The Impact of Checkout Congestion on Purchasing Behavior

Takanobu Nakahara
Senshu University
Yutaka Sakuma
National Defense Academy
Katsutoshi Yada
Kansai University
Michel Wedel
The University of Maryland

Acknowledgements:
This work was supported by JSPS KAKENHI Grant Number JP22243033.

Cite as:
Nakahara Takanobu, Sakuma Yutaka, Yada Katsutoshi, Wedel Michel (2021), The Impact of Checkout Congestion on Purchasing Behavior. *Proceedings of the European Marketing Academy*, 50th, (93801)

Paper from the 50th Annual EMAC Conference, Madrid, May 25-28, 2021

The Impact of Checkout Congestion on Purchasing Behavior

In this study, we analyze how customers' in-store purchasing behavior is affected by checkout congestion. Customers' in-store purchasing behavior varies depending on the time they can spend shopping.

Although previous studies have shown that checkout congestion reduces customer satisfaction, our hypotheses are that if the checkout area is congested, customers will spend less time shopping and reduce the total amount of money spent.

We use customer purchase history data and a hidden Markov model to define checkout congestion. Then, we show how purchasing behavior differs between crowded and uncrowded cash registers. The results provide support for our two hypotheses, namely, checkout congestion reduces shopping time and the total amount of money spend.

Keywords: *Hidden Markov model, Checkout congestion, Consumer behavior.*
Identify the track *Methods, Modelling & Marketing Analytics*

1 Introduction

Customers' in-store purchasing behavior depends on various factors, such as time pressure and congestion. Hui et al. matched RFID (radio frequency identifier) data with customer purchase history data and measured purchases in relation to time pressure and store congestion. The results showed that when customers entered the store and had plenty of time to shop, they walked all over the sales floor and purchased numerous items, but at the end of the shopping session, they changed to purposeful behavior in selecting items to purchase. It was also confirmed that while consumers would visit a crowded sales floor, they would not purchase items there [1].

Time pressure also affects price sensitivity, which is reduced in situations where shopping time is limited and achieving the shopping objective is more important than price. Therefore, customers will choose a product even if the price is slightly higher in this situation. In addition, a study on in-store shopping time found that when time spent in the store increased, consumers made more unplanned purchases, thereby increasing the total amount spent [2, 3].

Increased waiting times at the checkout registers were found to reduce customer satisfaction, with 60% of customers feeling stressed if they had to wait for more than three minutes at the checkout register in Japan [4]. Furthermore, checkout waiting times were reported to be an important factor in store selection [5]. Therefore, reducing waiting times at checkouts is an important means of achieving high levels of customer satisfaction in the retail industry.

In this study, we use a hidden Markov model (HMM) to quantitatively estimate checkout congestion using customer purchase history data (ID-POS). Using shopping path data obtained from RFID, we confirm congestion states estimated by the HMM and demonstrate the method's effectiveness. We then analyze the impact of checkout congestion on consumers' purchasing behavior to determine the relationship between the total amount spent and shopping time.

Hui et al. identified congestion on the sales floor using RFID and investigated the purchasing behavior of customers [1]. Our proposed method can predict congestion at the cash register using only ID-POS, which will enable stores to improve their operations and identify customers' purchasing behavior during periods of congestion without incurring additional investment costs.

2 Hidden Markov model

In this study, we develop a model to estimate checkout congestion using an HMM to learn about hidden states from observed variables.

The data x_t observed at time t are generated by hidden states $z_t \in \{1, 2, \cdots, K\}$ based on the distribution $p(x_t|z_t; \boldsymbol{\phi})$ where $\boldsymbol{\phi}$ is the parameter vector of the generative model which is assumed to be constant and independent of time t.

In this study, we use purchase quantities for the observational data x_t, and these follow a Poisson distribution over time.

The hidden state z_t transitions depend only on the previous state z_{t-1}, and the probability distribution is represented by $p(z_t|z_{t-1};\mathbf{A})$, where $\mathbf{A} = \{a_{i,j}|i,j = 1,2,\cdots,K\}$ is a transition matrix that is assumed to be constant and independent of t, and $\sum_j a_{i,j} = 1.0$. The initial state z_1 probabilities are contained in the vector $\boldsymbol{\pi}$. The model parameters are $\boldsymbol{\theta} = (\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\phi})$.

