $M/M/100$ system and a value of 0.445543 for the $M/D/100$ system. For $E(W_q|W_q > 0)$ and $\beta = 0.1$ we receive a value of 1.000002266 for the $M/M/100$ system and a value of 0.511930683 for the $M/D/100$ system. In both cases, $E(W_q)$ and $E(W_q|W_q > 0)$ are about half in the $M/D/100$ system in comparison with the $M/M/100$ system, as the traditional heavy traffic approximation predicts.

However, according to this approximation we were supposed to get the same results for the systems $M/M/100$ and $M/LN/100$ with $CV = 1$. That is clearly not the case as one can see in the Figures. Thus we conclude that the traditional heavy traffic approximations are not accurate in the *Halfin-Whitt* regime. In chapter 6 we present a special service time distribution that achieved these results in a much clearer fashion.

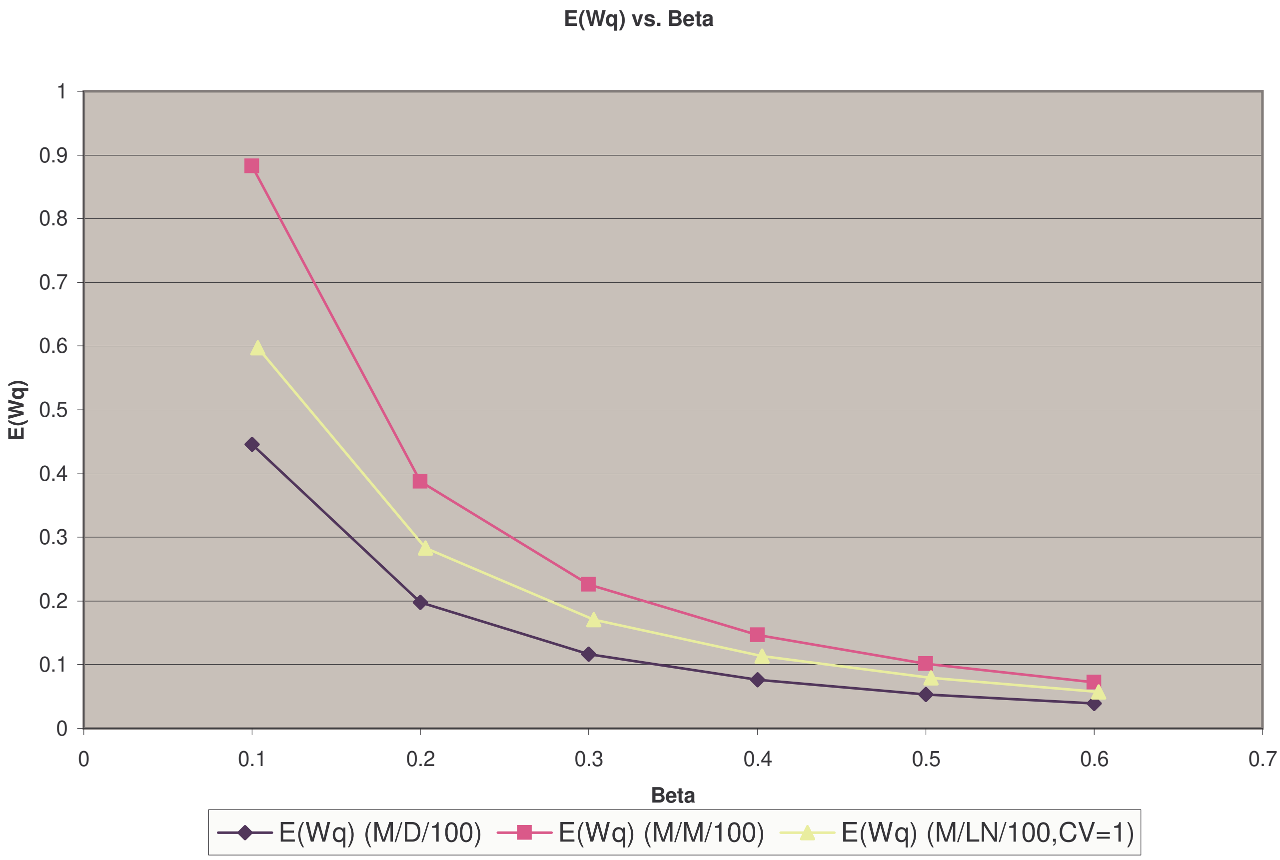

Figure 6: $E(W_q)$ vs. β ($M/M/100$, $M/D/100$ and $M/LN/100$ with $CV = 1$)

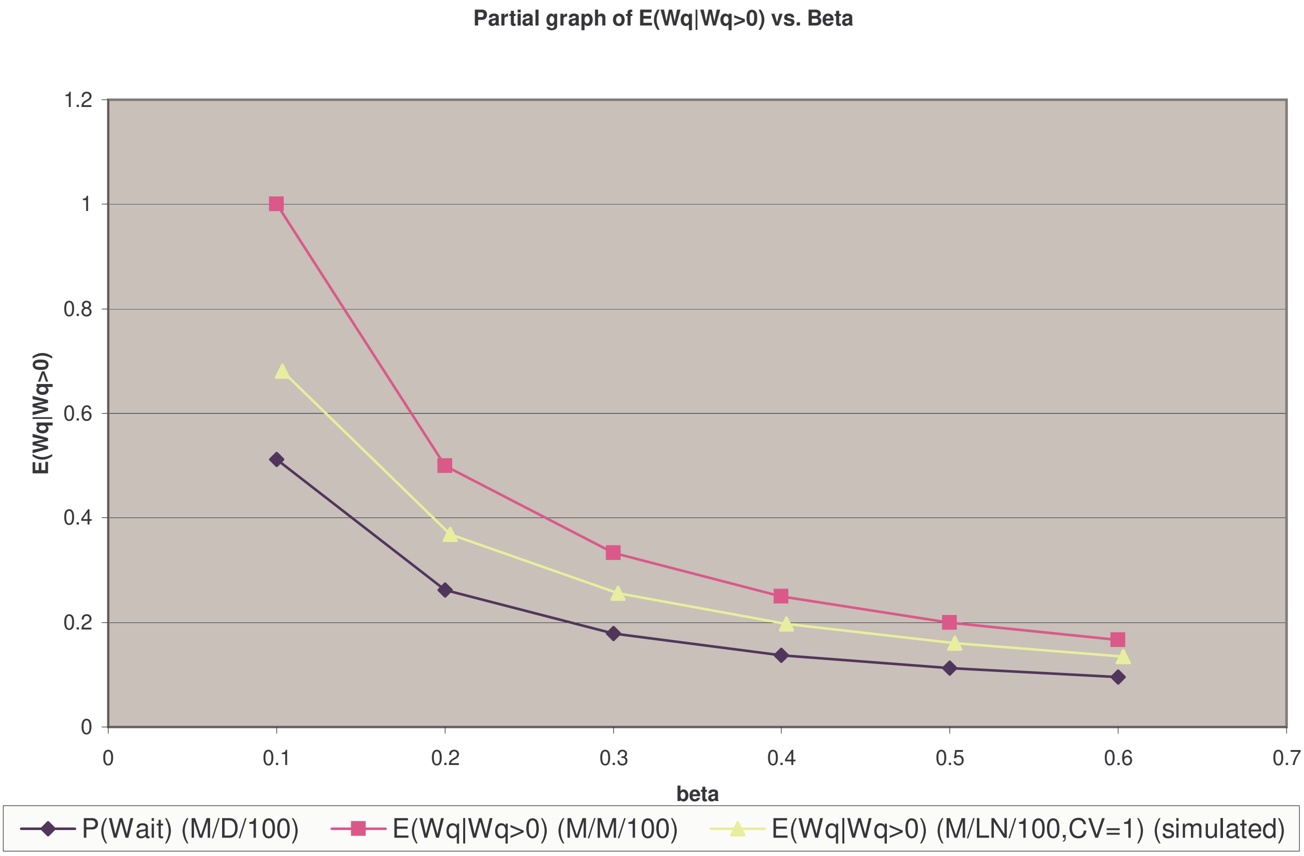

Figure 7: $E(W_q|W_q > 0)$ vs. β ($M/M/100$, $M/D/100$ and $M/LN/100$ with $CV = 1$)

In the previous chapters, we presented results concerning various statistics (all 'byproducts' of some steady-state distribution) but we have never displayed the distribution itself. In this chapter we try to do just that. We will present histograms for the waiting time given wait (i.e. $W_q|W_q>0$) in several cases. Of course we are interested in histograms of systems that are in the *Halfin-Whitt* regime.

In the case of the $M/M/S$ system it is known that waiting time given wait is distributed exponentially, hence we do not present results for such systems. It is also known from *Kingman*'s law that, in the classical heavy traffic regime, $W_q|W_q>0 \overset{d}{\approx} exp(mean = \frac{1}{1-\rho}\frac{1}{S}\frac{1}{\mu}\frac{C_a{}^2+C_s{}^2}{2})$ for any $GI/GI/S$ system as ρ approaches 1 (equivalently β approaches 0). The question is, whether this approximation still prevails in the *Halfin-Whitt* regime (and also for ρ's that approach 1). We will present results in which $W_q|W_q>0$ is clearly not distributed exponentially in this regime. As seen in the previous chapter, interesting phenomena appeared in the case of $M/LN/100$ systems. Thus we start with results for these systems. Then, in the next chapter we introduce a special service time distribution, for which we will also present histograms of $W_q|W_q>0$.

We will conduct the analysis of each histogram in two phases. In the first phase we compare the histogram with a theoretical exponential density function. The parameter chosen for this theoretical exponential function is: $\frac{1}{\widehat{E}(W_q|W_q>0)}$ (i.e. $\hat{\lambda}=\frac{1}{\widehat{E}(W_q|W_q>0)}$). $\widehat{E}(W_q|W_q>0)$ is obtained via the simulation. This phase enables us to see if the empirical histogram of $W_q|W_q>0$ resembles an exponential density function.

The second phase consists of an additional check. First, if f is an exponential density function with parameter λ (i.e. $f(t)=\lambda e^{-\lambda t}$, $\forall t\geq 0$), and F is the corresponding cumulative distribution function (i.e. $F(t)=\int_{s=0}^{t} f(s)ds = 1-e^{-\lambda t}$, $\forall t\geq 0$), then it can be easily shown that $-\frac{1}{\lambda}\ln(1-F(t))=t$, $\forall t\geq 0$. This fact provides us with an additional method of checking whether a given empirical density function is an exponential density function. The method is the following: computing the empirical cumulative distribution function (which we will refer to as $\hat{F}$) from the empirical density function (histogram), then calculating $-\frac{1}{\hat{\lambda}}\ln(1-\hat{F}(t))$. Since $\hat{\lambda}=\frac{1}{\widehat{E}(W_q|W_q>0)}$, we can conclude that: $(-\frac{1}{\hat{\lambda}}\ln(1-\hat{F}(t))) = (-\widehat{E}(W_q|W_q>0)\ln(1-\hat{F}(t)))$. The final stage of this method is thus to check whether the function $(-\widehat{E}(W_q|W_q>0)\ln(1-\hat{F}(t)))$ is a straight line with a slope of 1. As in previous chapters, the results were obtained by running 1000 simulations, each for a period of $10,000$ time units (where a time unit is the average service time).

First we will present results for the $M/LN/100$ system when the coefficient of variation CV of the service time is 1. These results appear in Figures 8 and 9.

