Inferring True Inter-Purchase Times From Incomplete Information: A Model Comparison

Abstract

The "incomplete information problem" describes the fact that most companies only possess their own customers' transaction data. Information about customers' purchases at competitors is usually unavailable. Yet, knowledge about the complete purchase history would enable companies to improve their targeting. The authors apply Chen and Steckel's (2012) parametric approach to model true inter-purchase times and share of wallet on the category-level to transaction data of a European grocery retailer. We propose a model simplification strategy and compare the estimates of the true inter-purchase times across models and product categories. In our validation tasks, both models perform equally well. Yet, our simplified model runs considerably faster. For our analysis, we use almost two years of transaction data from about 1,000 customers and nine product categories. Finally, we outline work-in-progress to apply the proposed model to improve the timing of personalized price promotions.

Keywords: inter-purchase times, share of wallet, hierarchical Bayesian estimation

Track: Methods, Modelling & Marketing Analytics

1 Introduction

The retail environment is characterized by fierce competition as retailers try to maintain and promote their customers' loyalty, often measured as share of wallet (SOW). Customers, in turn, rarely commit themselves to a single retailer. Instead, when shopping for groceries, 75% of customers frequently visit multiple retailers (Ho & Chen, 2005). Retailers, however, only observe purchases made at their own stores. If two customers, A and B, both exhibit equal purchase cycles and spend similar amounts of money on their shopping trips, then they seem to be equally valuable and loyal customers of retailer X. Yet, customer A might also visit the stores of a competitive retailer Y, while customer B only purchases at the focal retailer X. Yet, from their observed purchase characteristics, both customers seem undistinguishable to retailer X. This situation is known as the "incomplete information problem" as retailers lack information about purchases at other stores (Chen & Steckel, 2012). Shedding light on this unknown information enables a more effective targeting based on customers' true and unobserved purchase behavior (Chen & Steckel, 2012). Thus, retailers could use knowledge about true inter-purchase times (IPT)—based on observed and unobserved purchases—to provide customers with more relevant offers compared to one-size-fits-all price promotions.

In this paper, we investigate the "incomplete information problem" at the category level in grocery retailing. We apply the modelling approach proposed by Chen and Steckel (2012) to identify customers' true IPT and SOW from observations at a single company, in our case a retail chain. We analyze almost two years of data from about 1,000 customers and nine categories and benchmark the model by Chen and Steckel (2012) against a simplified version we propose. We find that both models seem to predict observed IPTs equally well. Yet, our model speeds up the estimation process. Strikingly, there are great differences with respect to SOW predictions and predictions of true category-level IPT. To (1) further evaluate the original model by Chen and Steckel (2012) against our simplified model and (2) improve the timing of personalized price promotions based on customers' true category IPT, we outline our plans to apply our results to a recommendation system at a European grocery retailer.

2 Model

Chen and Steckel (2012) use data on customers' credit card usage when buying groceries to model customers' true grocery IPT and the SOW of the focal credit card in the grocery category. To derive true category IPTs and SOWs in the grocery category, Chen and Steckel (2012) introduce two submodels.

2.1 *Submodel 1: Distribution of category-level IPTs*

Individual category-level IPTs are Erlang (2, β_{ij}) distributed (Chen & Steckel, 2012). For customer i and category j, the Erlang distribution with shape parameter 2 and rate parameter β_{ij} is given by:

$$f_{ij}\left(t;\,2,\,\beta_{ij}\right)=\beta_{ij}^{2}te^{-\beta_{ij}t},\quad t,\,\beta_{ij}\ \geq 0. \tag{1}$$

The Erlang-*k* distribution is a special case of the gamma distribution that requires an integer shape parameter *k* and also contains the exponential distribution (*k* = 1). Due to the additive property of the gamma distribution, the sum of m independent Erlang (k, β_{ij}) distributions is Erlang (km, β_{ij}) distributed. While the exponential distribution is memoryless and illustrates random IPTs, the Erlang-*k* distribution with *k* > 1 models regularity in timing patterns (i.e., increasing with *k*) (e.g., Chen & Steckel, 2012). The next purchase depends on the time since the last purchase, which has been empirically validated for grocery purchases on basket- (e.g., Chen & Steckel, 2012) and category-level (e.g., Herniter, 1971).

