Deployment of express checkout lines at supermarkets

Maarten Schimmel

Research paper Business Analytics
April, 2013
Supervisor: René Bekker

Faculty of Sciences
VU University Amsterdam
De Boelelaan 1081
1081 HV Amsterdam
Netherlands

Preface

This research paper is a compulsory part of the Master's program in Business Analytics at VU University Amsterdam. The goal of this course is to produce a paper that is geared towards a relevant business subject, based on mathematics and computer science skills learned during this study.

I would like to thank my supervisor René Bekker, for his valuable support during the making of this research paper.

Summary

One of the most common complaints at supermarkets is the long queues and waiting times at checkout lines. Shorter perceived waiting times increases service experience, customer loyalty and potentially increases market share. The perception of waiting time is important, customers with fewer items can perceive a longer waiting time, because the amount of items a customer buys has an influence on the tolerance towards waiting. This is a reason to implement express checkout counter, so customers with a few items do not have to wait long in line.

The range of possible expected waiting times depends on the worst and best case situations on queuing time. The waiting times are hard to determine exactly, but best and worst cases can be more easily identified. The worst case model assumes random selection of a queue when a customer arrives at checkout, and possible picks a line with a lot of waiting customers. The worst case is based on S times an M/G/1 queue, with S the amount of checkout counters.

The best case assumes the perfect world, where customers always pick the line with the least amount of work, and go directly to an available counter, not leaving customers wait in another queue while a cashier is idle. The best case is based on a single M/G/S queue, with S again the amount of checkout counters.

In practice, every counter has its own queue, suggesting that it functions as an M/G/1 queue. However, since customers join the shortest queue, or the queue with the least amount of work, it may also mimic an M/G/S queue. We have simulated a system where customers join the shortest queue. The results indicate that an M/G/S queue system is approached.

With the models for best and worst case a limit on the amount of items a customer is allowed to bring to the express checkout can easily be calculated. The most important constraint is that a high load on the system should be assumed, since with lower loads a large range of amount of items is allowed and a high load finds a small allowed range. In stores in general it is preferred to have cashiers helping customers than being idle, making it a valid assumption. Using simulation a better limit can be found, especially with lower loads on the system; when there are multiple limit possibilities.

When using express lines the combined expected waiting time of express and regular customers almost always increases, because there is only small variance in service times. This is in contrast with server farms which for instance can have large variation in service times. When express customers have a substantial decreased expected waiting time, regular customers have a relative high increase in waiting times. When this is compared to the service time regular customers have this seems more reasonable. This all increases the fairness by reducing the waiting time of express customers and increasing the expected waiting time of regular customers. Although the fairness may be increasing the combined expected waiting time also increases, making people waiting longer on average, only express customers benefit from this system because regular customers sometimes may have to wait considerable longer.

Contents

1 Introduction

Long queues and waiting times are one of the most common complaints at supermarkets. Shorter waiting times can increase service experience, customer loyalty and thereby potentially increases market share.

The main aim of this thesis is two-fold. First we determine models for the expected waiting time in the supermarket checkout queue. Secondly we look to the effect of express checkout counters on expected waiting time.

The perception of waiting time is important; customers with fewer items can perceive a longer waiting time than customers with more items even if the customers waiting time was the same. This is a psychological aspect of fairness: the more items you buy the longer you expect to wait. When you do not divide customers in groups, like express and regular customers, customers with a few items have the same expected waiting time as customers with large number of items. Supermarkets use multiple methods to reduce the waiting time at checkout. They use traditional methods, like express lines, and technology like self scan checkout technology to reduce the waiting time. This technology may help enhance the customer experience, because customers overestimate time on passive actions and underestimate time of active durations like scanning their products.

When supermarkets decide to use express checkout counters, supermarkets try to determine how many express and regular checkout lines are needed. Another issue will be the limit on the amount of items a customer can bring to an express checkout counter. Supermarkets constantly try to reduce costs, so they might want to know if cost reduction can be achieved while keeping the same service level.

To give an expectation of waiting time in the queue, we need to know how to approximate the waiting time. The exact customer behavior is difficult to model resulting in complex approximating models. Research showed that the expected waiting times have values between the waiting times of M/G/S and M/G/1 queues, where the M/G/1 queue has an arrival intensity of λ/S. Here M/G/s means one queue for S checkout counters and M/G/1 means each counter has a separate queue.

In Section 2 we give a literature overview on express checkout counters, and why the worst case and best case are lower and upper bounds. In Section 3 notation is introduced, some mathematical models are introduced, and assumptions are stated. Section 4 we show the influence of express checkout counters on the waiting times in the queue. A model is developed to simulate the situation and is compared to the theoretical best and worst case situations to determine the accuracy of the simple models. And the simulations can show the potential need for simulation or more complex theoretical models to determine the limit on items and other measures. Section 5 contains the conclusion and possible recommendations.

2 Literature

2.1 Use of express checkout lines

Waiting times perceived by customers is in general different from the real waiting time. This could be because people do not experience time linearly [Stevens (1957)]. When people wait, a passive action, they feel like the time moves slower, and when they are active or being served at checkout they experience time faster.

Another aspect of experienced time is the fewer items customers buy the faster customers are annoyed by waiting times [Maister (1985)]. Reactions of customer to waiting time are mainly based on perceived waiting time and not on real waiting time. This means that it is important to reduce the perceived waiting time rather than the real waiting time. This is one of the reasons express checkouts were introduced; to reduce the waiting time for customers with less items, because the service people wait for is important for perception.

The relative overestimation on waiting time is the same across all types of waiting lines, regardless of a snake line (one queue for all customers), separate queues for each counter or the use of express lines. The only significant difference between overestimation on waiting times could be found in the frequency customers shop and pleasure they perceive shopping [Hornik J(1984)].

2.2 Worst- & Best case

The advantages of worst- and best case models are the simplicity and intuitive upper and lower limits they give.

Customer behavior is hard to capture in a model, in this case because there are so many possible choices a customer can make in reality. But there are clear theoretical boundaries on the expected waiting time of this queuing checkout system, in example an upper and lower limit on expected waiting time.

The worst case is based on the queues as they are generally implemented; each counter has its own queue also known as M/G/1 queues. The upper limit is determined by a queue for each counter, where each customer randomly selects a queue, this can result in long waiting times in some queues. The best case is based on a single queue for all customers also known as M/G/S, or in this case one queue for all regular customers and one queue for all express customers. For an image of the best case see Figure 1, for worst case and reality see Figure 2, Figure 3.
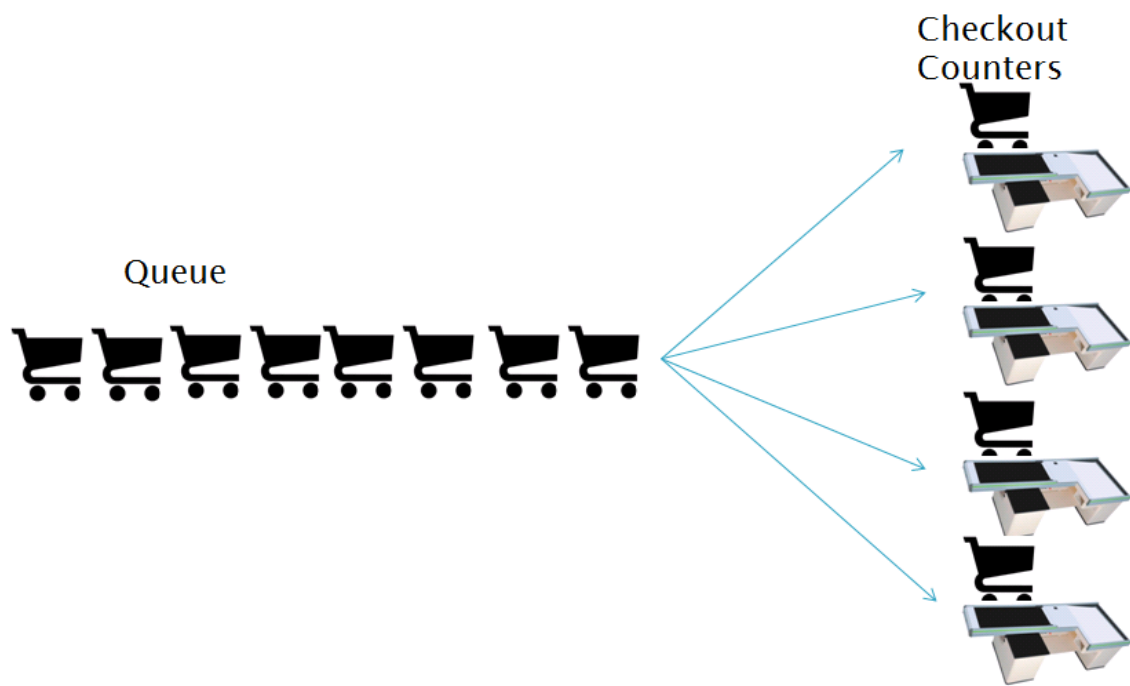
Figure 1: Best case (M/G/S)
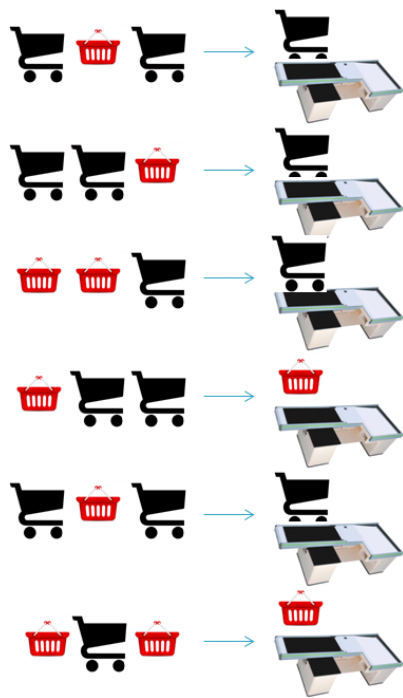
Figure 2: M/G/1 queues without Express
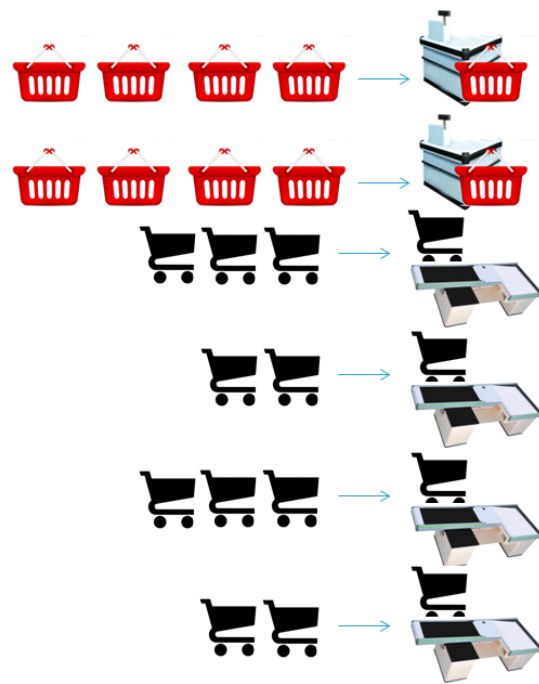
Figure 3: M/G/1 queues 2 Express lines 8 Regular lines