Note that the histogram in this case does not seem to be an exponential density function. Indeed, the histogram seems to be more concave than that of the exponential. This is expressed by the fact that there is less distribution mass near 0 in the histogram in comparison to the theoretical exponential density function, and as a consequence there is more distribution mass away from 0 in the histogram in comparison with the theoretical exponential density function. Note also that in the graph that presents our second phase check, it can be seen more clearly that this histogram is not an exponential density function. The line

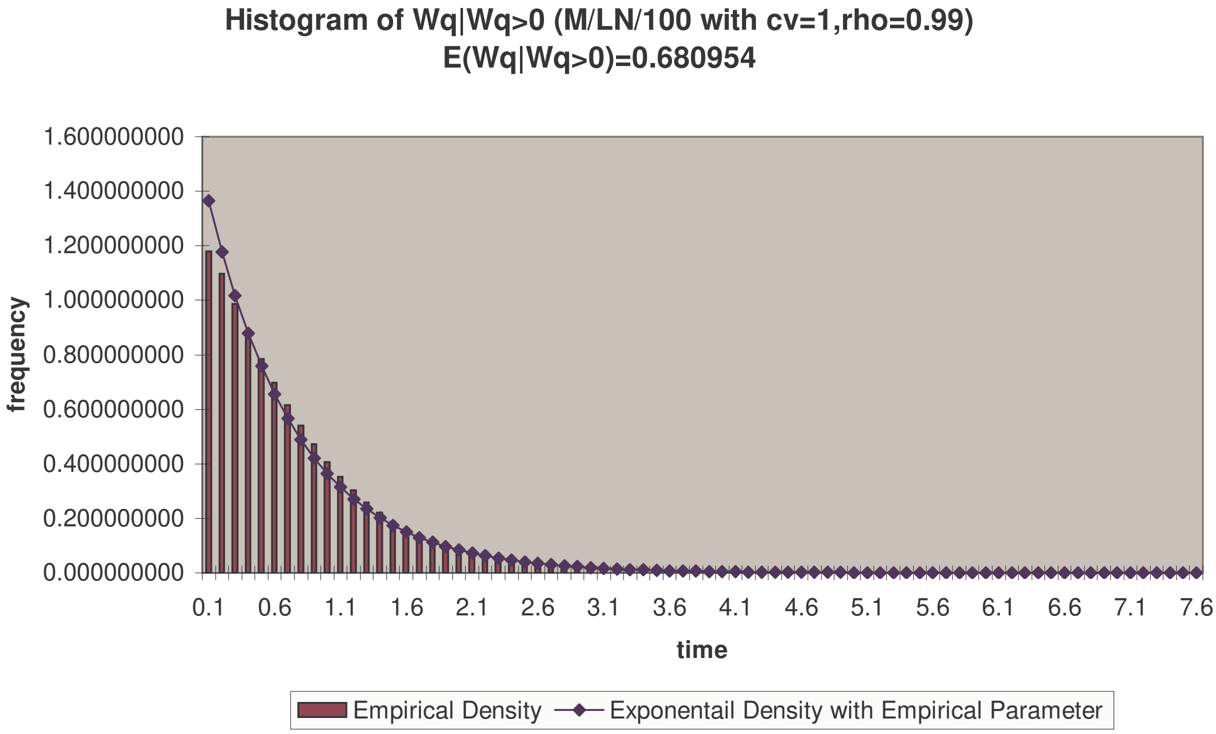

Figure 8: Histogram of $W_q|W_q > 0$ and exponential density ($M/LN/100$, $\rho = 0.99$, $cv = 1$)

that represents $(-\widehat{E}(W_q|W_q > 0)\ln(1 - \hat{F}(t)))$ is not a straight line as it curves upwards for large t values.

Notice that for t values for which $(-\widehat{E}(W_q|W_q > 0)\ln(1 - \hat{F}(t))) < t$ (i.e. the function $(-\widehat{E}(W_q|W_q > 0)\ln(1 - \hat{F}(t)))$ is smaller than the identity function), we can conclude that $\hat{F}(t) < F(t)$ where F is the cumulative exponential distribution function. Similarly, for t values for which $(-\widehat{E}(W_q|W_q > 0)\ln(1 - \hat{F}(t))) > t$ we can conclude that $\hat{F}(t) > F(t)$. Thus the function $\hat{F}(t)$ for waiting time given wait in the $M/LN/100$ system with $CV = 1$, for a value of t that is away from 0, is larger than $F(t)$ (where F is the cumulative distribution function in the exponential case). Hence we conclude that the distribution of waiting time given wait in the $M/LN/100$ system with $CV = 1$ has a lighter tail than an exponential distribution (which is no surprise since in the results presented in the previous chapter we showed that $E(W_q|W_q > 0)$ is smaller in the $M/LN/100$ system with $CV = 1$ and with $\rho = 0.99$ in comparison with the $M/M/100$ system with the same ρ, and in the latter it is known that $W_q|W_q > 0$ is distributed exponentially). Thus we can conclude that in the *Halfin-Whitt* regime with high ρ, waiting time given wait in the $M/LN/100$ system when the CV of the service time is 1 is clearly not distributed exponentially.

We will now examine the results for the $M/LN/100$ system when the CV of the service time is 10. We will show that in this case, even for not very high values of ρ (but the system is still in the *Halfin-Whitt* regime) the distribution of $W_q|W_q > 0$ is not exponential. We have arbitrarily chosen to show results for the following ρ values: 0.94, 0.96 and 0.98. For each ρ we will present two graphs, each one corresponds to the appropriate phase. The results appear in Figures 10 to 15.

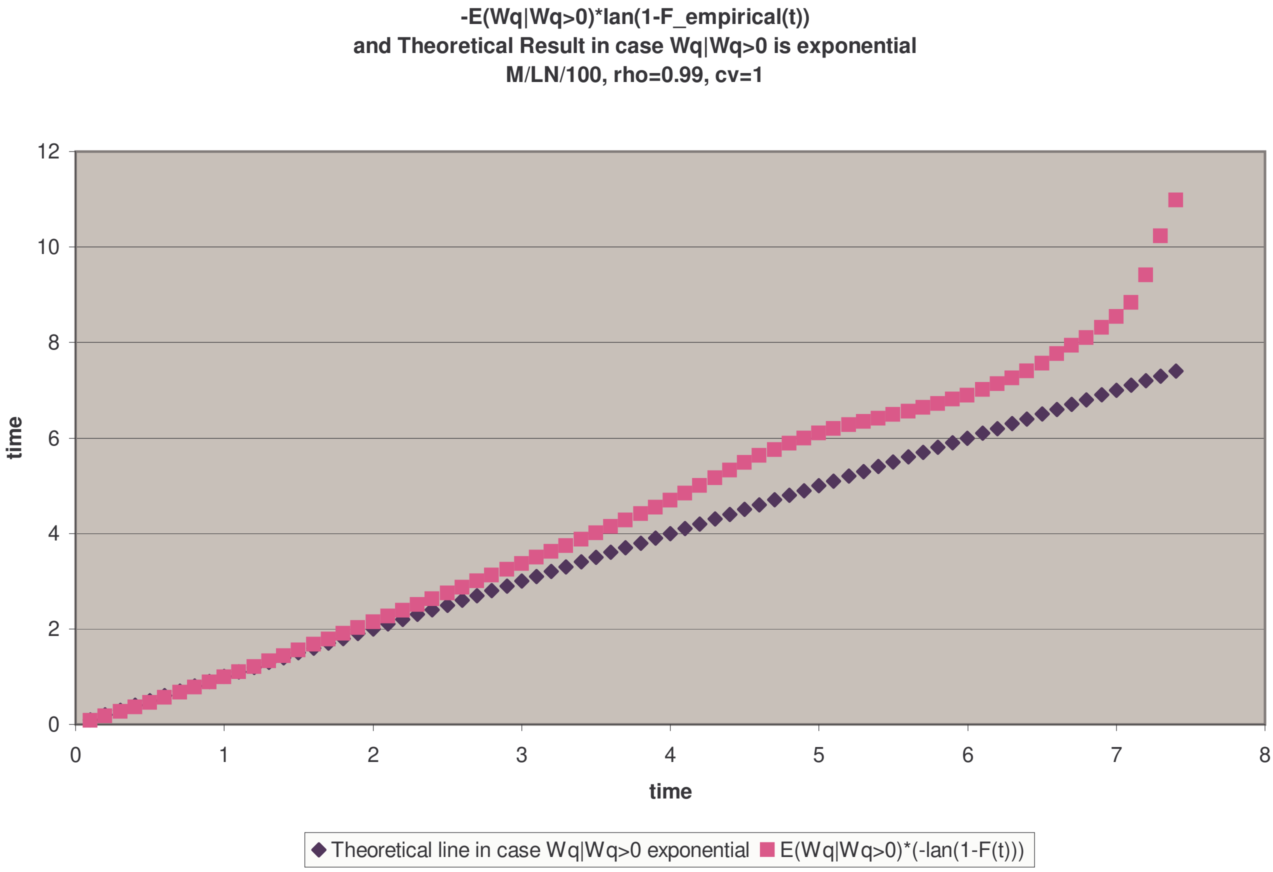

Figure 9: Second phase check ($M/LN/100$, $\rho = 0.99$, $cv = 1$)

Note that it is clear from this graphs, that the histograms of $W_q|W_q > 0$ do not represent exponential distribution functions. It can be seen that each one of the histograms has at least one local maximum point that is not 0. As a result these histograms are concave and convex alternatively. Also it can be seen in the graphs that present the second phase check, that the function $(-\widehat{E}(W_q|W_q > 0)\ln(1 - \hat{F}(t)))$ is convex near 0 and concave afterwards. Thus this function is clearly not a straight line with slope 1. If we try to conduct the same tail analysis as done in the case of the $M/LN/100$ system with $CV = 1$, we can state that as ρ approaches 1, the distribution of waiting time given wait in the $M/LN/100$ system with $CV = 10$ has a heavier tail than the exponential distribution. Note that this is no surprise since in the previous chapter we showed that $E(W_q|W_q > 0)$ is larger in the $M/LN/100$ system with $CV = 10$ in comparison with the $M/M/100$ system, in which it is known that waiting time given wait is exponential. So we can conclude that in the *Halfin-Whitt* regime, waiting time given wait in the $M/LN/100$ system when the CV of the service time is 10 is not distributed exponentially.

Summing up, we conclude that the *Kingman* approximation is probably not accurate in the *Halfin-Whitt* regime, since we have shown a system for which this approximation does not hold (i.e. $W_q|W_q > 0$ is not distributed exponentially).