In addition, Chen and Steckel (2012) assume that customers remain active over the complete observation period. In our case, retail customers could move, decide to no longer visit a particular retailer, or die. Yet, these cases will not be accounted for. Instead, we treat our observations as right censored.

2.2 Submodel 2: Observed purchases at a retailer

Similarly to Chen and Steckel (2012), we only possess data of a single company, in our case, a grocery retailer. Hence, the question arises which purchases the focal retailer observes. The original model proposes a two-state Markov process with the following transition matrix:

$$\begin{pmatrix} \phi_{ij} & 1\text{-}\phi_{ij} \\ \lambda_{ij} & 1\text{-}\lambda_{ij} \end{pmatrix} \tag{2}$$

to describe credit card usage in the grocery category. Thus, the probability of customer i using the focal credit card depends on whether she has used it for the previous grocery purchase in category j (ϕ_{ij}) or not (λ_{ij}). Chen and Steckel (2012) ascribe the necessity of this two-state Markov process to customers trying to balance their usage of several credit cards. In the case of grocery purchases, the two-state Markov process can be motivated by regularity in customers' purchases across retailers (Gijsbrechts, Campo, and Nisol, 2008). Thus, customers systematically visit a set of retailers on a regular basis. The SOW_{ij} can then be described by the steady state probability of the two-state Markov process and provides probability that we observe a purchase of customer i in category j in the long-run:

$$SOW_{ij} = \frac{\lambda_{ij}}{1+\lambda_{ij}-\phi_{ij}}\,. \tag{3}$$

A two-state Markov process is not the only possibility to model customers' behavior to visit multiple stores (Chen & Steckel, 2012). We argue that a simple Bernoulli process might be an alternative to model grocery purchases when customers engage in cherry-picking. Cherry-picking customers visit multiple stores in reaction to price promotions (Fox & Hoch, 2005). This behavior is in line with the results of a market research by The Nielsen Company (2015) that identifies price as a primary driver for store switching. Accordingly, customers have a set of retailers whose price promotions they consider. Yet, the decision where to buy is made anew every time. The following relation holds $\phi_{ij} = \lambda_{ij} = p_{ij}$. We observe a purchase of customer i in category j with probability $p_{ij,}$ which also reflects her SOW_{ij}. The Bernoulli process consequently simplifies the original model by Chen and Steckel (2012).

Which process better describes customers' shopping behavior across multiple stores depends on the composition of the heterogeneous customer base. As we expect a mixture of cherry-pickers and systematic store switchers at the focal retailer, we rely on the empirical analysis to determine which process most appropriately describes the observed purchases.

2.3 Derivation of a probability density function for observed IPTs

Following Chen and Steckel (2012), the combination of both submodels allows for a derivation of a probability density function for observed IPTs. For customer i, the distribution of observed IPTs t in category j can be described by:

$$g_{ij}(t|\phi_{ij}) = \sum_{m=0}^{\infty} Q_{ij}(m)\ \text{Erlang}[\,t|\ 2(m+1),\ \beta_{ij}]. \tag{4}$$

The term Erlang[t|2(m+1), β_{ij}] applies the additivity property of the Erlang-*k* distribution and describes the distribution of the observed IPTs given that m unobserved purchases lie between the observed purchases. The term $Q_{ij}(m)$ gives the probability that m unobserved purchases lie between two observed purchases of customer i in category j. In case of a two-state Markov process (Chen & Steckel, 2012), we write:

$$Q_{ij}(m)=\begin{cases}\phi_{ij} & , \ m=0\\ (1-\phi_{ij})(1-\lambda_{ij})^{m-1}\lambda_{ij}, & m>0.\end{cases} \tag{5}$$

When applying our proposed model simplification and replacing the two-state Markov process by a Bernoulli process, we set $\phi_{ij} = \lambda_{ij} = p_{ij}$.

The expected value of an Erlang (2, β_{ij}) distribution is given by $2/\beta_{ij}$. It describes a customer's expected category-level IPT across all purchases—observed and unobserved—and provides us with a solution to the "incomplete information problem". To receive the expected observed IPT of customer i in category j, we multiply her SOW_{ij} with the expected value of the Erlang (2, β_{ij}) distribution (Chen & Steckel, 2012):

$$E(\text{observed IPT})=\frac{2}{\beta_{ij}}\;\frac{1+\lambda_{ij}-\phi_{ij}}{\lambda_{ij}}. \tag{6}$$

When applying our model simplification strategy, we set $\phi_{ij} = \lambda_{ij} = p_{ij}$.