Both models are not totally accurate because customer behavior does not exactly mimic these models. A key factor of the best case (M/G/S) approach is the counters being used optimally so there are no idle servers when there are people in the queue. This gives an underestimation of the expected waiting time, also the order of serving has an influence but not on the expectation only on

the distribution. Since in reality there are multiple queues customers are not always able to pick the best line; for instance when there is an obstructed view they cannot always detect the exact amount of customers in queue and their amount of items. Changing queue when another queue seems better is not always possible or wanted because of the walking time and the risk of a new customer arriving at checkout, this will give an underestimation in waiting time.

The M/G/1 approximation, assumes random selection of a queue. In this case it is possible that customers go stand in line while some of the counters are idle. In reality people do not pick a line at random, people assess the lines and try to pick the best line [Rothkopf and Reich (1987)]. So in reality cashiers are hardly idle while there are waiting customers, because people jockey to those lines. This suggests that an M/G/S, Best case system gives more accurate results.

2.3   Why use simulation?

Simulation is commonly used in queueing theory, because the theoretical approach fast becomes complicated. Simulation is more easily adaptable and intuitively clear.

In real life people choose the shortest queue, switch lanes and choose the queue they think will go the fastest. The diverse and highly dynamic rules that customers use to choose a specific lane is hard to capture by theoretical models and such models are typically extremely hard to analyze. But with simulation you can make assumptions that keep performance analysis possible.

3   Model description

Before the whole model is being described we introduce some assumptions and variables. The assumptions used in these calculations are similar to [Koltai, Kalló & Lakatos (2008)].

We assume each customer has a fixed payment duration a and an identical processing time per item of b, resulting in a linear function with a constant time for payment. All the randomness in t_i comes from the amount of items a customer brings to checkout.

$$t_i = a + b \cdot i, \quad i = 1, \dots, K \quad ,$$

K the maximum amount of items a customer can buy, a the payment duration and b the processing time of a single item

We assume a payment duration of a= 0.75 minute and the processing time of a single item b=0.05 minute. Customers arrive at checkout according to a Poisson process with arrival rate λ. The limit on the amount of items a customer can bring to an express checkout counter is L. The maximum amount of items customers can buy is K. We assume ρ<1 so a steady state exists. We also assume that people who have less or equal amount of items as the limit always enters the express line and customers with more items than the limit always go to a regular checkout counter. For the item distribution see Section 3.2.

See Table 1 for all the variables.

λ	Total arrival rate
$λ_E$	Arrival rate of Express customers
$λ_R$	Arrival rate of Regular customers
L	Maximum items customer can bring to express checkout
K	Maximum amount of items a customer can buy
S	Total number of checkout counters
E	Number of express checkout counters
R	Number of regular checkout counters
B	Expected service time
p_i	probability a customer buys i items
t_i	Service time of a customer buying i items
a	Payment duration
b	Scan time one item
t_E	Expected service time of Express customers
t_R	Expected service time of Regular customers
B_E	Service time of Express customers
B_R	Service time of Regular customers
σ_E^2	Variation of service times in Express queue
σ_R^2	Variation of service times in Regular queue
t_q	Weighted average waiting time in queue
$t_{qE}\{M/G/1\}$	Expected waiting time Express queue in Worst Case
$t_{qR}\{M/G/1\}$	Expected waiting time Regular queue in Worst Case
$t_q\{M/G/1\}$	Combined expected waiting time in Worst Case
$t_{qE}\{M/G/s\}$	Expected waiting time Express queue in Best Case
$t_{qR}\{M/G/s\}$	Expected waiting time Regular queue in Best Case
$t_q\{M/G/S\}$	Combined expected waiting time in Best Case

c_b^2	Squared coefficient of variation
c_{bE}^2	Squared coefficient of variation of Express customers
c_{bR}^2	Squared coefficient of variation of Regular customers
ρ	Load on system
ρ_E	Load on Express checkout
ρ_R	Load on Regular checkout

Table 1: Variables used

3.1 Splitting in customer types

When express lines are used distinction need to be made between express and regular customers. Here we split by the amount of items a customer brings to checkout.

Poisson arrivals can be split in multiple Poisson arrival processes. So one stream of customers arriving according to a Poisson process can in this case be separated in two groups respectively express and regular customers and both groups have again a Poisson arrival processes. There are S checkout counters of which there are E express counters and R regular counters (E+R=S)

$$\lambda_E = \lambda \sum_{i=1}^{L} p_i \qquad \text{and} \qquad \lambda_R = \lambda \sum_{i=L+1}^{K} p_i \ ,$$

with p_i the probability a customer buys i items

The average service time of express checkout counters (t_E) is based on customers with at most L items and the average service time of regular customers (t_R) is based on customers with more than L items.

$$\boldsymbol{E}(B_E) = t_E = \frac{\sum_{i=1}^{L} p_i(a + b \cdot i)}{\sum_{i=1}^{L} p_i} \qquad \text{and} \qquad \boldsymbol{E}(B_R) = t_R = \frac{\sum_{i=L+1}^{K} p_i(a + b \cdot i)}{\sum_{i=L+1}^{K} p_i}$$

Because the service time is a function of the amount of items a customer brings to checkout, we assume a general distribution, a known variance in service time. Because a constant payment service time every customer with the same amount of items will have the same service time. This leaves only variation in service times between customers with different number of items. For convenience we calculate the variation based on a randomly generated set of items according to the chosen distribution.

$$\sigma_E^2 \ \text{and} \ \ \sigma_R^2 \ \text{will be calculated based on a generated data set}$$

3.2 Item distributions

In this section we consider two distributions used for the amount of items customers bring to the checkout counters, namely a lognormal distribution in Section 3.2.1 and a combined lognormal distribution in Section 3.2.2

3.2.1 Lognormal

We assume the amount of items customers buy will be estimated by a Lognormal distribution, with μ=1.98 and σ=1. This gives an average of about 12 items per customer (11.941 to be precise) see Figure 1Figure 4. We use this assumption because a lack of real supermarket item distribution data based on the average checkout value of 22.10 euro for the year 2011 in supermarkets in the Netherlands [GfK Supermarktkengetallen Oktober 2012]. The only data we could find was a geometric item distribution as suggested by a Colombian study on supermarket cashiers. But this study has no clear reason for this distribution other than that it is a discrete function.
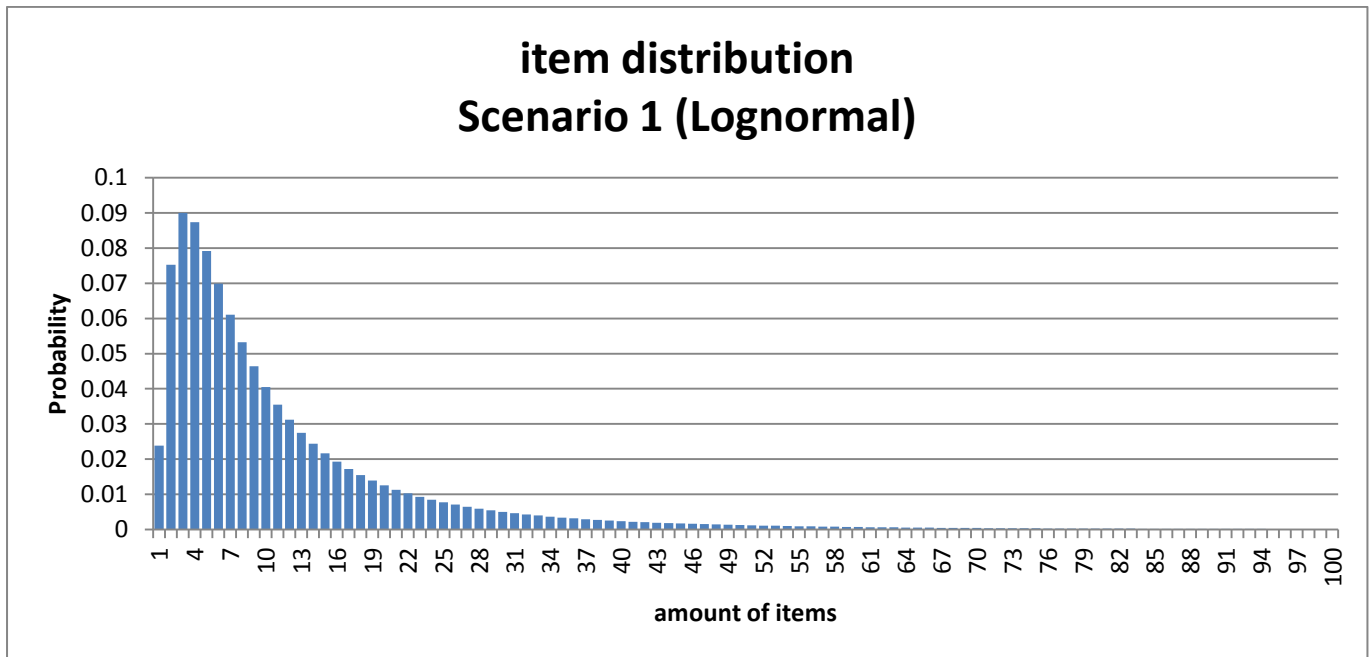