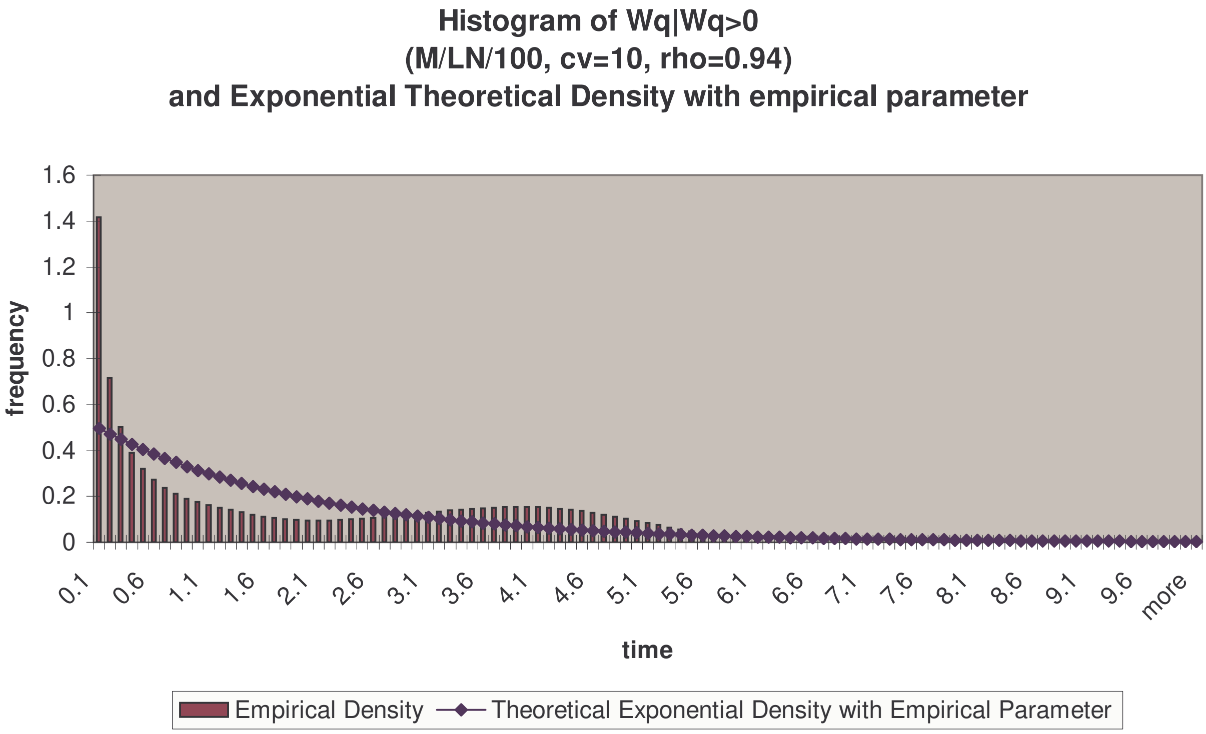

Figure 10: Histogram of $W_q|W_q > 0$ and exponential density ($M/LN/100$, $\rho = 0.94$, $cv = 10$)

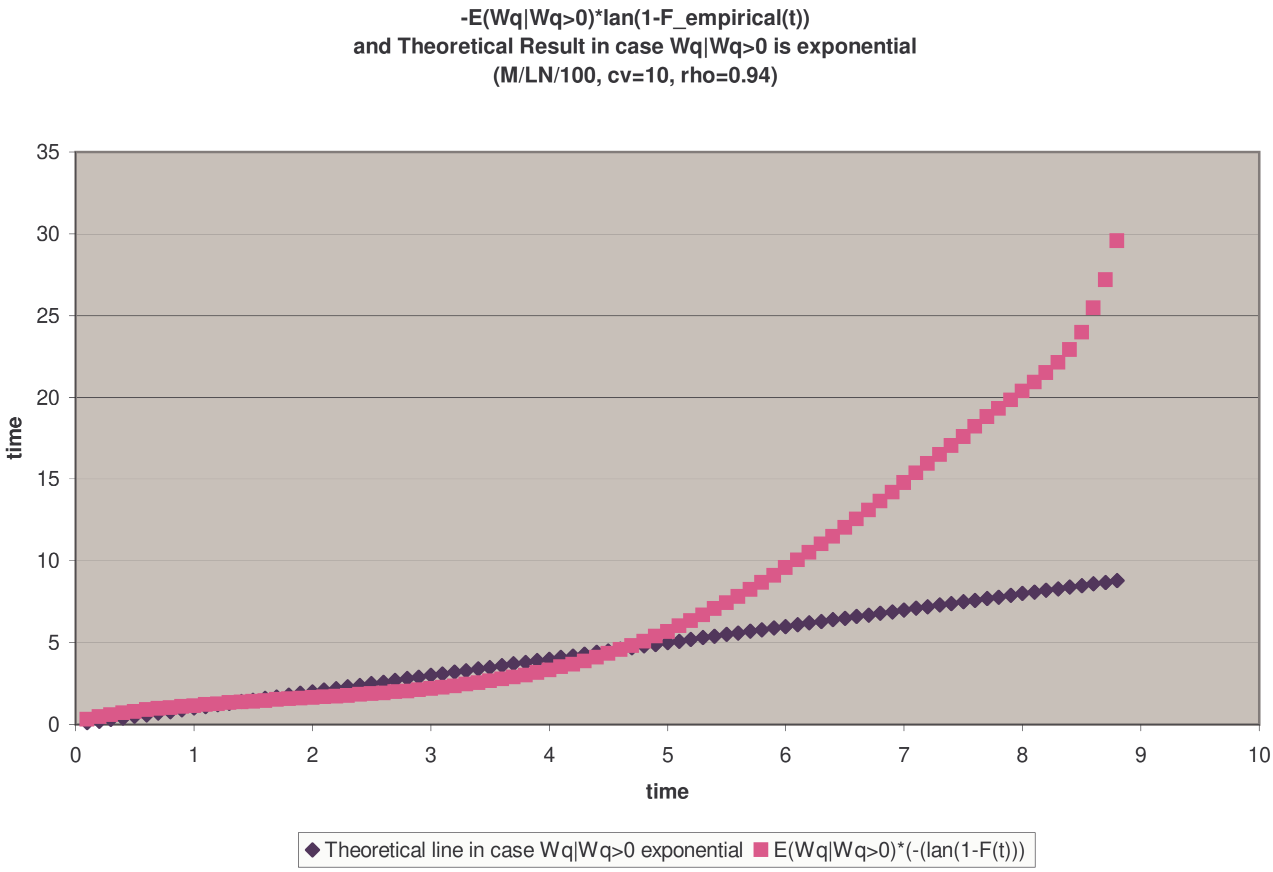

Figure 11: Second phase check ($M/LN/100$, $\rho = 0.94$, $cv = 10$)

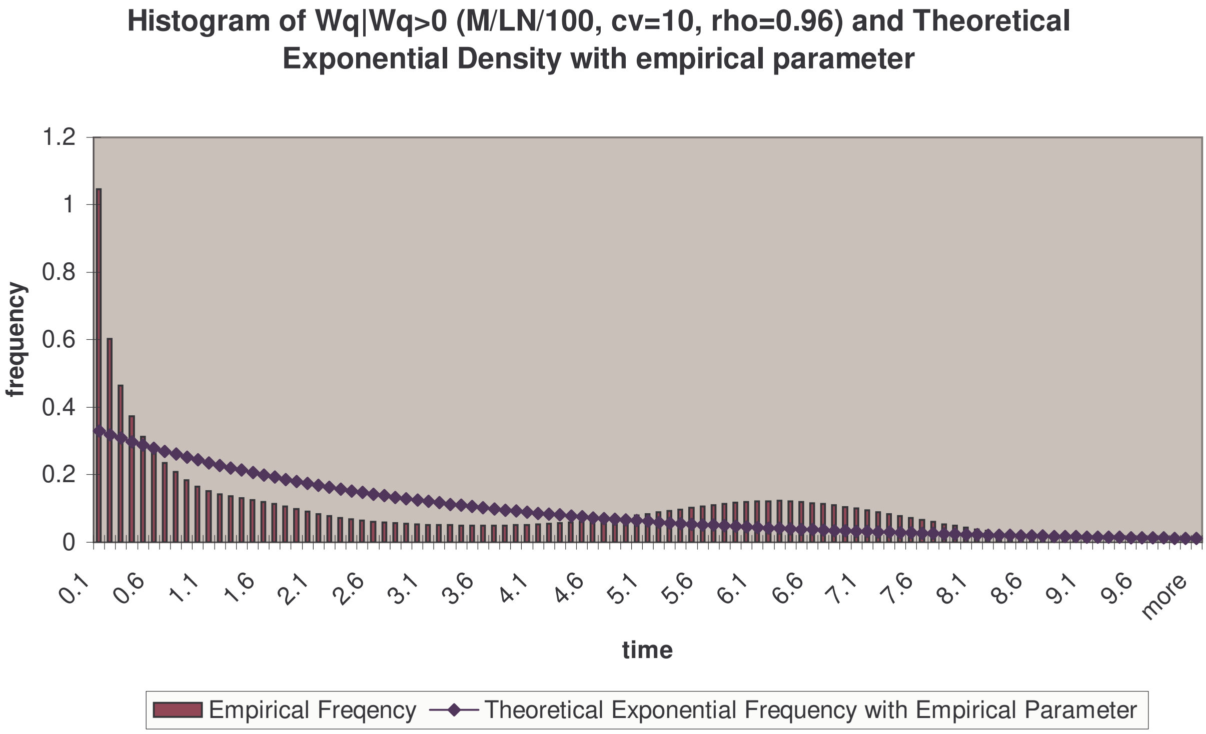

Figure 12: Histogram of $W_q|W_q > 0$ and exponential density $(M/LN/100,\ \rho = 0.96,\ cv = 10)$

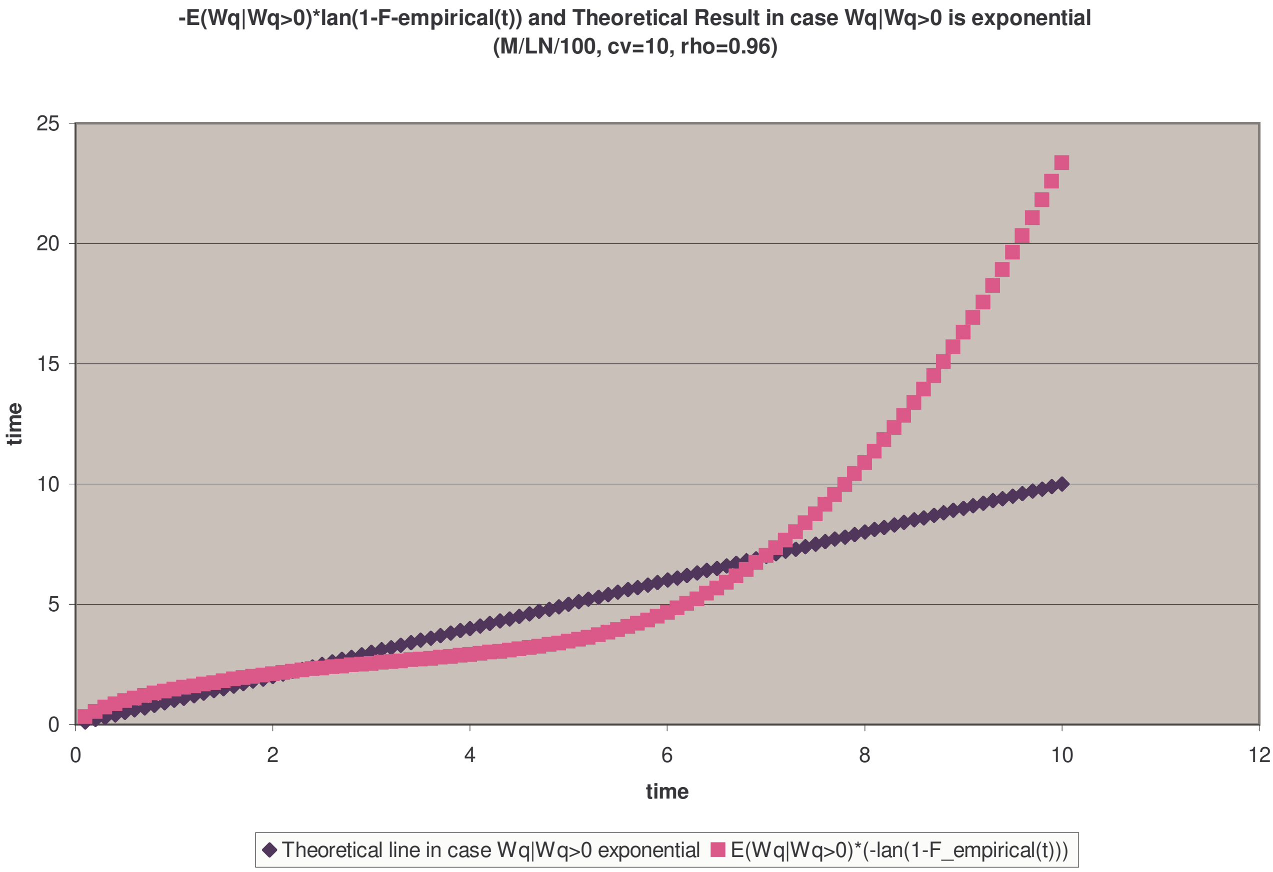

Figure 13: Second phase check ($M/LN/100$, $\rho = 0.96$, $cv = 10$)