Thus, the simultaneous probability of the observed sequence $X = \{x_1, x_2, \cdots, x_T\}$ and the hidden state sequence $Z = \{z_1, z_2, \cdots, z_T\}$ are given by Equation 1[6].

$$p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) = p(z_1|\boldsymbol{\pi}) \left[\prod_{i=2}^{T} p(z_i|z_{i-1};\mathbf{A}) \right] \prod_{j=1}^{T} p(x_j|z_j;\boldsymbol{\phi}). \tag{1}$$

The EM (expectation–maximization) algorithm is used to maximize the following likelihood function with $P(\boldsymbol{Z}|\boldsymbol{X};\boldsymbol{\theta})$ as the posterior probability of $\boldsymbol{Z}$.

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}') = \sum\nolimits_{\boldsymbol{Z}} P(\boldsymbol{Z}|\boldsymbol{X};\boldsymbol{\theta}') \log P(\boldsymbol{X}, \boldsymbol{Z};\boldsymbol{\theta}). \tag{2}$$

where $\boldsymbol{\theta}'$ is a tentative parameter vector. In the EM algorithm, the maximum likelihood estimate is calculated by repeating the steps used in calculating the Q function given the provisional parameter $\boldsymbol{\theta}'$ (E step), finding the $\boldsymbol{\theta}$ that maximizes it, and updating it as a new parameter (M step).

In this study, the observed data sequence $\mathbf{X}$ corresponds to the number of product scans per unit of time at all cash registers, and four states ($K = 4$) were used to describe the degree of checkout congestion.We determined the number of hidden states is four based on the BIC (Bayesian information criterion). The table 1 shows the BIC values.

Table 1: Number of hidden states and BIC values

No. of states	BIC
2	5889.431
3	5197.666
4	**3932.766**
5	3996.349

The data generation model assumes a Poisson distribution which is a non-negative integer with the average number of product scans λ_{z_i} as a parameter:

$$p(x_j|z_j;\boldsymbol{\phi}) = \frac{\lambda_{z_i}^{x} \mathrm{e}^{-\lambda_{z_i}}}{x!}. \tag{3}$$

3 Estimating checkout congestion using an HMM

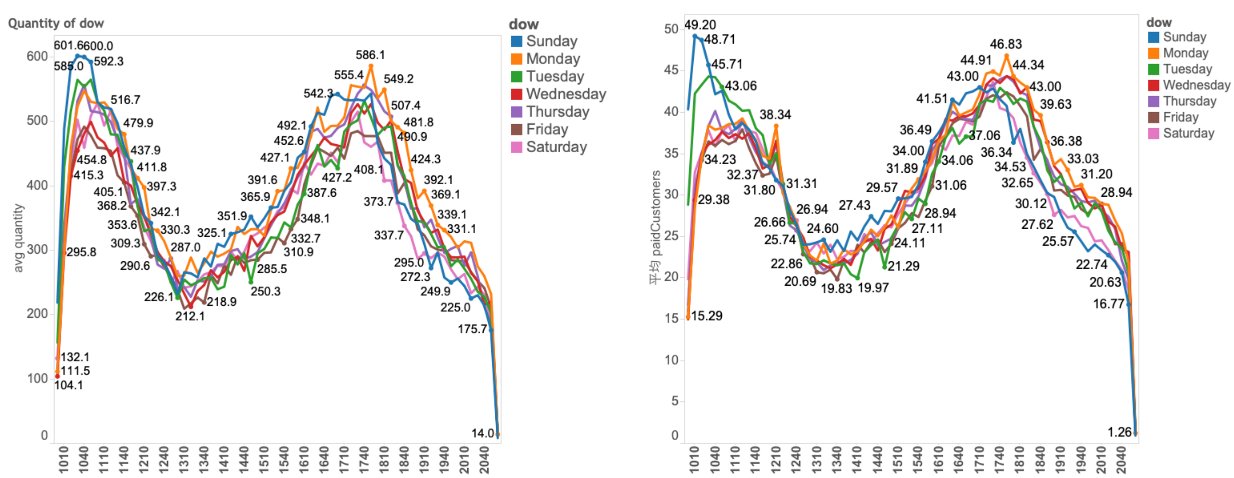
The data used in this study were obtained from a supermarket chain in Japan. Both POS and ID-POS data were collected for the eight-month period from April to November 2012.