Figure 4: Item distribution lognormal

3.2.2 Combined lognormal distribution

For the second scenario we let the amount of items customers buy will be estimated by two Lognormal distributions, the first one with a weight of 40 percent with parameters μ=1.98 and σ=1 and one with parameters μ=4 and σ=0.25 for the other 60 percent, See Figure 5. This gives a scenario with two types of customers, one group with a few items and a group with a large amount of items; to create a more spread in the amount of items customers bring to the checkout counters.
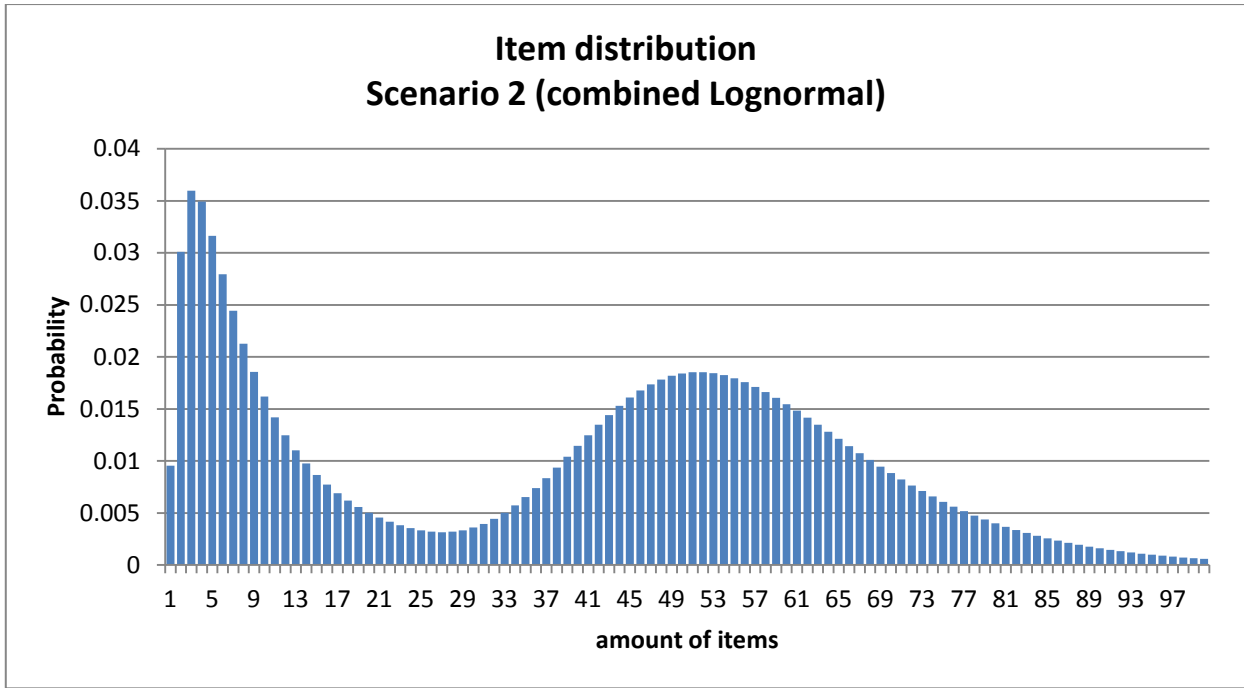
Figure 5: Combined lognormal item distribution

3.3 Scenarios

Here we explain the characteristics of the used scenarios; all the scenarios use a total of 10 checkout counters.

3.3.1 Scenario 1 (Lognormal distribution)

We assume the amount of items customers buy will be estimated by a Lognormal distribution, with $\mu=1.98$ and $\sigma=1$, as shown in Section 3.2.1

Scenario 1.1 (high load)

Scenario 1.1 consists of 2 express checkout counters and 8 regular checkout counters. Total of 400 customers arrive per hour at the checkout. The amount of items brought to checkout is distributed by a lognormal distribution as showed in Section 3.2.1 with a mean of 11.941 items bought per customer. This gives a total load on the system of 0.898

Scenario 1.2 (medium load)

Scenario 1.2 has a lower amount of arrivals than scenario 1.1; this scenario has a total of 300 arrivals per hour at checkout. This results in an overall load on the system of 0.674

Scenario 1.3 (low load)

Like scenario 1.2 only the amount arrivals per hour changes, this time to 200 arrivals per hour. Resulting in a total load of the system of 0.449

3.3.2 Scenario 2 (Combined Lognormal distribution)

We also used a scenario with almost the same load on the system as the high load scenario 1 but now we combine two lognormal distributions; see Section 3.2.2; to create a more spread in the amount of items customers bring to the checkout counters. With a total of 200 arrivals per hour,

which means there are less customers per hour than scenario 1 but they buy more items. We still use a configuration of 2 express- and 8 regular checkout counters. This gives an overall load on the system of 0.893

3.4   Worst case (M/G/1) scenario

The worst case scenario is based on the M/G/1 model, where each checkout counter has a separate individual queue. This case is based on random selection of a line when a customer arrives at checkout.

As mentioned before the arrivals of customers are assumed to be Poisson. This means the arrivals, when randomly split in groups still have Poisson properties. So here customers with few items and customers with large amount of items, repectively express and regular customers can be split in separate groups. Both groups will have again Poisson arrivals

We use the Pollaczek–Khinchine formula to calculate the mean waiting time in the M/G/1 situation

$$\text{With } c_b^2 = \frac{Var(\pmb{E}(B))}{\pmb{E}(B)^2}$$

This means for the express queue expected waiting time per queue:

$$t_{qE}\{M/G/1\} = EW_{QE} = \frac{1}{2}\big(1 + c_{bE}^2\big)t_E \frac{\rho_E}{1 - \rho_E}$$

And the formula for the regular queue:

$$t_{qR}\{M/G/1\} = EW_{QR} = \frac{1}{2}\big(1 + c_{bR}^2\big)t_R \frac{\rho_R}{1 - \rho_R}$$

This gives the following combined expected waiting time in the queue

$$t_q\{M/G/1\} = t_{qE}\{M/G/1\} \sum_{i=0}^{L} p_i + t_{qR}\{M/G/1\} \sum_{i=L+1}^{K} p_i$$

3.5   Best case (M/G/S) Scenario

The best case scenario is based on the M/G/S model; with one queue for all counters (see Figure 1). The M/G/S model underestimates the real waiting times of customers. The M/G/S model can still be used because in real life situations customers do not choose a random queue, but try to enter the queue with the shortest waiting time, and may act the same as an M/G/S queue. [Rothkopf and Reich (1987)] .

For the approximation we use the method as proposed in [Whitt 1993,1999],

$$EW(M/G/S) \approx \frac{(1+c_b^2)}{2} EW(M/M/S) = \frac{(1+c_b^2)\pmb{E}(B)}{2s(1-\rho)} P(\ W(M/M/S) > 0),$$

We approximate P(W(M/M/S) > 0) based on the Sakasegawa approximation [Sakasegawa 1977]

$$P(W(M/M/S)>0) \approx \rho^{\sqrt{2(S+1)}-1}$$

Although the Sakasegawa approximation approaches the M/G/S situation closely for high loads, at lower loads it highly overestimates the probability of waiting and thus the expected waiting time. Because we are interested in high traffic situations this is acceptable.

Now we apply this to the two types of customers. Because we assume Poisson arrivals, we can split the arrivals and calculate the expected waiting time in the queue for both the express and the regular line.

The express queue expected waiting time:

$$t_{qE}\{M/G/S_E\} = EW_{QE} \approx \frac{(1 + c_{sE}^2)t_E}{2s_E(1 - \rho_E)} \rho_E^{\sqrt{2(S_E+1)}-1}$$

And the formula for the regular queue:

$$t_{qR}\{M/G/S_R\} = EW_{QR} \approx \frac{(1 + c_{bR}^2)t_R}{2s_R(1 - \rho_R)} \rho_R^{\sqrt{2(S_R+1)}-1}$$

We can combine both to calculate the average waiting time in the queues.

$$t_q\{M/G/S\} = t_{qE}\{M/G/S\} \sum_{i=0}^{L} p_i + t_{qR}\{M/G/S\} \sum_{i=L+1}^{K} p_i$$

3.6   Assumptions simulation model

We used a computer simulation tool named Rockwell Automation Arena 14 to implement the simulation model of the queuing system.

The model uses Poisson arrivals at checkout. We use single-server queues; each checkout desk has a separate queue. Customers join the express line if they have less than or the set limit of items. Then the customer joins the shortest queue of the express or regular queues respectively. Customers do not jockey (change line once they have picked one).

4　Results

In this section we first show the results based on a comparison of best- and worst case scenarios, after that we will show the results of the simulations used.