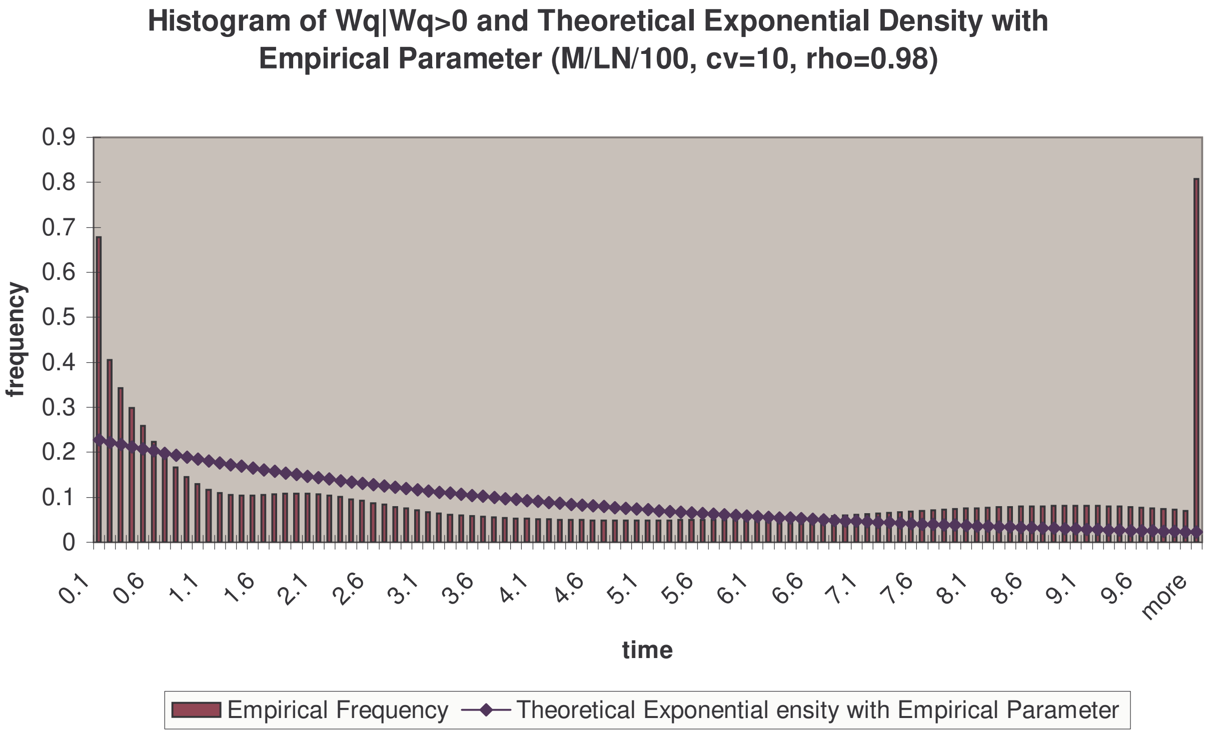

Figure 14: Histogram of $W_q|W_q > 0$ and exponential density $(M/LN/100,\ \rho = 0.98,\ cv = 10)$

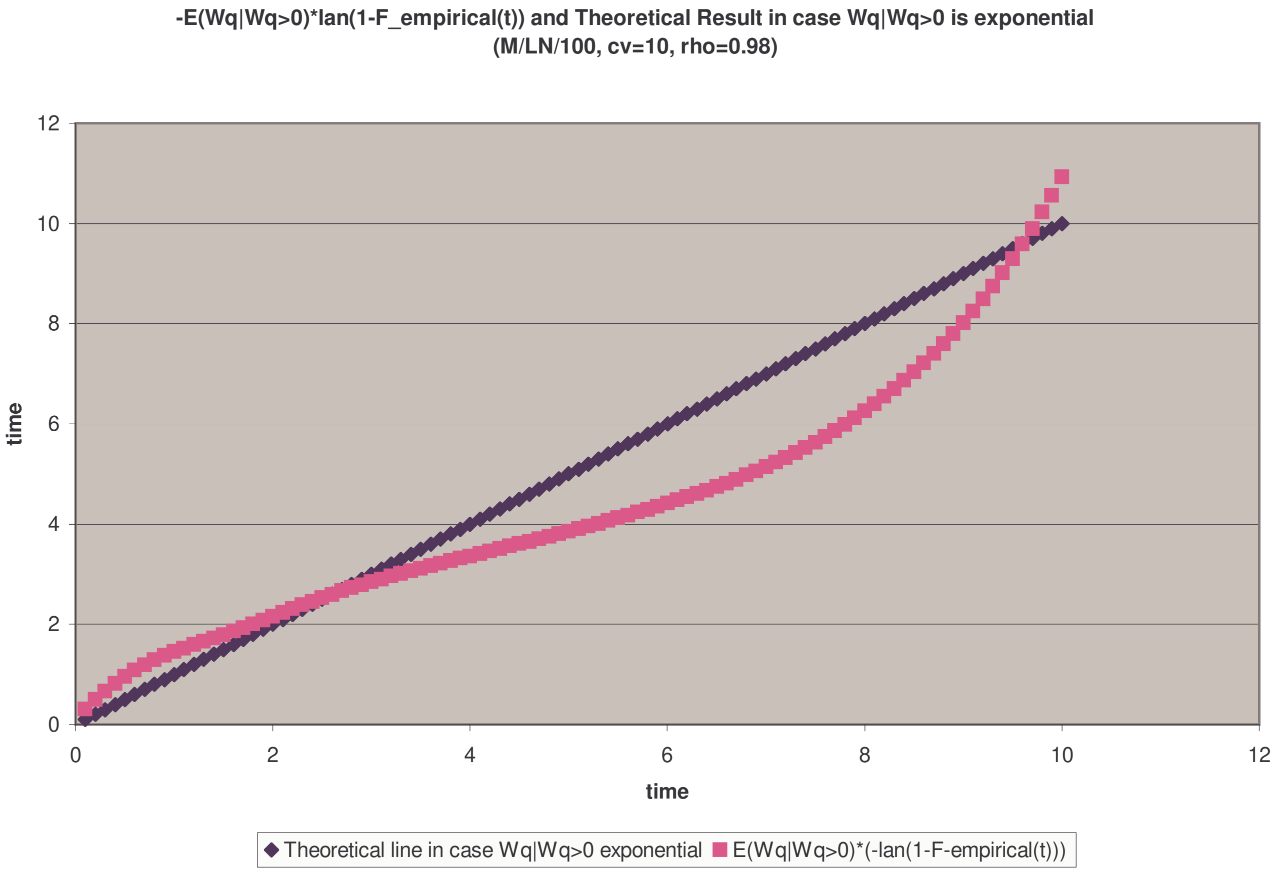

Figure 15: Second phase check ($M/LN/100$, $\rho = 0.98$, $cv = 10$)

Performance of $M/G/S$ vs. β and $W_q|W_q > 0$ Histograms for Special Service Time Distribution

As mentioned in the chapter that contains the performance of $M/G/100$ vs. β, we will now present a special distribution that we will use as a service time distribution. Using this special distribution we will be able to show in a much clearer fashion the phenomena which appeared in the two previous chapters.

In this chapter we will present the various statistics vs. β for different parameter values of the special distribution. We will compare the results received for the special distribution with the results received for the *regular* distributions (exponential, deterministic and log-normal). Additionally we will compare these results of the special distribution for several parameter values of the special distribution itself. Finally we will present histograms for $W_q|W_q > 0$ for the different parameter values of the special distribution, compare these histograms between the different parameter values and compare the special distribution case with the *regular* distributions cases.

Special Service Time Distribution

Definition 1 *We will say that a random variable X is distributed Special(p) for any p such that $\frac{1}{2} \leq p < 1$ iff $P(X = a) = p$ and $P(X = b) = 1 - p$ where $a = 1 - \sqrt{\frac{1-p}{p}}$ and $b = 1 + \sqrt{\frac{p}{1-p}}$.*

We will use the notation $X \sim$ *Special*(p) to indicate that X is distributed *Special*(p), i.e.

$$X = \begin{cases} 1 - \sqrt{\frac{1-p}{p}} & \text{wp } p \\ 1 + \sqrt{\frac{p}{1-p}} & \text{wp } 1-p \end{cases}, \qquad \frac{1}{2} \leq p < 1.$$

Lemma 1 *If $X \sim$ Special(p) for some p such that $\frac{1}{2} \leq p < 1$, than $E(X) = 1$ and $Var(X) = 1$.*

Can be easily shown using simple arithmetics.

In our simulations we used three special cases of the special distribution, each case corresponds to a different value of the parameter p. Two of these cases were chosen as *extreme* cases, in which the special distribution behaves in a manner that is close to some simple distribution. The third case is an intermediate case between these two *extreme* cases. Here are the three cases:

- $p = 0.9999$. In this case $a \approx 0.9899995$ and $b \approx 100.9949999$. Note that in this case the special distribution resembles a deterministic distribution of the value 1. With a probability that is very close to 1 (more accurately a probability of $p = 0.9999$) the special distribution returns a value that is very close to 1 (in similar to the deterministic distribution of the value 1). Note that with a very small probability, the special distribution with $p = 0.9999$ returns a very large value.

- $p = 0.5001$. In this case $a \approx 0.0002$ and $b \approx 2.0002$. Note that the case in which $p = 0.5$, the special distribution is exactly a distribution that with probability of $\frac{1}{2}$ returns a value of 0 and with a probability of $\frac{1}{2}$ returns a value of 2. Since we are dealing with service times, we want to keep all service times to be of positive values. Thus we have chosen p to be close to 0.5, and not to be exactly 0.5. Note that even though p is not exactly 0.5, the special distribution with $p = 0.5001$ is still very similar to the distribution that with equal probabilities returns the values of 0 or 2.

- $p = 0.75$. In this case $a \approx 0.4226497$ and $b \approx 2.7320508$. Note that this case where $p = 0.75$ is not similar to some other interesting case. We have chosen this value of p as a third case, since it is an intermediate case of the two first cases.

Performance of $M/G/100$ vs. β

In this section we will examine the performance of the $M/G/100$ system. We will present the results obtained from the simulation for the various statistics in the three different cases of the special distribution. Like previous chapters, the results were obtained by running 1000 simulations, each for 10000 time units (where a time unit is equal to the average service time). β was calculated as before: $\beta = \sqrt{N}(1 - \hat{\rho}_N)$ (for more accuracy we used $\hat{\rho}$, since $\hat{\rho}$ is almost always virtually identical to ρ).

Recall that for the *regular* distributions we received the surprising result that in the *Halfin-Whitt* regime, $E(W_q|W_q > 0)$ does not depend only on the first two moments of the distribution of the service time (was show two chapters before). We showed results in which the system $M/LN/100$ (where the CV of the service time distribution is 1) had lower waiting probabilities than the $M/M/100$ system (where the CV of the service time distribution is also 1) for each value of β. This difference became much clearer as ρ was closer to 1 (which is equivalent to β getting closer to 0). We received similar results for $P(Wait > 0)$ and for $E(W_q)$ statistics. According to the *Khintchin Polatchek* approximation, as ρ is closer to 1, $E(W_q)$ (and also $E(W_q|W_q > 0)$) are supposed to be approximately the same for the $M/M/100$ and $M/LN/100$ systems mentioned above. However our results show that for systems in the *Halfin-Whitt* regime this is not necessarily accurate.