Shopping path data were obtained on 30 and 31 August, 1–5 and 27–30 September, and 2–10 October 2012 using RFID tags attached to shopping carts. In total, shopping path data were obtained for approximately 8,000 customers. Since RFID tags have to be attached to the shopping carts, shopping path data were obtained only for customers who used a shopping cart, while ID-POS data were available for all customers. When we used shopping path data, the ID-POS data and shopping path data were merged based on the existence of common IDs.

The observation data sequence $\mathbf{X}$ used to estimate the HMM comprises the total number of product scans at all cash registers during each 10-minute period. The maximum number of cash registers in operation at any given time was eight. For parameter estimation, we used data from June to July 2012 as training data. In addition, data from August to November 2012 were used as test data to predict the hidden state sequence.

3.1 Basic analysis of the number of product scans

Figure 1 (a) shows the number of product scans that were used as the observed data series for the HMM. The vertical axis shows the average number of scans, which is the average of all cash registers during each time period. The horizontal axis represents the various time periods. The store is open from 10 am to 9 pm, a total of 66 10-minute periods.



(a) Average number of scans (b) Average number of customers

Figure 1: The average numbers of scans and customers in each 10-minute period on each day of the week

It can be seen that on most days of the week, there are peaks at around 11 am and 5–6 pm, while on Sundays, the first peak occurs a little earlier, at around 10:30 am. There is little difference among the various days of the week, although the number of product scans is higher on Sundays and Mondays, with an average of 379 scans per period on Sundays and 391 scans per period on Mondays.

Figure 1 (b) shows the average number of customers who passed through the checkout area per period. The vertical axis represents the average number of customers, which is the average of all cash registers during each time period. The horizontal axis represents the various time periods.

It can be seen that the peaks in relation to the number of customers occur at roughly the same times as those in relation to the numbers of products scanned (Figure 1 (a)). Therefore, it appears that the number of product scans is influenced more by the number of customers than by the quantity of items purchased by customers.

3.2 Applying the HMM to estimate congestion

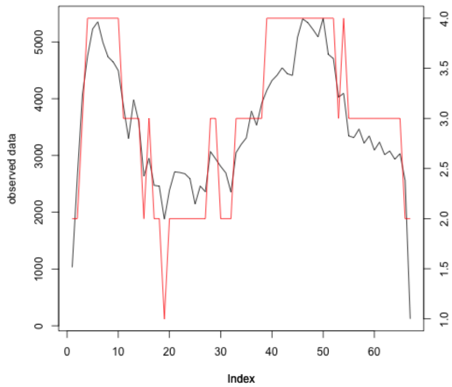
For the observed sequence of training data, we used the total number of product scans during each time period on each day of the week for modeling using the Poisson distribution. The training data were obtained over two months and appeared eight times on Wednesdays, seven times on Thursdays, and nine times on each of the other days. There were fewer Thursdays because of regular holidays.

Figure 2 (a) shows the results of applying the HMM to Monday's training data and estimating the parameters. The horizontal axis shows the time from 10 am to 9 pm, divided into 10-minute periods. The left-hand vertical axis represents the total number of scans at all cash registers during each period. Note that this value is the sum of the same multiple days of the week. The right-hand vertical axis shows the value of the estimated hidden state sequence z_i where 1.0, 2.0, 3.0 and 4.0 represent quiet, regular, crowded, and heavily crowded, respectively. The mean of the observed data corresponding to each hidden state, λ_{z_i}, is 1909.634, 2377.131, 3371.949, and 4834.144, respectively.

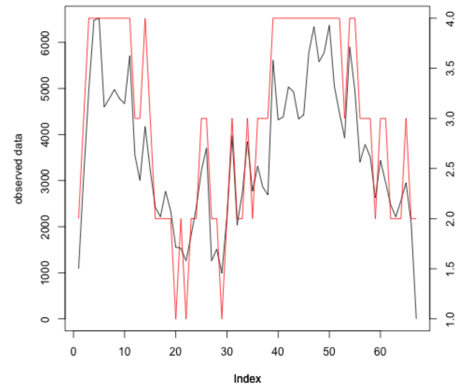
Figure 2 (b) shows the results of applying the model to the test data using the parameters estimated from the training data. The figure shows the results for Monday 27 August. The number of product scans in the observational data sequence was multiplied by nine to match the training data. The estimated hidden state captures the increase or decrease in the observed data sequence. The mean absolute error using the observed values of all test data and the HMM ' s predicted values was 391.27 and the mean absolute percentage error was 38.97%.