4.1　The gain

The squared coefficient of variation of service time decreases for both regular and express customers; this is because the range of possible service times decreases.
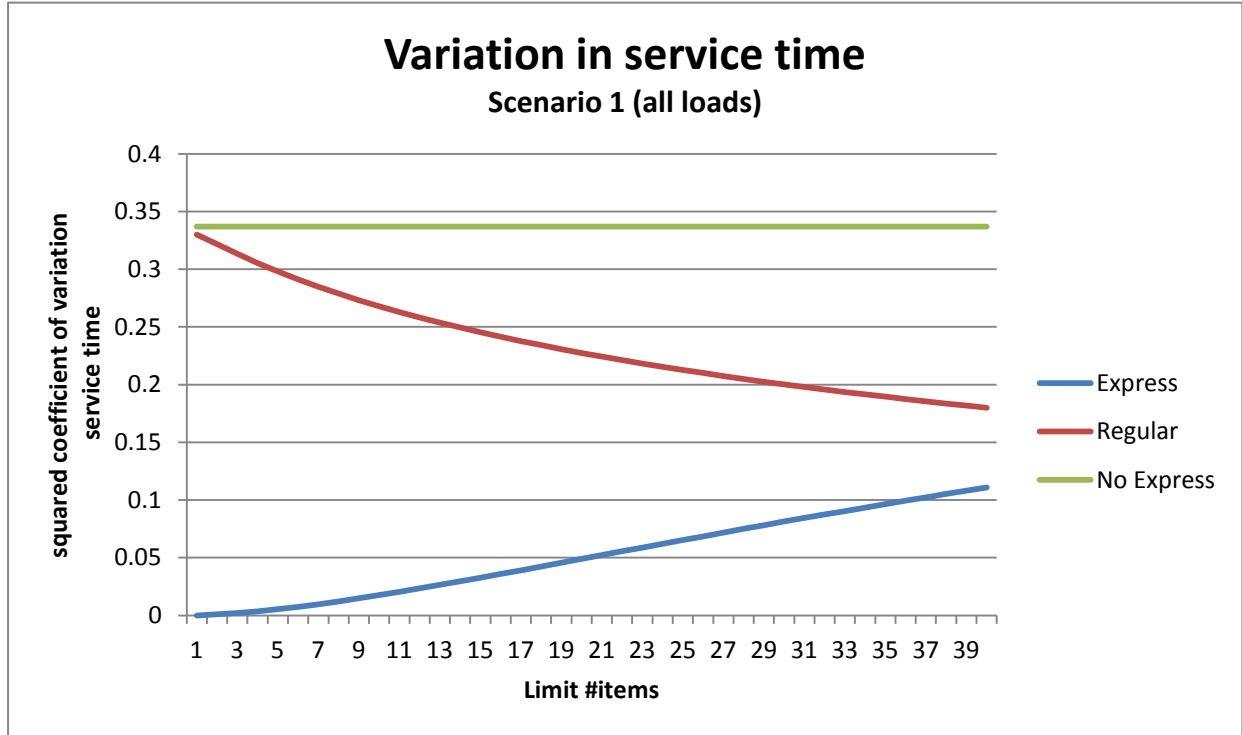

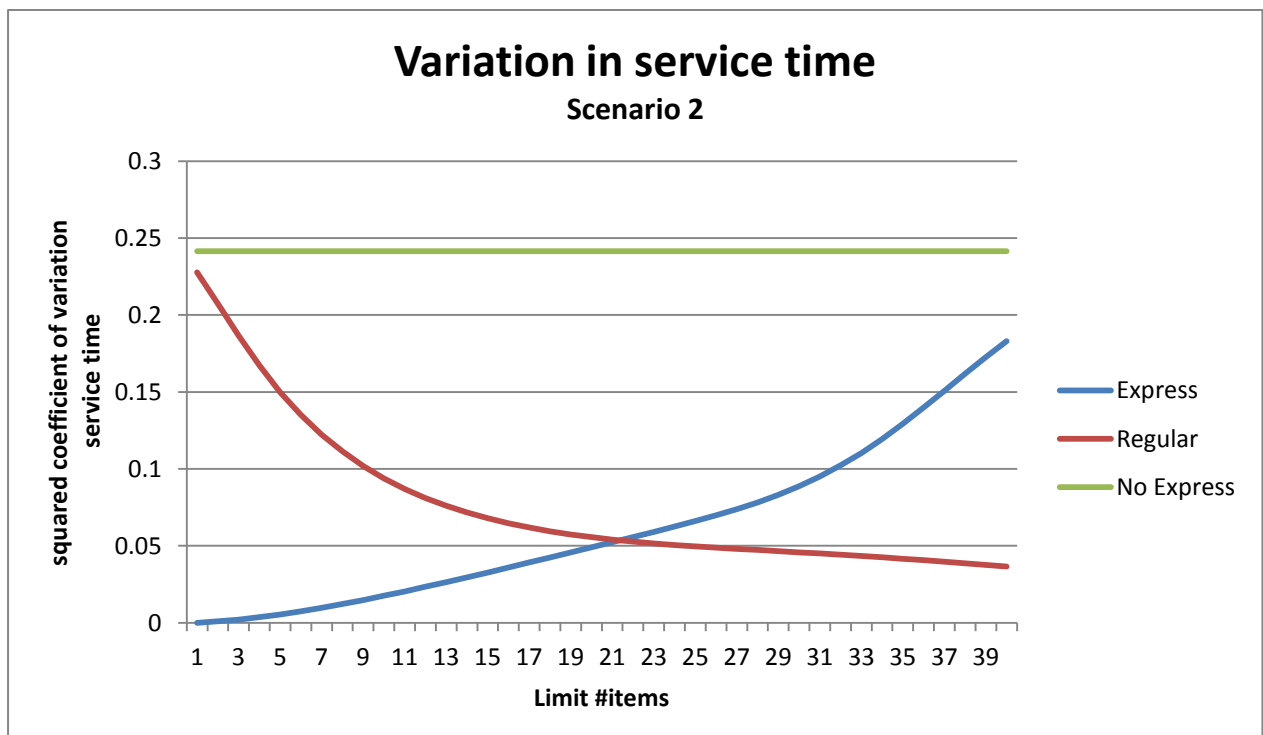Figure 6: squared coefficient of variation in service time (Scenario 1)

As can been seen more easily in the combined lognormal case, the squared coefficient of variation only depends on the range of items allowed in either the express line or regular line. So it is independent of the load on the system, so all sub scenarios of 1 have the same squared coefficient of variation. What can be seen is that with a smaller range of items the variance in service time of express customers is lower in our case, which seems logical there are less possible service time durations, and they are closer together. But for instance server farms can have a large variation in service time with the same amount of items; some files are much larger and take considerably more time to process than other files, while with checkout counters all items take about the same time to process.

4.2   The problem

Next we show the range of possible limits on the amount of items a customer can bring to an express checkout counter with a particular amount of express counters. First we show the ranges for scenario 1, in the order low, medium and high load. With an arrival rate of 200, 300 and 400 customers per hour. This gives respectively a load of 0.449, 0.674 and 0.898 of the system. The space between the red and blue lines is the stability region, See Figure 8, Figure 9 and Figure 10. For computational reasons the range of valid items is not calculated further than a 90 items limit in the express line, as can be seen in the lower load scenarios; in the lower load scenarios the limit can be infinite because with one or two counters less it is still could be possible to serve all customers.
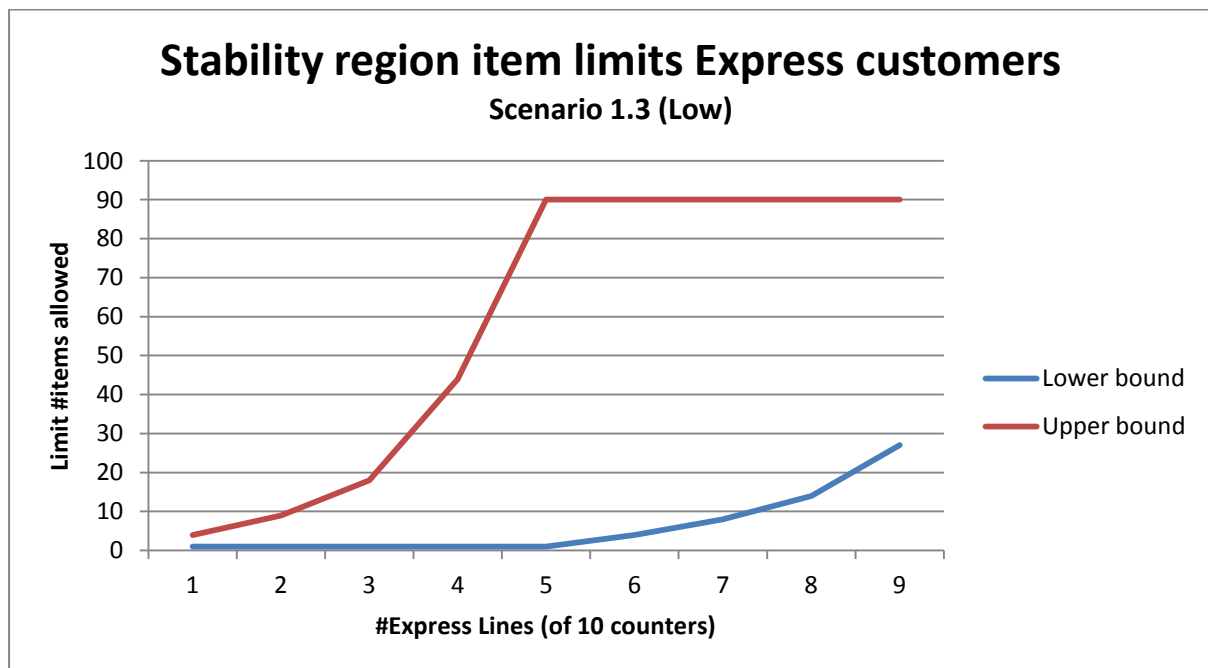