Let us first examine $P(Wait > 0)$ vs. β. Recall that for the *regular* service time distributions, we received only one surprising result in the case of the $M/LN/100$ system (where the CV of the service time distribution is 1). In that case, $P(Wait > 0)$ was slightly smaller in the $M/LN/100$ system in comparison to the $M/M/100$ system for each value of β. The results for the three cases of the special service time distribution are presented in Figure 16.

Additionally, the full results for the special service time distribution and the *regular* distributions appear in Figure 17.

We obtained that for the two first cases (where $p = 0.9999$ and $p = 0.5001$), $P(Wait > 0)$ is smaller then in the case of $p = 0.75$, for each value of β. It can be intuitively explained why $P(Wait > 0)$ in the case where $p = 0.9999$ is smaller than the case where $p = 0.75$. The reason is that in the case in which $p = 0.9999$, the service time distribution is similar to a deterministic distribution concentrated 1. And since we obtained the lowest $P(Wait > 0)$ for a deterministic time service distribution for each β, among all the *regular* distributions, this result is no surprise. Additionally notice that the results of the special distribution in

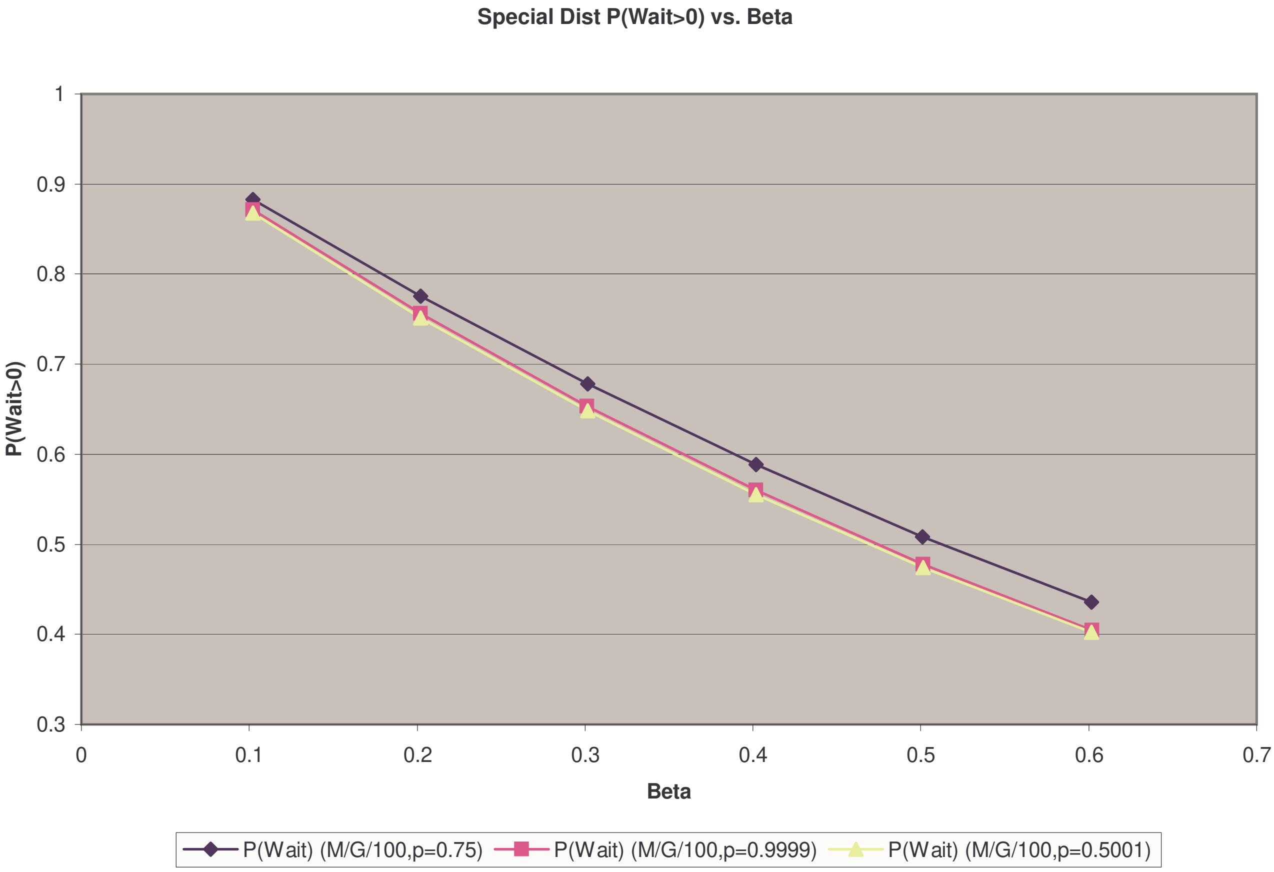

Figure 16: $P(Wait > 0)$ vs. β for special service time distribution

the case that $p = 0.75$ are almost identical to the results of the $M/M/100$ system, thus making the intuitive explanation more relevant.

Another surprising result concerns the special distribution in the case where $p = 0.5001$. In this case $P(Wait > 0)$ is low relative to **all** the other systems checked for all values of β. Though the difference can be negligible when comparing with $M/D/100$ and $M/G/100$ (where $p = 0.9999$) systems, it is still surprising that this in fact occurs. We have no explanation as to why this phenomenon happens (but it might be argued that since the case where $p = 0.5001$ is similar to a distribution with values 0 or 2 in equal probabilities, then such a distribution might cause the system to behave in a similar fashion to the system where the service time distribution is deterministic and equals to 1).

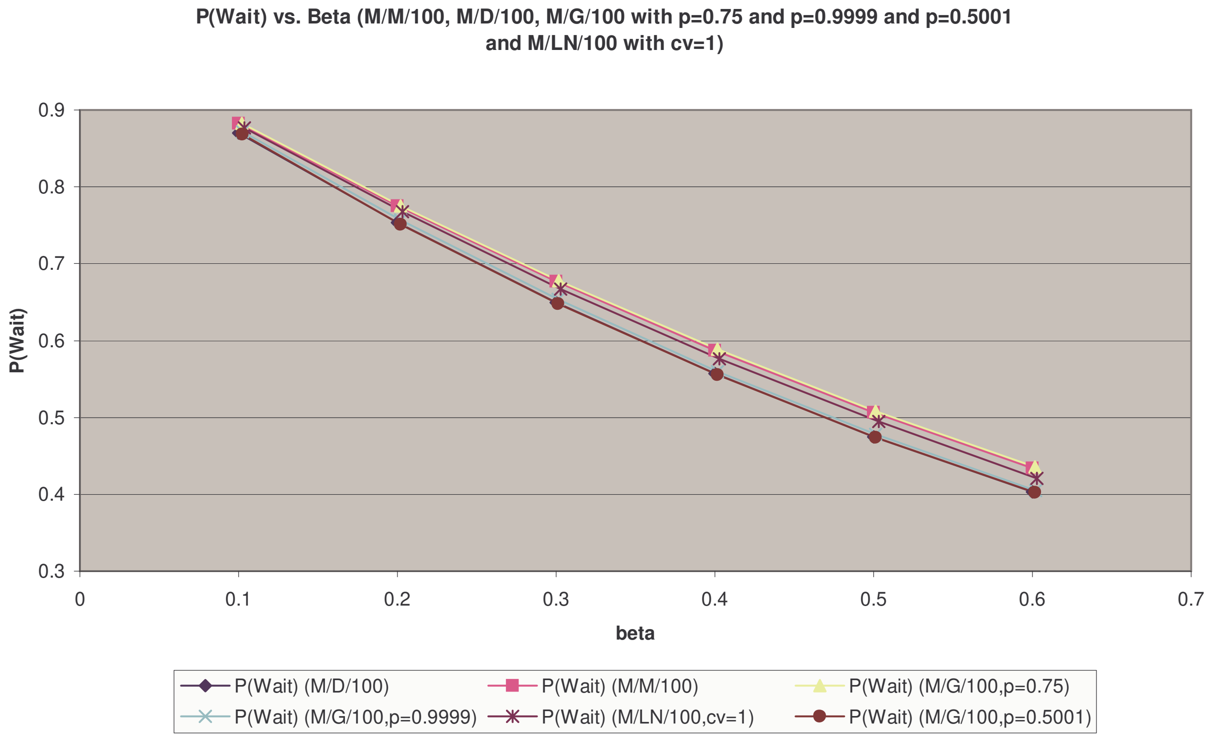

Figure 17: $P(Wait > 0)$ vs. β for special service time distribution and *regular* distributions

Let us examine now $E(W_q|W_q > 0)$ vs. β. Recall that for the *regular* service time distributions, like in the case of $P(Wait > 0)$ vs. β, we got only one surprising result in the case of the $M/LN/100$ system (where the CV of the service time distribution is 1). In that case, $E(W_q|W_q > 0)$ for the $M/LN/100$ system mentioned above was smaller than in the $M/M/100$ system, for each value of β. The conclusions of this phenomenon appear in a previous chapter (and were mentioned in this chapter only brief by). The results for the special service time distribution appear in Figure 18.

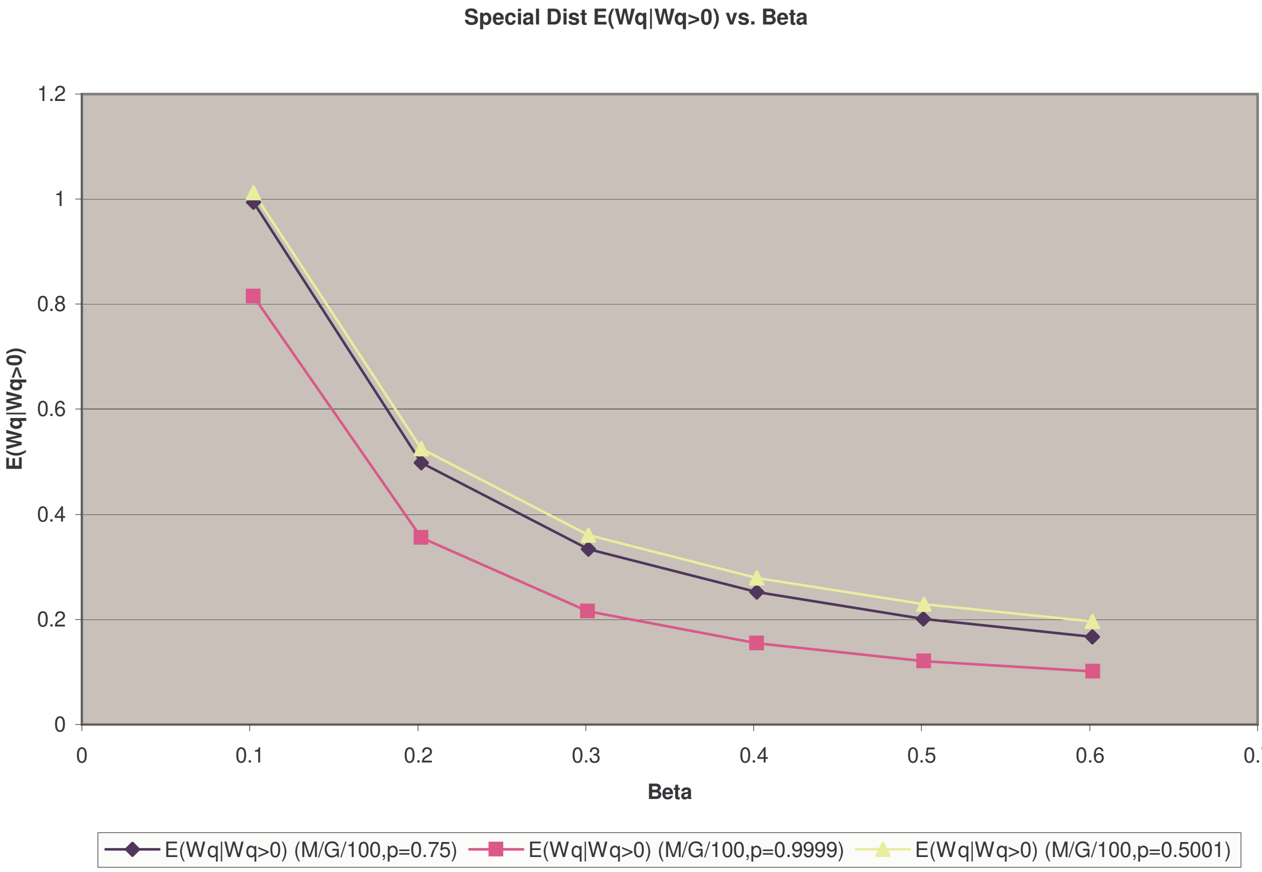

Figure 18: $E(W_q|W_q > 0)$ vs. β for special service time distribution

In addition, the total results for the special service time distribution and the *regular* distributions appear in Figure 19.