3.3 Verification of results using shopping path data

To verify whether the hidden state sequence estimated using the HMM accurately reflected the level of cash register congestion, we used shopping path data to check the level of congestion.



(a) The HMM applied to Monday's training data



(b) Prediction based on test data

Figure 2: Results of applying the HMM to training data and predictions based on test data

Congestion can be verified using shopping path data. However, while most retailers have POS systems in place, they do not collect shopping path data, and acquiring these data would require additional investment.

In this study, only POS data were used to identify congestion, which is an important initiative regarding the level of cash register congestion. Our method can improve store operations, such as increasing the number of operating cash registers in advance during busy periods, because congestion at the cash registers can be predicted. It can also elucidate consumer behavior during periods of congestion.

Shopping path data allow us to observe the amount of time a customer spends at the cash register. POS data, on the other hand, is recorded the time a customer has paid.

Figure 3 shows the average customer waiting time based on the shopping path data. The waiting times at the cash register peaked at roughly three minutes on every day of the week, but waiting times extended to more than six minutes in some cases. The mean for the entire week was 195.7 seconds, with a standard deviation of 81.37 seconds. However, it should be noted that this only included customers who used shopping carts.

Table 2 shows the results of ANOVA testing using waiting times in the four hidden states at the 5% significance level. There are differences among the mean waiting times in the four hidden states ($p < 0.0001$). A comparison of waiting times by day of the week was also significant at the 5% level ($p < 0.0001$). These results show that waiting times at the cash register differed among the four hidden states. Basically, the waiting time tends to increase as the degree of congestion increases. Therefore, the hidden states estimated using only the number of products scanned at the cash register reflect the waiting time.

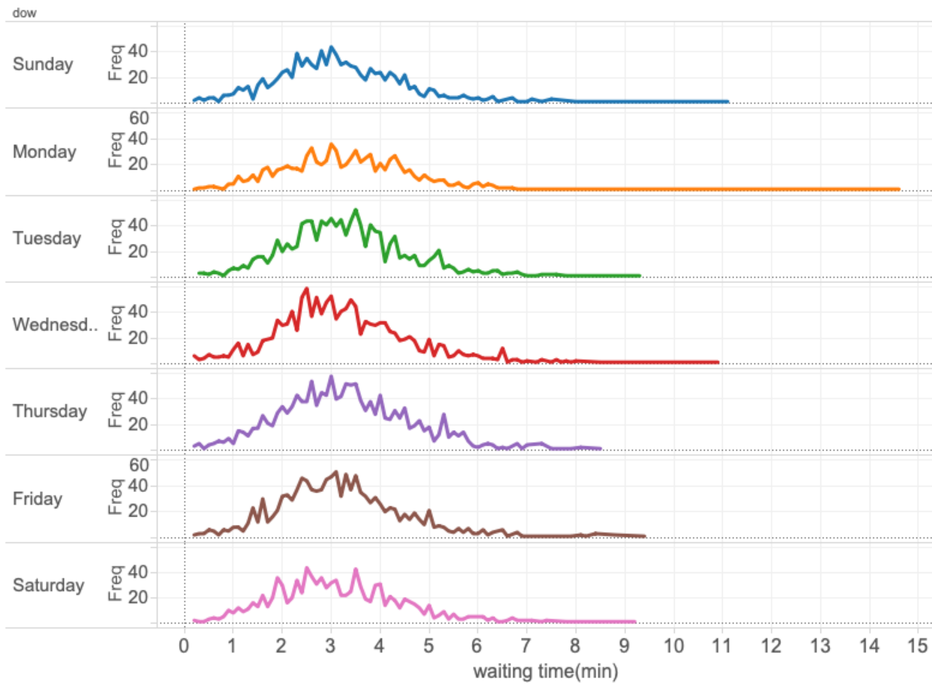


Figure 3: Average waiting time at the cash register

4 Analysis of purchasing behavior during periods of checkout congestion

We used both POS and shopping path data to analyze the impact of checkout congestion on purchasing behavior. A previous study [2] found that increased shopping time promotes unplanned purchases and increased sales. However, the question is, does it have the same effect when there is checkout congestion? To clarify the impact of checkout congestion, we formulated the following hypotheses:

H1: When the cash registers are congested, consumers want to finish their shopping quickly, which reduces their shopping time.

H2: When the cash registers are crowded, both the shopping time and the amount of money spent are reduced.