Figure 8: stability region; Scenario 1.3 Low load
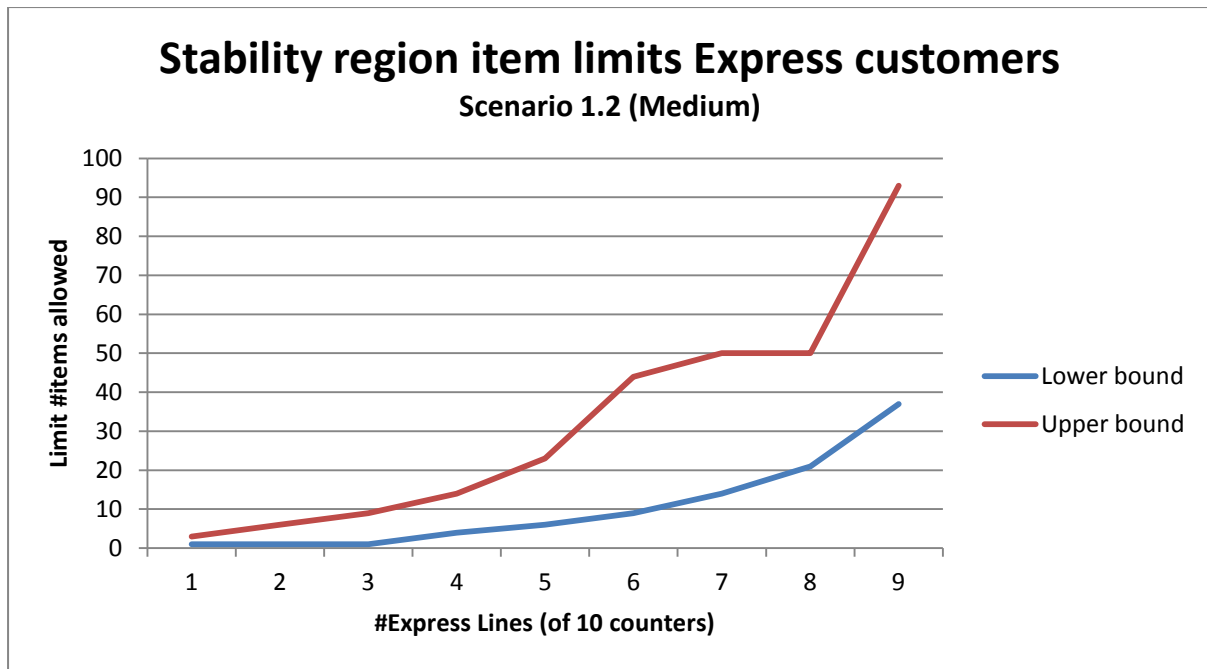
Figure 9: stability region; Scenario 1.2 Medium load
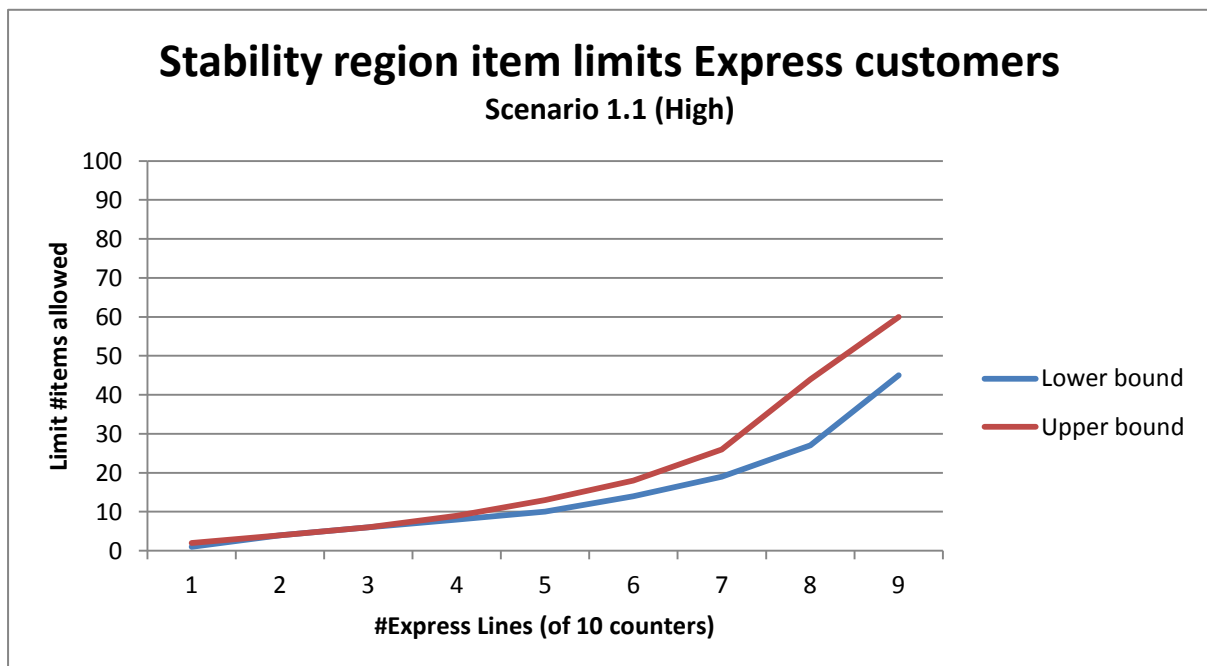
Figure 10: stability region; Scenario 1.1 High load

To show some clear graphs here we use a combined lognormal distribution introduced in Section 3.2.2 and Scenario 2 (Combined Lognormal distribution), which combines a high load with large range of valid item limits.

We made the same figure for Scenario 2, see Figure 11, where the arrow indicates the stability region of allowed item limits for the express checkout counter when 2 express and 8 regular

counters are used.

Stability region item limits Express customers
Scenario 2 (Combined Lognormal distribution)
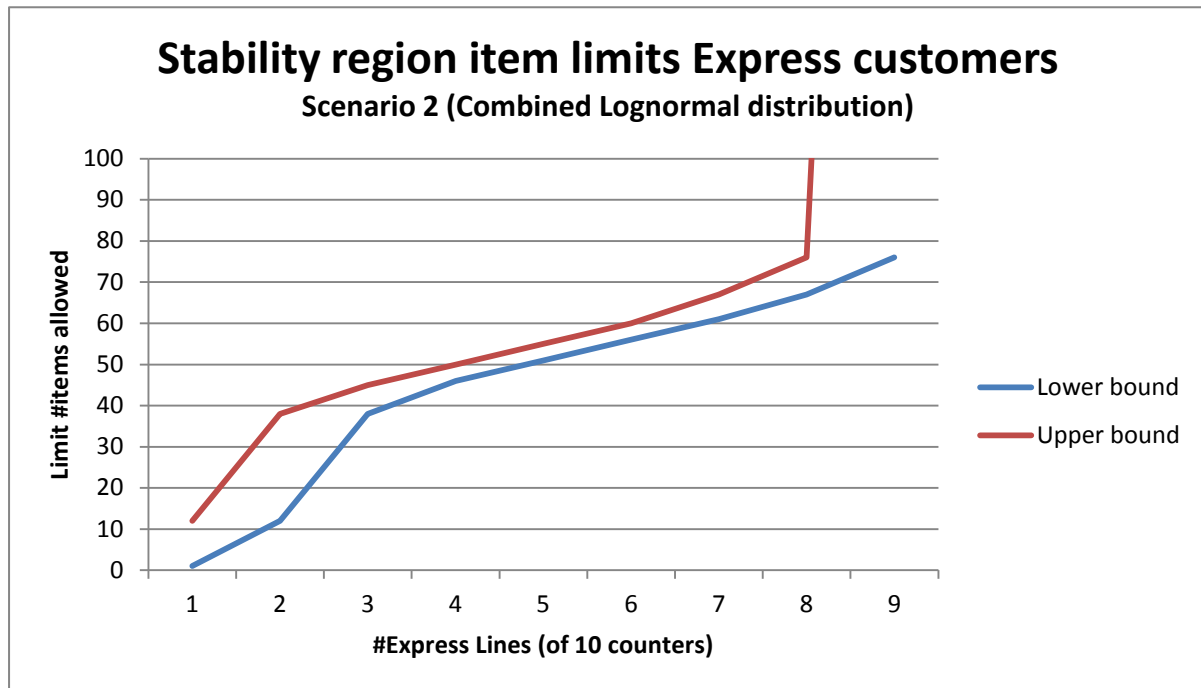
Figure 11: Range of valid item limits; Scenario 2 Combined lognormal distribution

We also made a graph of the gap between best and worst case found limit on the optimal item limit, when the lowest combined expected waiting time is optimized. In the low and medium load figures an arrow is drawn to show this gap between best- and worst case, see Figure 12, Figure 13 and Figure 14

With a higher load on the system the difference between the best and worst case optimal limit on the amount of items converges to the same amount of item limits, as can been seen in Figure 14. Partly because the stability region decreases with increased load.