Note that for $E(W_q|W_q > 0)$, the order between the three cases of the special service time distribution has changed in comparison to the $P(Wait > 0)$ statistic (we define order according to the height of the lines in the graph). In $P(Wait > 0)$, the highest values aroused for the case $p = 0.75$ and the lowest for the case $p = 0.5001$. However, with $E(W_q|W_q > 0)$ the highest values are for $p = 0.5001$ and the lowest for $p = 0.9999$. The intuitive explanation for the fact that the lowest results were for $p = 0.9999$ is the same explanation that was given before as to why the case $p = 0.9999$ yields smaller waiting probabilities in comparison to the case where $p = 0.75$ (that the case $p = 0.9999$ is similar to the deterministic distribution with value of 1 and we know that such a deterministic distribution has yielded the lowest results for the *regular* distributions). However, we do not have any explanation as to why

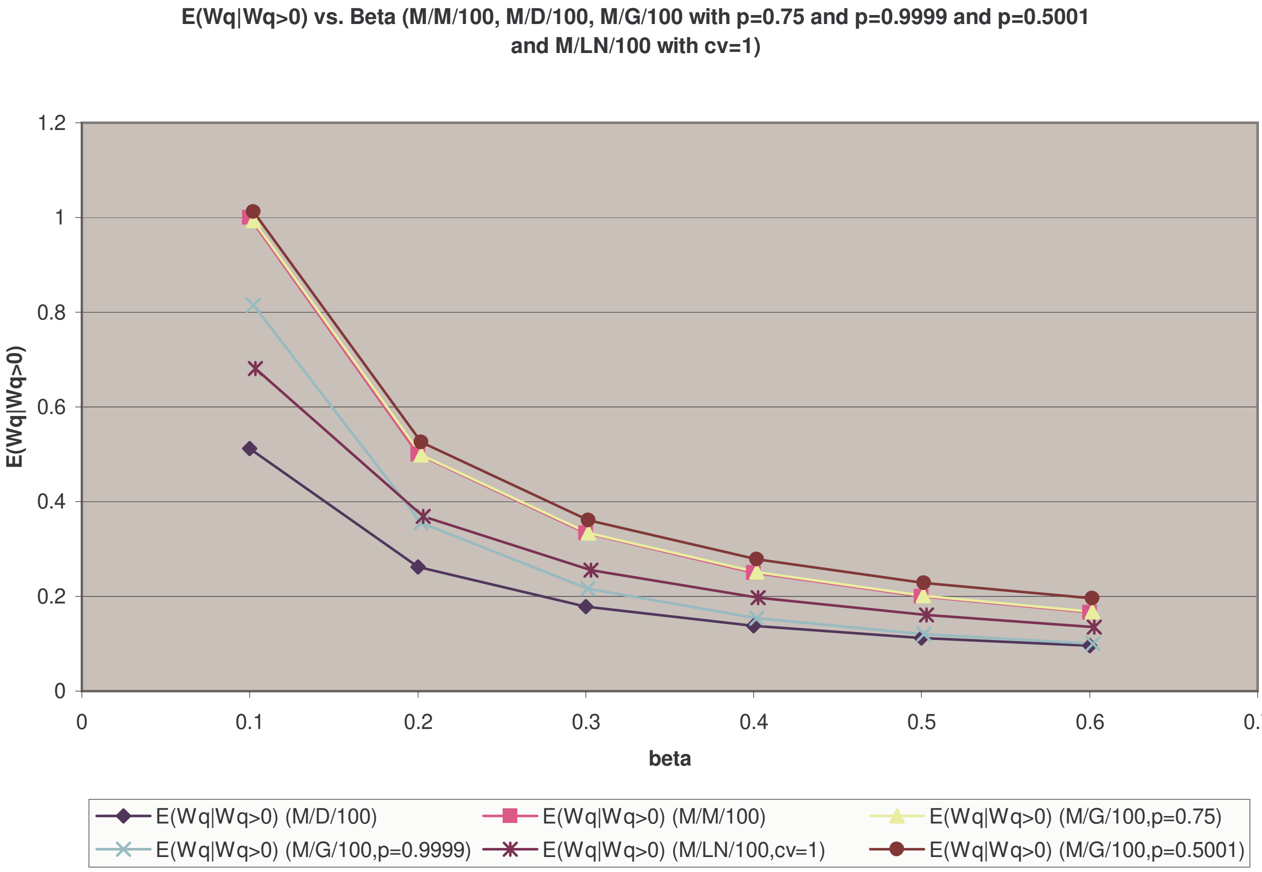

Figure 19: $E(W_q|W_q > 0)$ vs. β for special service time distribution and *regular* distributions

the case $p = 0.5001$ has the highest results in the $E(W_q|W_q > 0)$ case.

Notice that in all three cases of the special distribution, the slope of the corresponding line of each case becomes steeper as ρ gets closer to 1 (equivalently as β gets closer to 0) than the slope of the lines that correspond to the *regular* distributions. It might be worth mentioning that $E(W_q|W_q > 0)$ is the only statistic for which we have obtained lines that intersect with each other, when making graphs for a statistic vs. β (i.e. the only instance that line intersect is in Figure 19). Again, we have no explanation for this phenomenon (it might be some error in the simulation or a numerical inaccuracy , but we have not been successful in finding one). We would like to point out, that this intersection of lines is most significant in the three following systems: $M/LN/100$ (with coefficient-variance of 1), $M/D/100$ and $M/G/100$ (where $p = 0.9999$). Note that the line that represents $E(W_q|W_q > 0)$ in the case of $M/G/100$ (where $p = 0.9999$), for small ρ's is very close to the line that represents the case of $M/D/100$. However for high ρ's, this line is higher than that of the $M/LN/100$ system.

As in $P(Wait > 0)$, notice that the case $p = 0.75$ is very close to the $M/M/100$ system. Notice also that in the case $p = 0.5001$, $E(W_q|W_q > 0)$ is slightly higher, for each β value, than the $M/M/100$ case (as it was mentioned before, the order between the cases has changed).

The graph of $E(W_q)$ vs. β appears in Figure 20. Notice that the order of the lines in this graph is the same as the order of the lines in the graph of $E(W_q|W_q > 0)$.

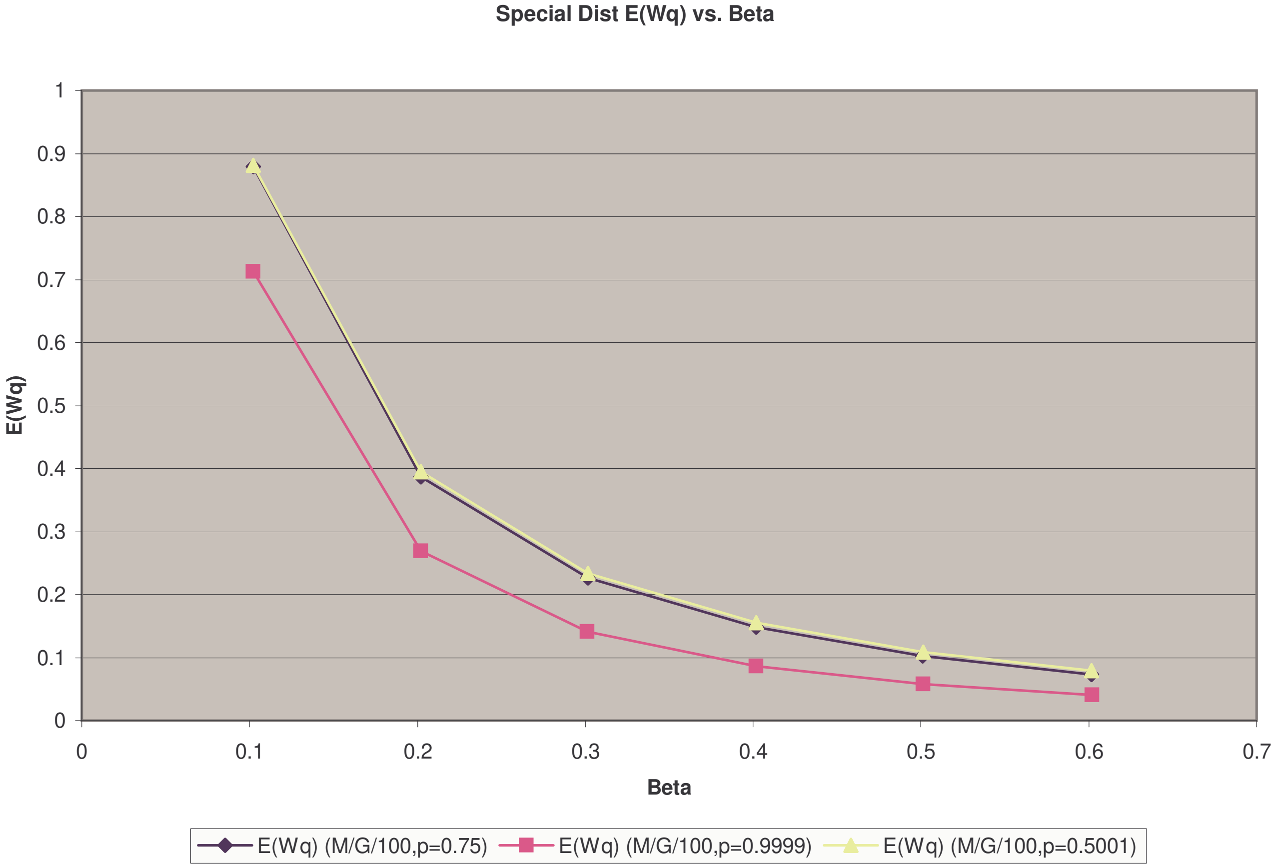

Figure 20: $E(W_q)$ vs. β for special service time distribution

Histograms of $W_q|W_q > 0$ for Special Service Time Distribution

In this section we will review the histograms of $W_q|W_q > 0$ which were obtained from the simulations. Since approximations of the distribution of $W_q|W_q > 0$ are asymptotically correct in classical heavy traffic as ρ is closer to 1 (for example take into consideration $Kingman$'s approximation), we will present the histograms for systems in the *Halfin-Whitt* regime with the highest ρ simulated (equivalent to the smallest β that appeared in the previous graphs).