To test these hypotheses, we compared the differences in purchasing behavior based on the four hidden states at the cash register when customers entered the store. Specifically, we looked at the overall Shopping Time, Travel Time (time when the cart is moving), and Stationary Time (time when the cart is still). We also looked at the amount of money spent and the amount of money spent per product.

Table 3 shows these results. P-values represent the significance probability of the analysis of variance for each indicator after applying a logarithmic transformation.

Table 2: Comparison of average wait times (in seconds) in the four states of congestion.

DOW	Quiet	Regular	Crowded	Heavily Crowded
Overall	169.5180	170.5425	192.7286	212.1775
Sunday	169.5455	179.4872	185.6361	210.0000
Monday	166.3333	159.8477	190.0110	222.1087
Tuesday	199.0000	178.7670	197.9718	218.8665
Thursday	181.5500	168.6050	201.2529	211.3527
Wednesday	173.1644	178.0093	188.0298	224.4675
Friday	116.7000	149.5124	190.5199	202.2845
Saturday	*	160.1931	193.6068	201.8460

Table 3: Purchasing behavior of customers who entered the store during the four states

Average	Quiet	Regular	Crowded	Heavily crowded	P-value
Shopping Time	1370.59	1258.80	1256.88	1205.49	0.002 **
Travel Time	479.35	456.40	455.23	432.21	0.006 **
Stationary Time	723.20	698.14	708.35	693.54	0.668
Quantity of purchased	21.42	21.65	21.14	20.13	0.039 *
Amount of money	3827.11	3732.52	3683.76	3500.78	0.069 .
Amount of per product	182.89	175.22	177.71	176.37	0.415

Shopping time, travel time, and quantity of items purchased were significant at the 5% level. The total amount of money spent was significant at the 10% level. Shopping time tends to be shorter when the store is crowded, as does travel time. However, there was no significant difference in relation to stationary time. This suggests that when congestion occurs, either the shopping route, and thus travel time, becomes shorter, or the travel becomes faster. Therefore, Hypothesis 1 was supported.

The average quantity of products purchased decreased by about one when the store was heavily crowded. This may be related to the reduced shopping time. The total amount of money spent also decreased as conditions became more crowded. Therefore, Hypothesis 2 was supported.

When the cash registers are crowded, customers want to finish their shopping quickly. As a result, shopping time is reduced, as are the total amount of money spent and the number of items purchased.

Hui et al.[1] found that people do not buy anything when they encounter a crowded sales floor, and our findings confirm this loss of sales opportunities as a result of crowding. Therefore, stores should avoid crowding at the cash registers wherever possible by considering the operation of the cash registers.

5 Conclusion

In this study, we used an HMM to define the level of congestion in stores and identify its impact on cash register congestion. In particular, since the congestion status can be estimated from the number of purchases made by customers using POS data, no new investment is necessary to identify the level of congestion, such as acquiring RFID data.

Although previous studies have shown that checkout congestion reduces customer satisfaction, the results of this study showed that checkout congestion leads to a reduction in shopping time, the number of products bought, and the total amount of money spent. Thus, stores should reduce checkout congestion as much as possible by improving checkout operations because reduced shopping time as a result of congestion leads to lost sales opportunities. Our model can be a useful tool to identify and predict congestion based on POS data.

Future research will focus on applying a queueing model with the aim of improving checkout operations.

References

[1] Hui, S.K., Bradlow, E.T., & Fader, T.S. (2009). Testing behavioral hypotheses using an integrated model of grocery store shopping path and purchase behavior. *Journal of Consumer Research*, 36, 478–493.

[2] Spies, K., Hesse, F., & Loesch, K. (1997). Store atmosphere, mood, and purchasing behavior. *International Journal of Research in Marketing*, 14(1), 1–17.

[3] Bell, D.R., Corsten, D., & Knox, G. (2011). From point of purchase to path to purchase: How preshopping factors drive unplanned buying. *Journal of Marketing*, 71, 31–45.

[4] Citizen, (2013). Waiting Time Awareness Survey, (Jan. 3rd 2021 last accessed) `http://www.citizen.co.jp/research/time/20130603/01_02.html`

[5] Popkowski Leszczyc, P. T. L., & Timmermans, H. (2001). Experimental choice analysis of shopping strategies, *Journal of Retailing*, 77(4), 493–509.

[6] Bishop, C.M. (2006). *Pattern recognition and machine learning.* Berlin: Springer-Verlag.