Figure 12: Gap between best- and worst case; Scenario 1.3 Low load

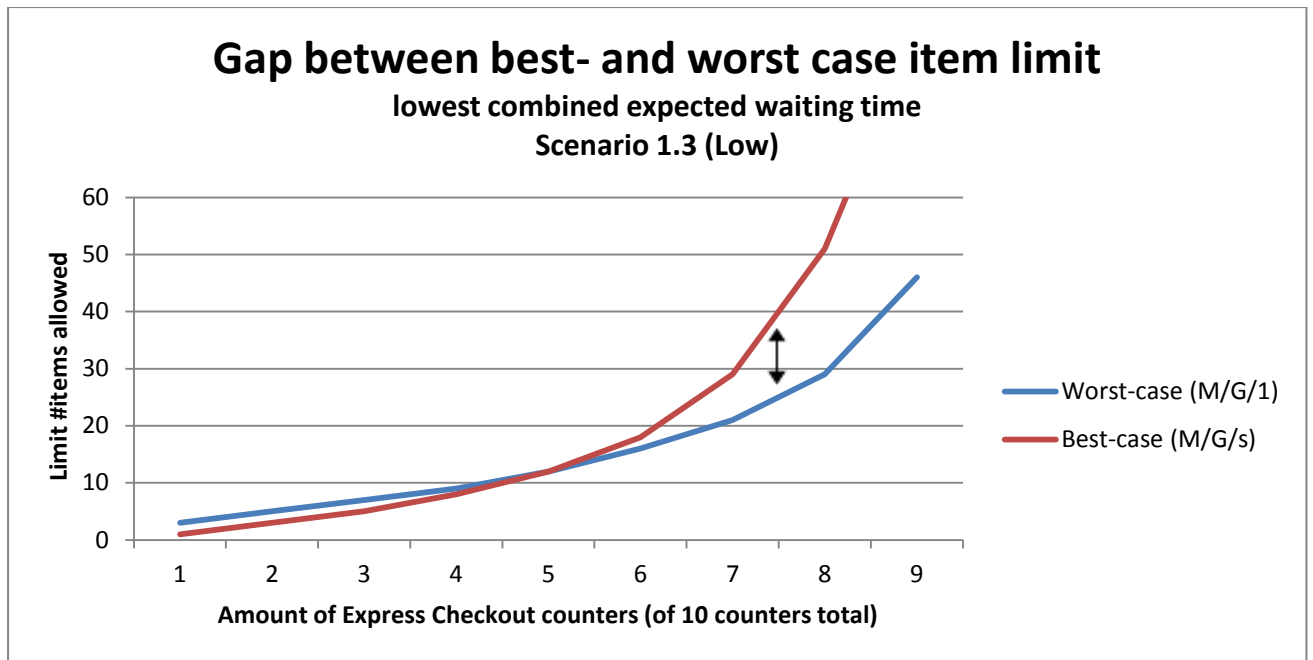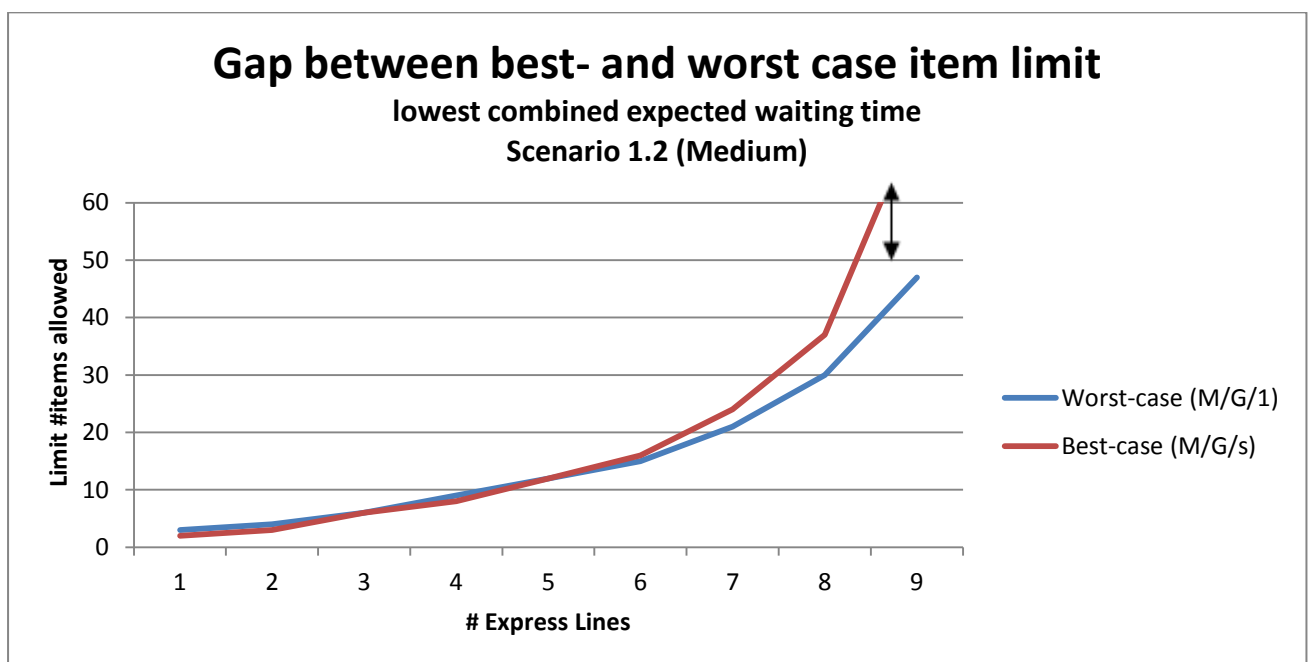**Figure 13: Gap between best- and worst case; Scenario 1.2 Medium load**

Figure 14: Gap between best- and worst case; Scenario 1.1 High load

Using a scenario with 10 checkout counters, 5 arrivals per minute, a lognormal item distribution with a mean of 12 items, based on that the expected service time is 1.3473 minutes, composed of 12 times a service time per item of 0.05 minute and a payment time of 0.75 minute.

Scenario 1.1 (high load) has a load on the system of 0.898, when we use this load only the only two limits on the amount of items allowed in express checkout are 3 and 4 items; all other limits create an unstable system when we use 2 express checkout lines. When the limit is set at 4 items the express customer expects to wait longer than regular customers so there is only one valid limit left of 3 items see Figure 15.

To be able to show a wider range of limits we use Scenario 1.2 (medium load); with fewer arrivals the load of the system decreases, creating a wider range of possible limits on the amount of items allowed in express checkout. The new overall load of the system is 0.674. This results in a limit range from 1 to 5 items. You can see in Figure 16 that the limit on the amount of items people can bring to express checkout 3, the same as in the high load scenario. When looking at the best case with this load, the express checkout has almost always a higher expected waiting time than when no express lines are used. Although the limit with the minimum combined waiting time is the optimal at higher load, this cannot be found at low loads. Because we are mostly interested in high load situations it is hardly a problem but you should take it into account. The worst case shows a clear lower expected waiting time for the express customers see Figure 17, and compared to the limit of 3 and 4 of Scenario 1.1 (high load) it suggests that maybe 4 is this time a limit that results in lower waiting times for express customers than regular customers.
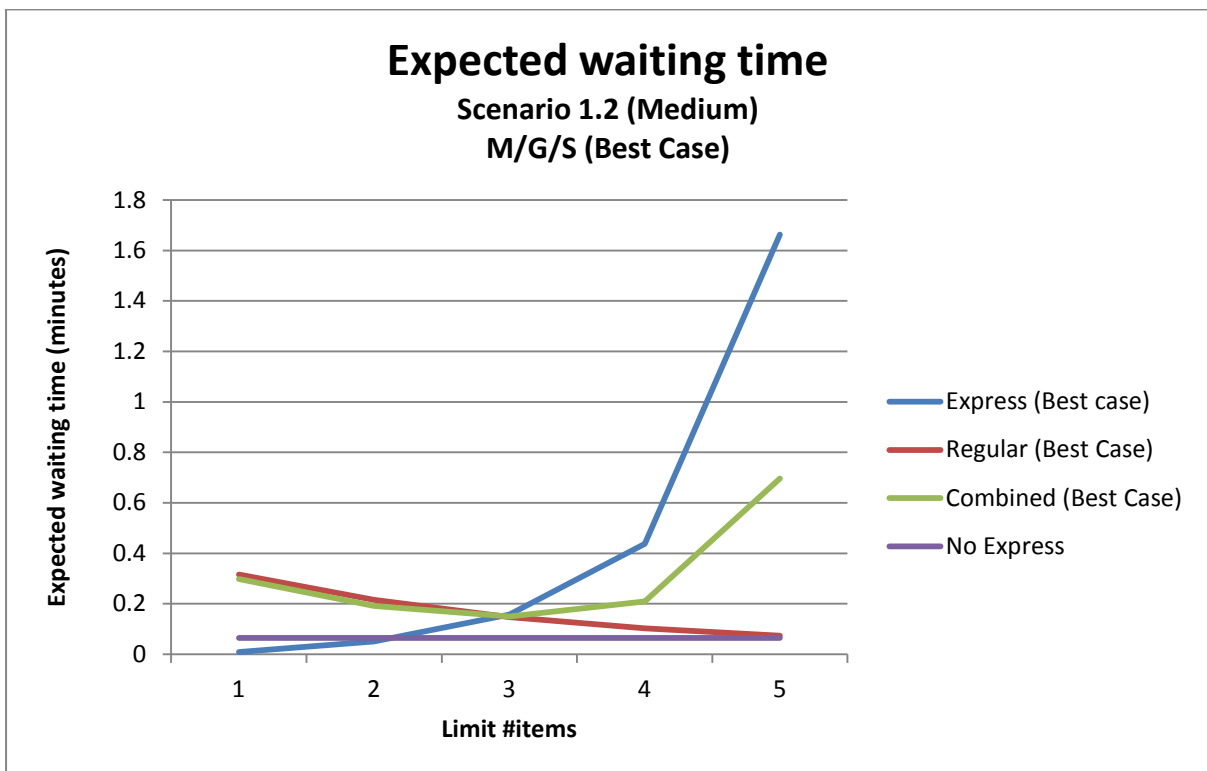
Figure 15: Scenario 1.1 (High) Best case

Figure 16: Scenario 1.2 (Medium) Best case
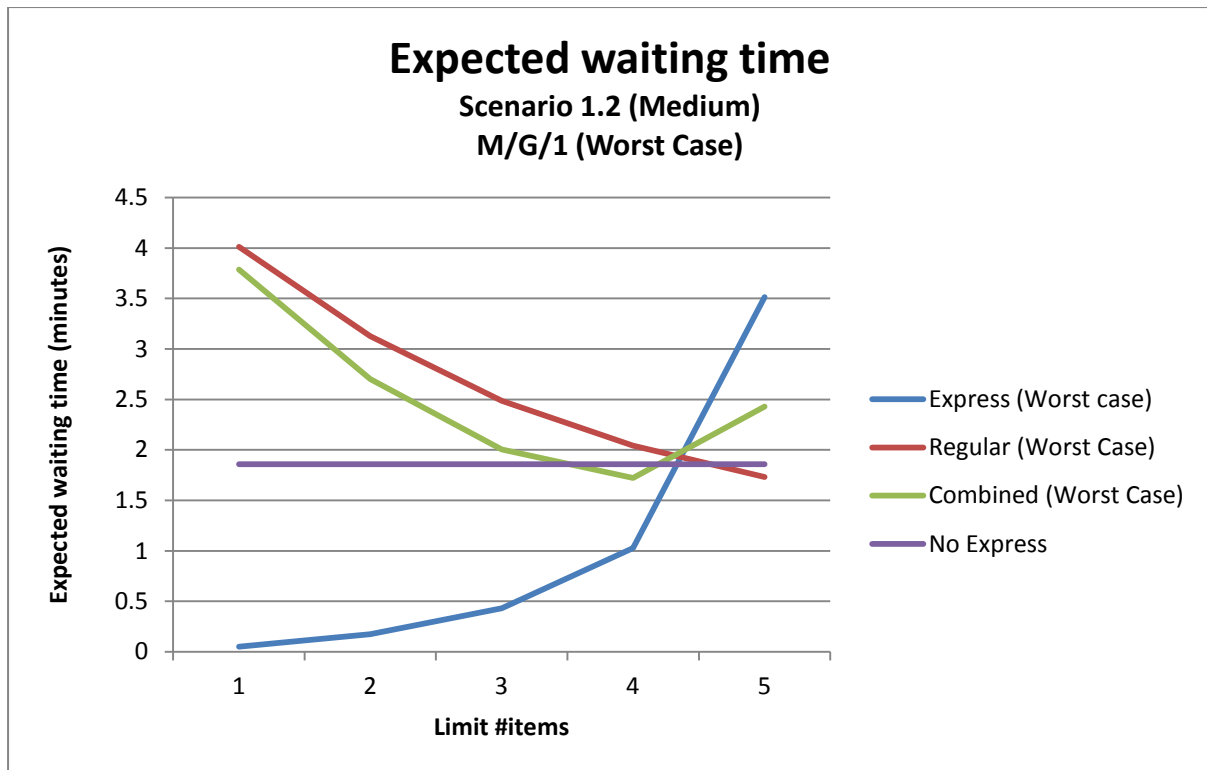
Figure 17: Scenario 1.2 (Medium) Worst case

4.3   Accurate Approximations

To be able to tell if either the best- or worst case approach makes any sense, we decided to create a simulation of the express checkout system.