According to $Kingman$'s law, the distribution of waiting time given wait will be exponential as ρ gets closer to 1. We will try and see if this is indeed the case with the different three cases of the special distribution of the service time in the *Halfin-Whitt* regime. Like in the previous chapter, we will conduct this check in two phases. In the first phase we will compare the histogram with a theoretical exponential density function and in the second phase we will check whether the function $(-\widehat{E}(W_q|W_q > 0)\ln(1 - \hat{F}(t)))$ is a straight line with a slope of 1 (the detailed description of these two check phases appeared in the previous chapter).

We will now present the results for the different three cases of the special distribution. We will start with the case $p = 0.9999$ which appears in Figures 21 and 22.

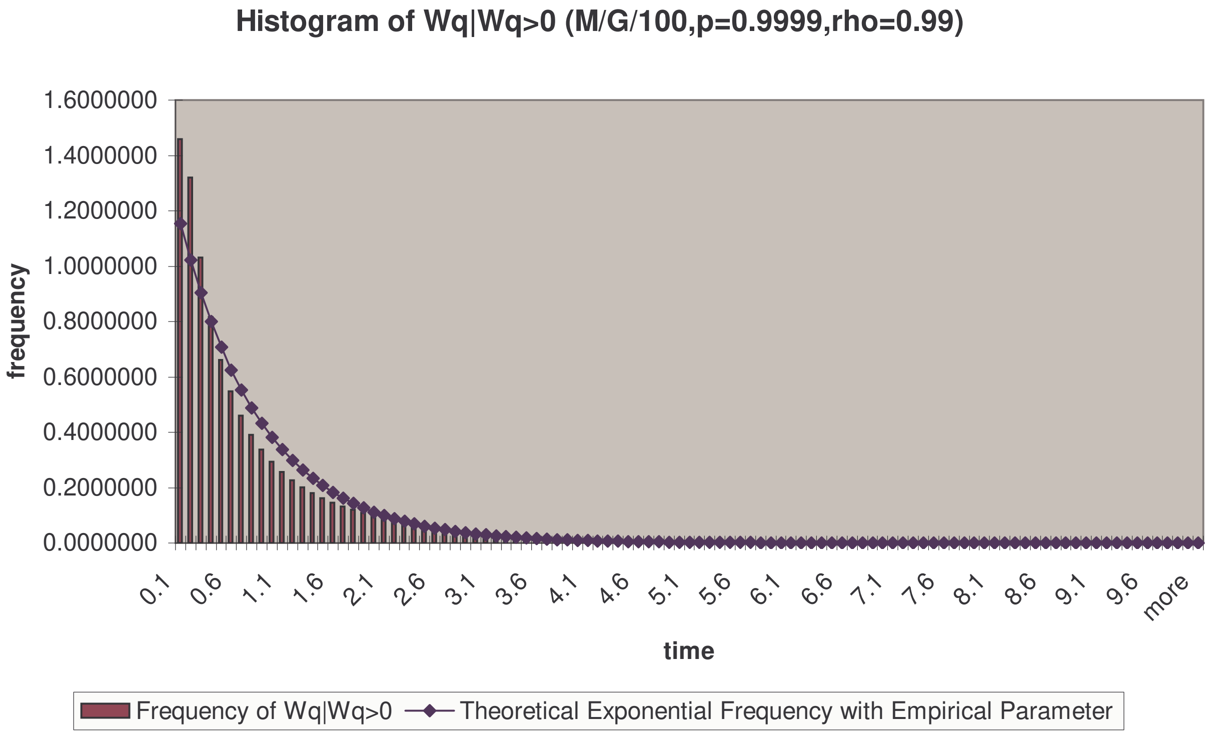

Figure 21: Histogram of $W_q|W_q > 0$ ($M/G/100$, $p = 0.9999$, $\rho = 0.99$)

Note that the empirical histogram of the density of $W_q|W_q > 0$ seems more concave than the theoretical exponential density function with the empirical parameter. This is manifested by the fact that the empirical histogram has a larger amount of distribution mass near 0 in comparison with theoretical exponential density function. However, note that the histogram still keeps its concave shape. In the second graph we can see that the

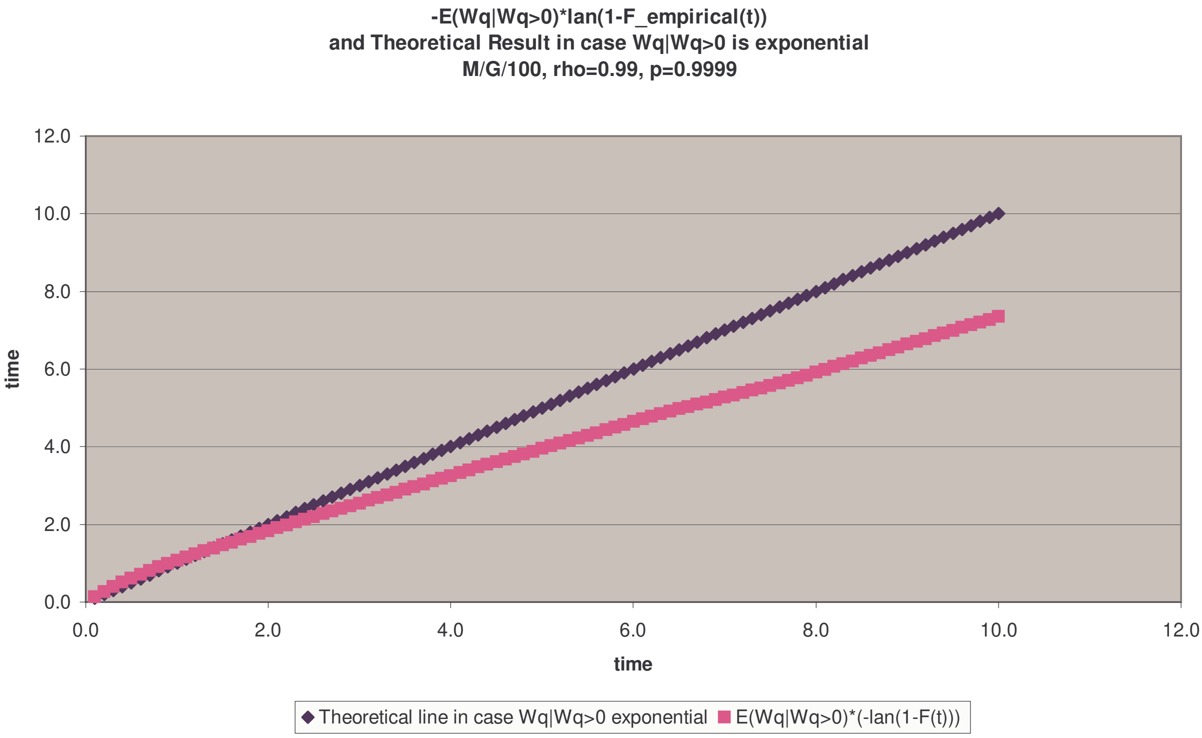

Figure 22: Checking Histogram of $W_q|W_q > 0$ ($M/G/100$, $p = 0.9999$, $\rho = 0.99$)

line that represents $(-\widehat{E}(W_q|W_q > 0)\ln(1 - \hat{F}(t)))$ is not a straight line with an inclination of 1, since it has a smaller inclination than 1 (even though it seems like a straight line except near 0). Conducting the analysis of the tail of the distribution of $W_q|W_q > 0$ as done in the previous chapter, it can be concluded that the distribution of waiting time given wait in the case $p = 0.9999$ has a heavier tail than an exponential distribution. The reason for this conclusion is that $(-\widehat{E}(W_q|W_q > 0)\ln(1 - \hat{F}(t))) < t$ for almost every t (except maybe for t values that are very close to 0), hence $\hat{F}(t) < F(t)$ (where F is a cumulative exponential distribution function).

The results for the case $p = 0.5001$ appear in Figures 23 and 24.

Note that the empirical histogram of the density of $W_q|W_q > 0$ does not seem to be similar to an exponential distribution. The reason is that there is less distribution mass near 0 in comparison with the theoretical exponential density function with the empirical parameter. The histogram does not have the concave shape that an exponential distribution function has, near 0 it is convex and far from 0 it is concave. To strengthen the assertion that in the case where $p = 0.5001$ the histogram of $W_q|W_q > 0$ is not exponential, we can look at the second graph and see whether the function $(-\widehat{E}(W_q|W_q > 0)\ln(1 - \hat{F}(t)))$ is a straight line with an inclination of 1. It can be clearly seen that the mentioned function is not such a straight line, since for large values of t it curves upwards (thus the above function is not even a straight line). Conducting the analysis of the tail of the distribution of $W_q|W_q > 0$ as done before, we can conclude that the distribution of waiting time given wait in the case $p = 0.5001$ has a lighter tail than an exponential distribution.

Summing the two first cases of the special service time distribution, it can be stated

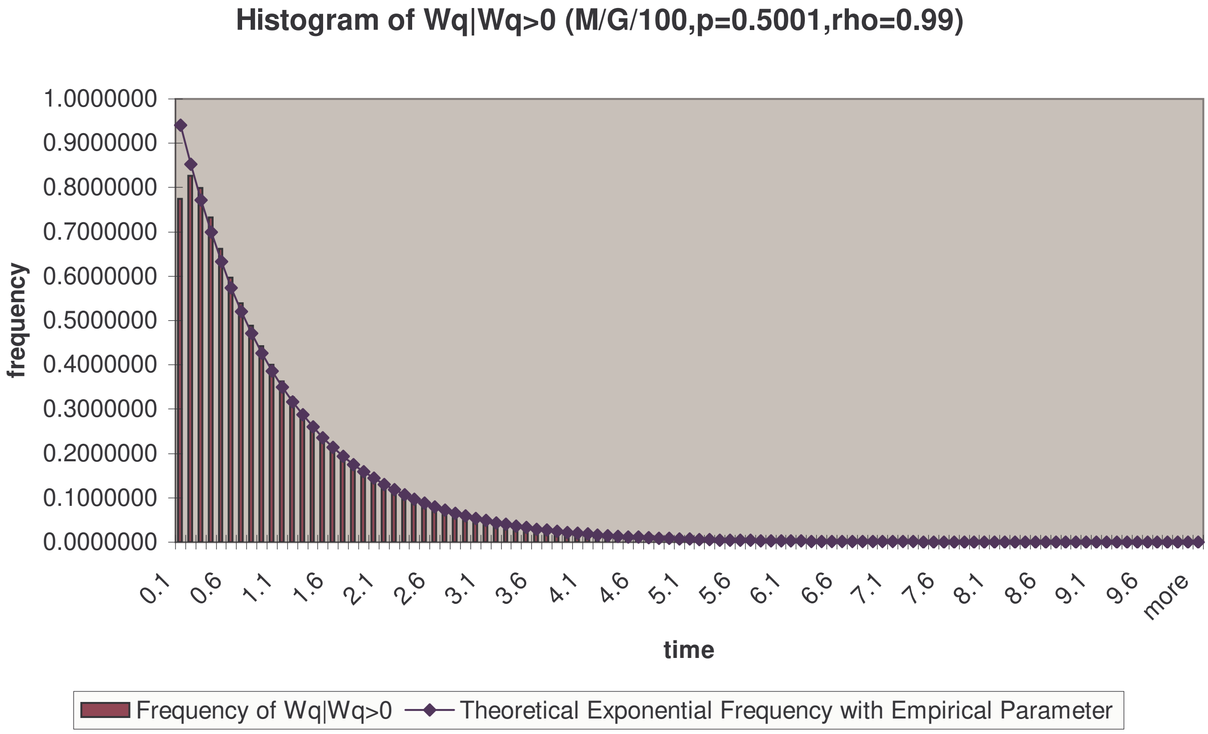

Figure 23: Histogram of $W_q|W_q > 0$ ($M/G/100$, $p = 0.5001$, $\rho = 0.99$)

that waiting time given wait is not necessarily exponentially distributed. Thus we have strengthened the claim that was made in the previous chapter, that in the *Halfin-Whitt* regime *Kingman*'s approximation may not be accurate.