4.3.1 Validation simulation model.

The simulation is based on M/G/1 queues where customers join the shortest queue (JSQ), whereas the worst case is based on M/G/1 queues with random selection of a queue. When you look at Figure 19 you can see that the simulation projects a lower waiting time than the worst case, when random queue selection is used the expected waiting time is equal to the worst case. In Figure 19 this is shown for Scenario 2 except for the random queue assignment part.

  The best case is already a close approximation to the simulation, in reality customers also take the amount of items customers in front have into account and change lines to an empty counter if they can reach it, possibly even decreasing the gap between simulation and best case.
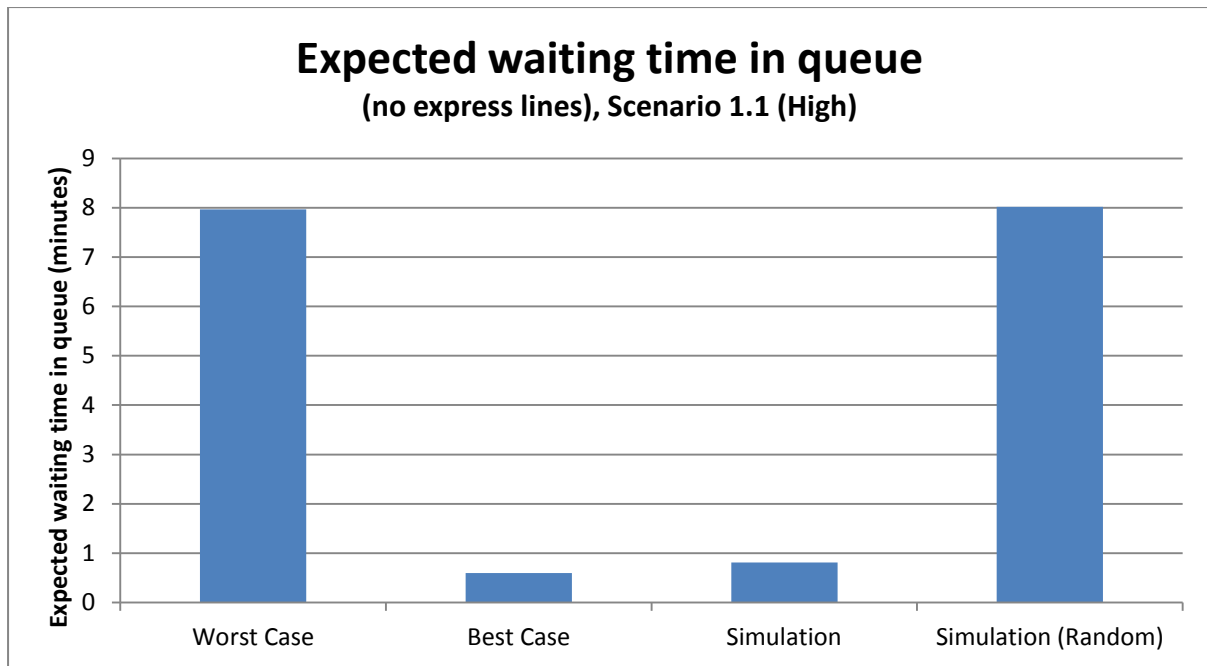
Figure 18: Comparison simulation when no express lines (Scenario 1.1 High load)
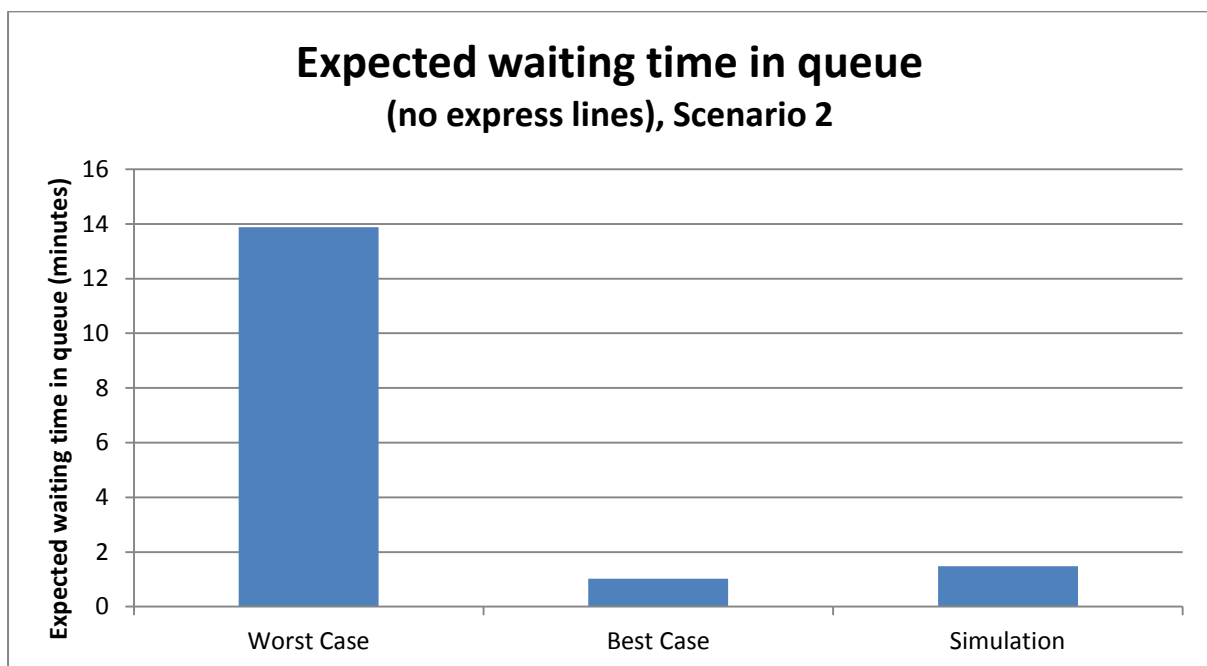
Figure 19: Comparison simulation when no express lines (combined lognormal distribution; Scenario 2)

When we compare the simulated expected waiting times with the best and worst case on the expected combined waiting time for the high load scenario 1.1 you see in Table 2 that the best case approximates the simulation really good. We did this just for the high load situation because this is the most interesting situation.

To be able to create a strong visible graph we compare the simulated expected waiting times with the best and worst case for scenario 2 you see in Figure 20, and Table 3 that the best case approximates the simulation really good, see the italic values for the data of Figure 20. Because of

time constraints on simulation only the simulations from the limit from 26 up to and including 35 are gathered. The same holds for the express and regular lines as can been seen in Table 3.

Limit # items Express	Combined			Express			Regular			No express		
	Worst Case	Best Case	Simul ation	Worst Case	Best Case	Simul ation	Worst Case	Best Case	Simul ation	Worst Case	Best Case	Simul ation
3	16.42	1.92	2.21	0.86	0.36	0.39	21.17	2.39	2.76	7.98	0.60	0.81
4	7.50	1.47	1.64	6.08	2.94	2.94	8.17	0.79	1.03	7.98	0.60	0.81

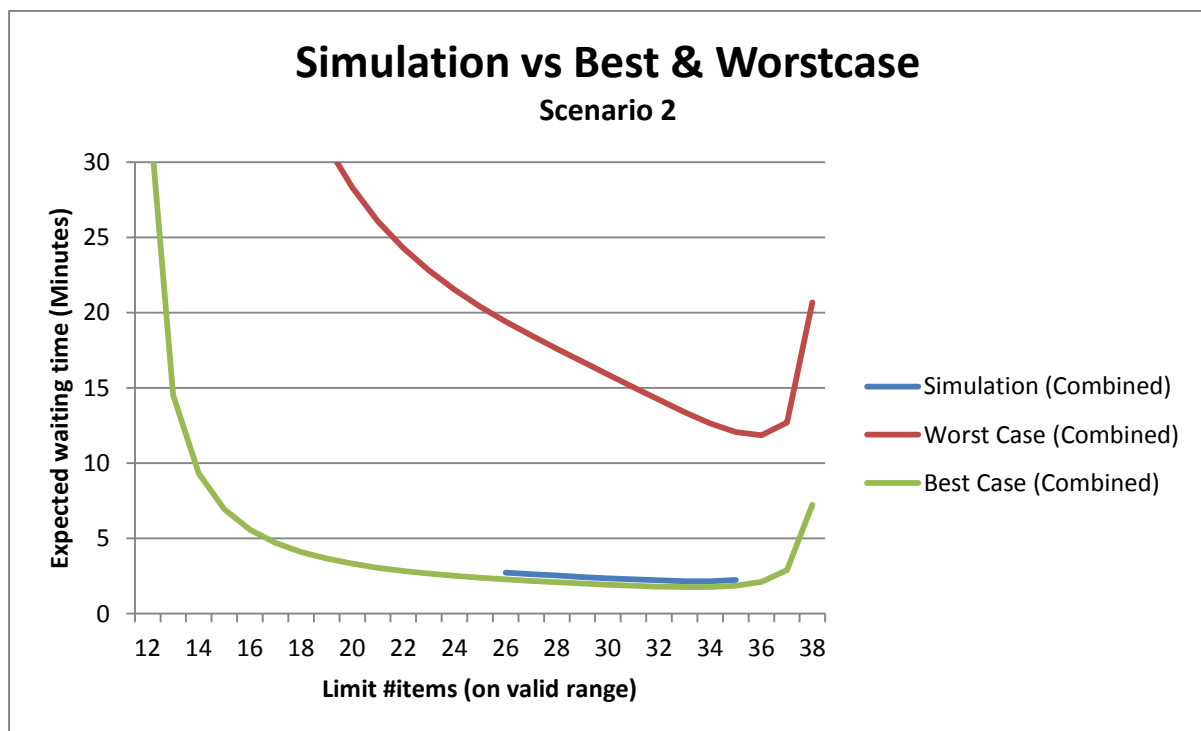**Table 2: Worst Case, Best Case simulation expected waiting times; Scenario 1.1 (high load)**

Figure 20: Worst Case, Best Case simulation expected waiting times; Scenario 2

Limit # items Express	*Combined*			Express			Regular			No express		
	Worst Case	*Best Case*	*Simul ation*	Worst Case	Best Case	Simul ation	Worst Case	Best Case	Simul ation	Worst Case	Best Case	Simul ation
26	*19.41*	*2.28*	*2.74*	1.45	0.62	0.67	29.61	3.23	3.91	13.88	1.02	1.48
27	*18.48*	*2.18*	*2.63*	1.55	0.66	0.72	28.23	3.06	3.73	13.88	1.02	1.48
28	*17.61*	*2.09*	*2.53*	1.66	0.72	0.78	26.92	2.90	3.55	13.88	1.02	1.48
29	*16.75*	*2.01*	*2.44*	1.79	0.78	0.85	25.61	2.73	3.38	13.88	1.02	1.48
30	*15.90*	*1.93*	*2.35*	1.96	0.86	0.93	24.30	2.57	3.21	13.88	1.02	1.48
31	*15.05*	*1.86*	*2.27*	2.17	0.96	1.04	22.95	2.41	3.03	13.88	1.02	1.48
32	*14.21*	*1.80*	*2.21*	2.46	1.10	1.19	21.56	2.24	2.85	13.88	1.02	1.48
33	*13.39*	*1.76*	*2.16*	2.86	1.30	1.40	20.13	2.06	2.66	13.88	1.02	1.48
34	*12.65*	*1.77*	*2.16*	3.46	1.59	1.70	18.68	1.88	2.47	13.88	1.02	1.48
35	*12.06*	*1.85*	*2.25*	4.43	2.07	2.22	17.22	1.70	2.27	13.88	1.02	1.48

In Table 3 you can see the expected waiting time for four groups. The combined expected waiting time is the weighted average of express and regular waiting times. The express and regular waiting times are the expected waiting times of an individual customer waiting in either the express or regular lane. The no express data is the expected waiting time without the use of express counters.

4.4 The trade off

The use of express lines has advantages and disadvantages as can be seen in the following section.

4.4.1 Express counters and the optimal threshold on allowed items

The stability region on the amount of items a customer can bring to the express checkout counter decreases with increasing load, to only a few valid values.

In some cases both express and regular customers have a longer expected waiting time. In the case of Scenario 2 limit 34 and 35 have higher expected waiting times for express and regular customers as can been seen in Table 5. In high load Scenario 1 both express and regular customers expect to wait longer when the limit is set at 4 items as can been seen in Table 4.

In general when reducing the waiting time for express customers for instance with 50%, the regular customers have an increase in expected waiting time. In the case of scenario 1 with a high load an increase in waiting time of 242% is created for regular customers, and an increase of 152% in waiting time in scenario 2 for regular customers, see Table 4. This sounds much but maybe justified if waiting times are compared to service times of the customers. Without express lines customers with a small amount of items may have to wait relative long, and customers with large amount of items have a relatively short waiting time. This may be a factor to consider, see Table 5 and Table 4. In both situations the squared coefficient of variation decreases, the expected service times for both the express and regular customer will have lower variation in both situations, see Figure 6 and Figure 7.

In Table 5 and Table 4 the first part is the expected waiting time of express and regular customers compared to the waiting time they would have had without express lines. The second part of the table shows waiting time divided by their expected service time and in the last two cases the time express or regular customers would have to wait, if there were no express lines. For instance a express customers would always wait longer than their service time if there are no express lines in Scenario 2, see Table 5 the second to last column, compared to the 4th column where express customers wait shorter than their service time.

	Expected change in waiting time compared to no express		Relative waiting time compared to service time				
Limit	Express Customer	Regular Customer	Express/ Express	Regular/ Regular	No Express/ No Express	No Express/ Express	No Express /Regular
3	-52%	242%	45%	185%	60%	94%	54%
4	264%	28%	333%	66%	60%	92%	52%

Table 4: Expected change in waiting time & relative waiting time compared to service time; Scenario 1.1 (high load)

	Expected change in waiting time compared to no express		Relative waiting time compared to service time				
Limit	Express Customer	Regular Customer	Express /Express	Regular /Regular	No Express /No Express	No Express /Express	No Express /Regular
26	-54%	165%	58%	110%	55%	127%	42%
27	-51%	152%	62%	105%	55%	126%	42%
28	-47%	141%	66%	100%	55%	125%	42%
29	-43%	129%	71%	95%	55%	124%	41%
30	-37%	117%	78%	90%	55%	123%	41%
31	-30%	105%	86%	85%	55%	122%	41%
32	-20%	93%	97%	79%	55%	121%	41%
33	-6%	80%	113%	74%	55%	119%	41%
34	15%	67%	135%	68%	55%	117%	41%
35	51%	53%	174%	63%	55%	115%	41%

Table 5: Expected change in waiting time & relative waiting time compared to service time; Scenario 2

The simulation estimates greater reduction in waiting time of express customers and a relative lower increase in expected waiting time of regular customers than the best case.

4.4.2 Possibility to achieve similar service level for all customers

As can been seen in the previous paragraph using express checkout counters can considerably reduce the service level of regular customers when express customers have an improved service level.

People associate long queues with long waiting times, although this does not need to be true as with express lines, so the amount of customers in front is psychologically important for customers, but using express lines increases in these cases the probability of seeing 3 people in front for both regular and express customers. See Table 6. For the express customer the expected waiting time decreases, but there is a larger chance of more people in front than without express lines, although in the express line there are customers with a small number of groceries. For regular customers there is a larger change of more people in front, with on average longer processing times than when no express lines are used.

	Express Customer		Regular Customer		No Express lines	
Limit	P(3 customers in front)	P(4 customers in front)	P(3 customers in front)	P(4 customers in front)	P(3 customers in front)	P(4 customers in front)
3	3.1%	0.7%	29.5%	17.5%	2.9%	0.6%
4	52.5%	39.6%	4.2%	1.1%	2.9%	0.6%

Table 6: Chance of seeing 3 or 4 people in front in the queue; Scenario 1 (high load)

5   Conclusion and discussion

Thoughts about queues and waiting times at supermarkets are important, so this research considers some intuitively understandable best case and worst case model, and a simulation model. The best case waiting time model performs fairly well, and is more intuitively understandable than the simulation model. Although when low load situations are encountered the use of the best case M/G/S waiting time approximation is overly positive. In general high load situations are considered because low load means idle staff. To establish the optimum value for the maximum amount of items customers can bring to the express checkout on high load all three models can be used best case, worst case and simulation perform well. As with low load the simulation model is again the best model to be used.

In supermarkets express checkout counters can have a positive impact on the expected waiting time of express customers, but can strongly increase the expected waiting time of regular customers. When express customers wait on average 50 percent shorter, regular customers can have an increase in expected waiting time of more than 100 percent.

Although this creates fairness with respect to the amount of items a customer buys it creates a longer combined expected waiting time. And both express and regular customers have a higher chance of seeing larger queues, although the express customers have to wait for a shorter time than without express lines.

A better solution to reduce perceived waiting times could be, self scan counters or hand terminals in store. This possibly decreases the passive waiting time at checkout, in contrary to the use of express checkout lines where regular customers have to wait longer than without express lines. So the optimal solution could be a combination of express checkout and some form of unmanned checkout. A combination of both possibly only slightly increases or may even decrease the need for staff to create the same perceived waiting times and so increasing the service level. Because when only express checkout counters are implemented the service level of express customers can increase, the service for regular customers can deteriorate strongly. But more research is needed to possible combinations of express lines and self checkout.

Bibliography

GfK Foodservice, GFK Supermarktkengetallen Oktober 2012

Hornik J, *Subjective vs. Objective Time Measures: A Note on the Perception of Time in Consumer Behavior*, Journal of Consumer Research (11 June 1984).

Jorge A. Alvarado, Luis M. Pulido, *Simulation and experimental design applied to sizing supermarket cashiers in Colombia*, 2008

Kalló N, Koltai T, A *review of management issues related to express line systems*, Social and Management Sciences 16/1 (2008) 21–32.

Maister D.H., *The Psychology of Waiting Lines*, The Service Encounter (Cziepel J A, Solomon M R, SURPRENANT C F, eds.), Lexington Books, Lexington, 1985, pp. 113-123.

Rothkopf H. M, Reich P, *Perspectives on Queues: Combining Queues is not Always Beneficial* , Operations Research 35 (1987), no. 6, 906-909.

Sakasegawa, H., *An approximation formula L q ≃ α· ρ ᵝ/(1-ρ)*, Ann. Inst. Statist. Math. 29 , 67-75 (1977)

Stevens S.S., *On the Psychological Law*, The Psychological Review 64 (1957), no. 3, 153-181.

Whitt W., *Partitioning customers into service groups*. Management Science 45, 1579–1592 (1999)

Whitt W., *Approximations for the GI/G/m queue*. Production and Oper. Management 2, 114–161. (1993)