The results for the case $p = 0.75$ appear in Figures 25 and 26.

Note that the empirical histogram of the density of $W_q|W_q > 0$ seems to be very close to an exponential density function. By looking at the two graphs, one can see that in the first graph the empirical histogram and the theoretical exponential density with the empirical parameter seem almost identical. In the second graph one can see that the function $(-\widehat{E}(W_q|W_q > 0)\ln(1-\hat{F}(t)))$ seems like a straight line with an inclination of about 1. Even though it seems that the inclination is slightly larger than 1, we can say in general fashion that the histogram of the density of $W_q|W_q > 0$ seems to be an exponential distribution.

Summing the results of this chapter, one can say that there are still many phenomena that do not have an explanation (by explanation we of course mean a mathematical proof). Not only do the checked statistics not behave as predicted in the *Halfin-Whitt* regime, but the distribution of $W_q|W_q > 0$ seems to be not exponential in several cases.

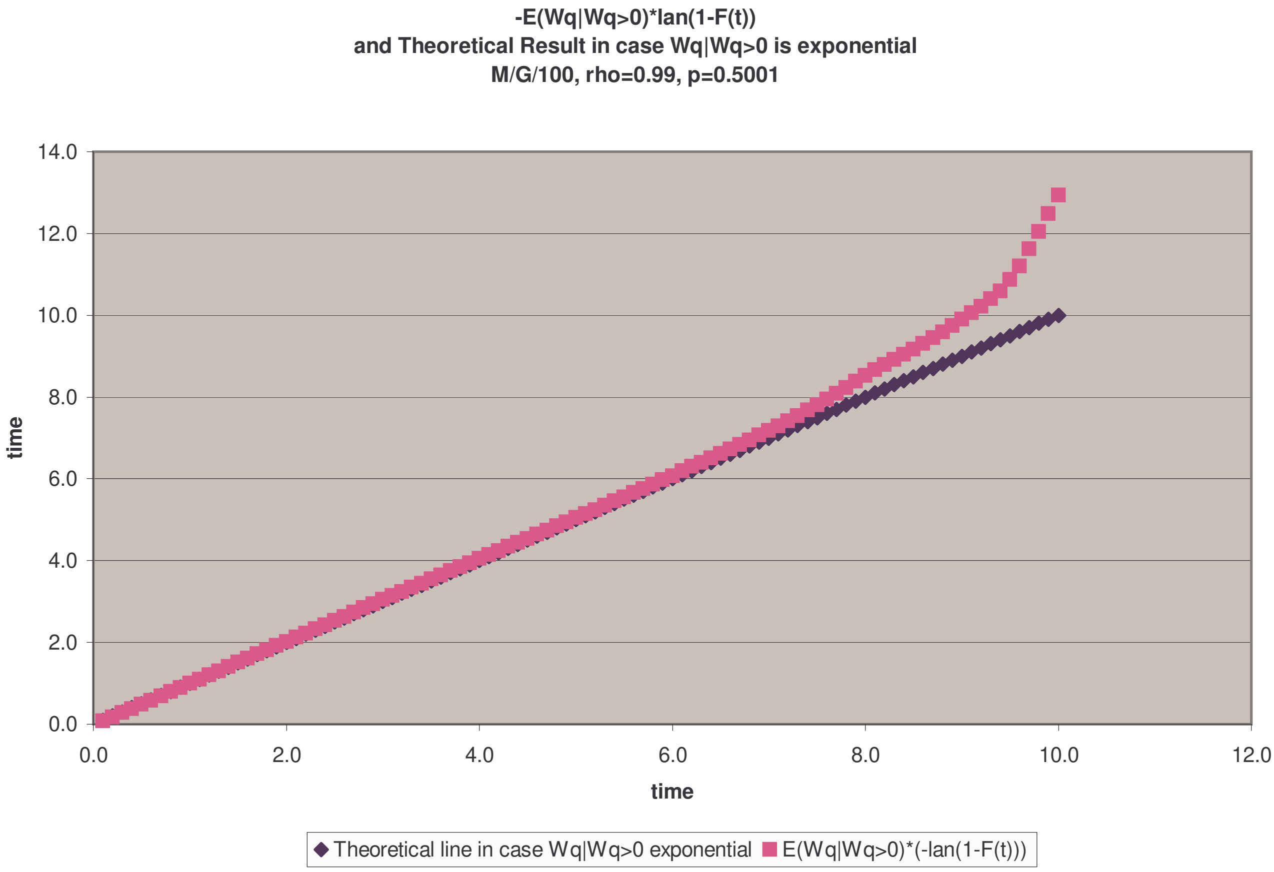

Figure 24: Checking Histogram of $W_q|W_q > 0$ ($M/G/100$, $p = 0.5001$, $\rho = 0.99$)

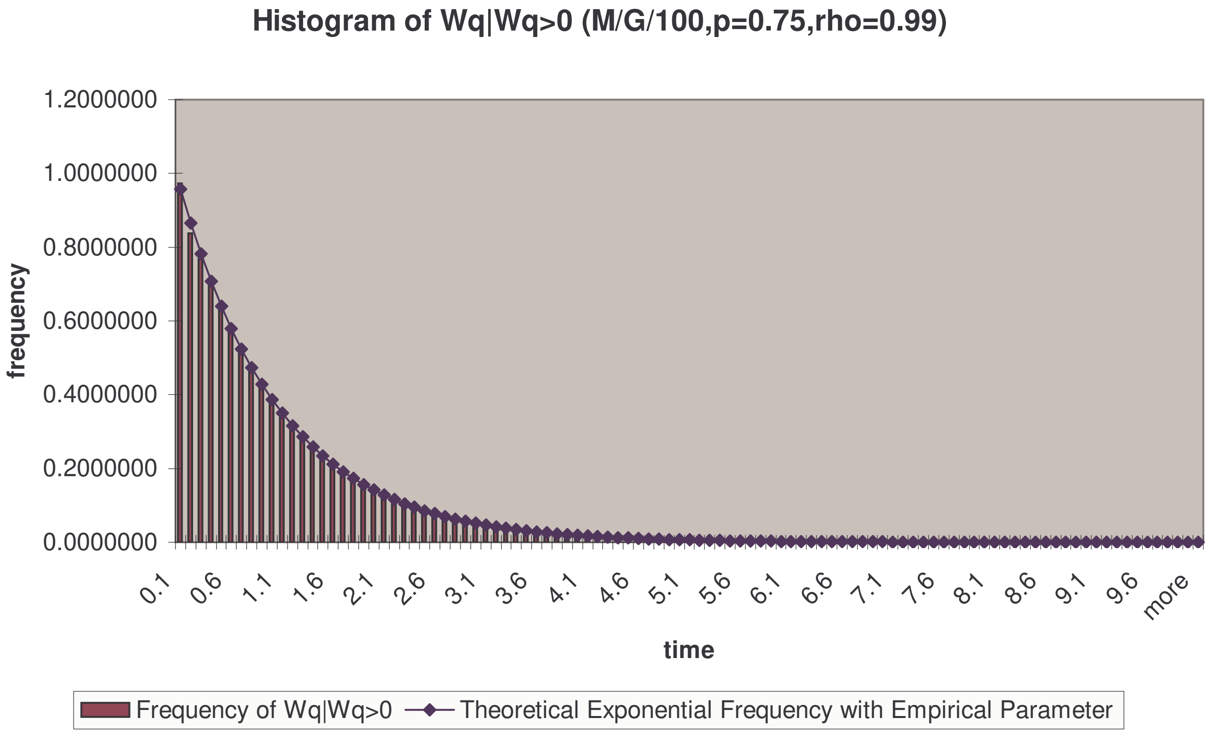

Figure 25: Histogram of $W_q|W_q > 0$ ($M/G/100$, $p = 0.75$, $\rho = 0.99$)

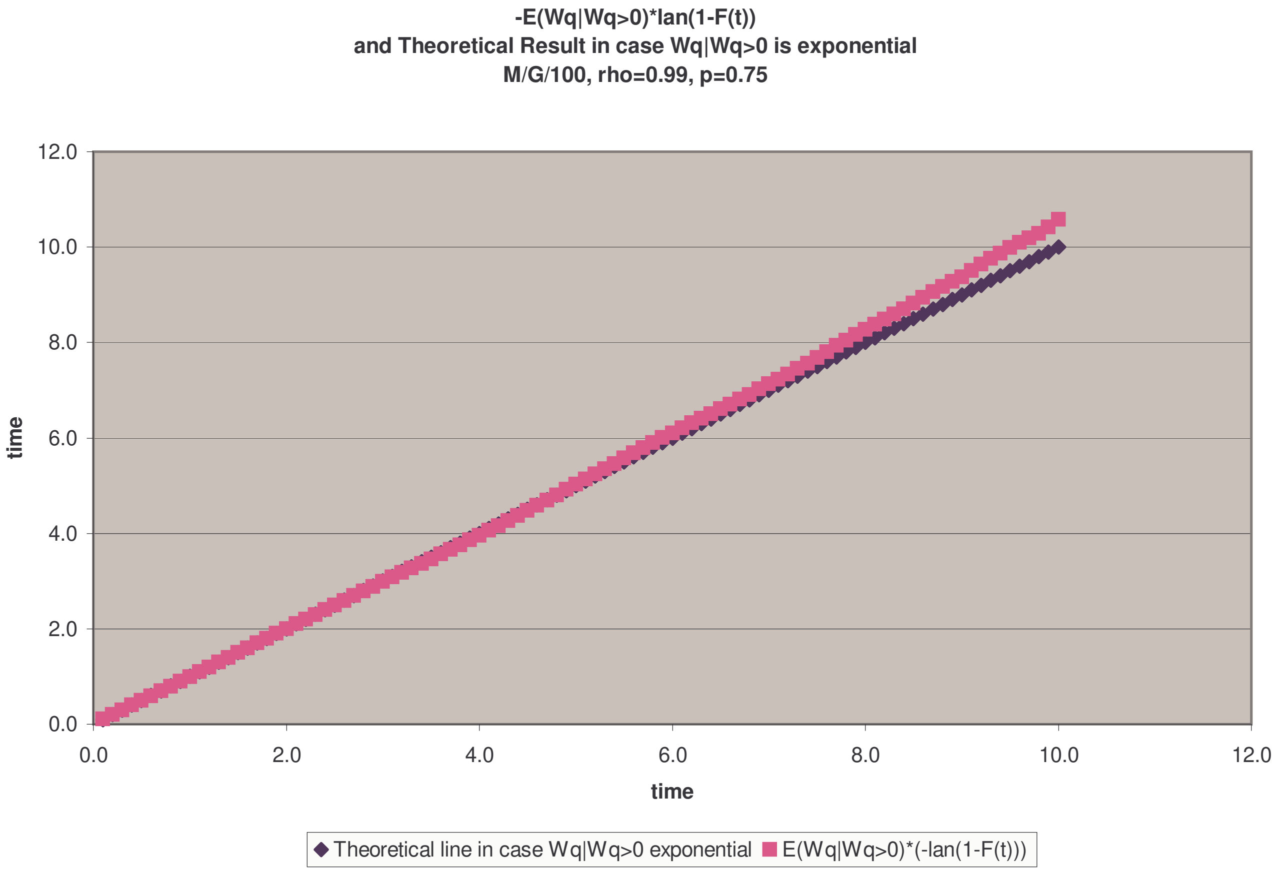

Figure 26: Checking Histogram of $W_q|W_q > 0$ ($M/G/100$, $p = 0.75$, $\rho = 0.99$)