2.4 Specification of the Hierarchical Bayesian Model

In the original paper by Chen and Steckel (2012), the model requires the estimation of three parameters on an individual level β_{ij}, ϕ_{ij}, and λ_{ij}. In contrast, our proposed simplification requires only two parameters, β_{ij} and p_{ij}. As sufficient data per customer is needed for the individual-level estimation, Chen and Steckel (2012) introduce a hierarchical Bayesian approach. Thus, the parameters to be estimated are modelled as functions of customer-level information. To ensure that β_{ij} is positive and the parameters of the two-state Markov (ϕ_{ij} and λ_{ij}) and Bernoulli process (p_{ij}) are bounded by 0 and 1, we rewrite them as:

$$\beta_{ij}=\exp(\theta_{\beta ij}) \tag{7}$$

$$p_{ij}=\frac{\exp(\theta_{pij})}{1+\exp(\theta_{pij})} \quad \text{(equivalent for } \phi_i\text{, and } \lambda_i\text{)}. \tag{8}$$

$\Theta_{ij} = (\theta_{\beta ij}, \theta_{\phi ij}, \theta_{\lambda ij})'$ and $\Theta_{ij} = (\theta_{\beta ij}, \theta_{pij})'$ contain the transformed parameters of the original model (Chen & Steckel, 2012) and our proposed model simplification. We model Θ_{ij} as a function of customer- and category-level information:

$$\begin{aligned}\theta_{\beta ij} = {} & \gamma_0+\gamma_1(\text{category penetration}) + \gamma_2(\text{category coefficient of price variation}) + \\ & \gamma_3(\text{avg. quantity per category per purchase}) + \gamma_4(\text{share of brands per category}) + \\ & \gamma_5(\text{avg. basket value per user}) + \gamma_6(\text{avg. category purchase frequency per user}) + \\ & \gamma_7(\text{avg. number of categories per basket per user}) + \varepsilon_{\beta ij} \\ & (\text{equivalent for } \theta_{pij}, \theta_{\phi ij}, \text{and } \theta_{\lambda ij}).\end{aligned} \tag{9}$$

When estimating the model (Chen & Steckel, 2012), we set up the following likelihood function for observed IPTs t_{ijr}, where i (to n) denotes the index for the customer, j (to J) denotes the index for the product category, and r (to R_{ij}) denotes the r^{th} observed IPT:

$$L=\prod_{i=1}^{n}\prod_{j=1}^{J}\left\{\left[\prod_{r=1}^{R_{ij}} g_{ij}(t_{ijr}|\Theta_{ij})\right]\left[1-\int_0^{t_h} g_{ij}(t|\Theta_{ij})dt\right]\right\}. \tag{10}$$

The second part of the likelihood function accounts for right censoring and describes the probability of observing no purchase between the last observed purchase and the end of the observation period (t_h) (Chen & Steckel, 2012).

3 Model Application

Our current project is not limited to the simplification of the model introduced by Chen and Steckel (2012). Instead, we want to use the insights gained from modelling the true IPTs to improve the timing of personalized price promotions. At our focal retailer, loyalty program members receive personalized price promotions. These promotions result from a dynamic scoring model that considers customers' previous purchases but no timing patterns.

However, the timing of a promotion is crucial for its redemption. Previous research has shown that including the timing of customers' observed category-level purchases, can help to improve the personalization of price promotions (e.g., Vuckovac, Wamsler, Ilic, and Natter, 2016). We plan to go one step further and use knowledge about (unobserved) true category IPTs in order to improve the timing and thus, the redemption of personalized prize promotions. This might prove to be a competitive advantage for retailers and enable them to trigger purchases that a customer would have made at a competitor.

As the application discussed in this section outlines work-in-progress, we only present the model comparison of the original model by Chen and Steckel (2012) to our proposed model simplification[1].

4 Empirical Analysis

4.1 Results

We received the transaction data for the empirical analysis from more than 100 stores of a brick-and-mortar grocery retailer that are located in and around a major European city. We analyze 22.5 months of data from almost 1,000 customers and nine different product categories. All customers in our dataset take part in the retailer's loyalty program and identify themselves at the check-out via their loyalty card. However, the retailer does not collect any personal information (e.g., age, gender, address) of the loyalty card holders. Only transaction data is provided.

The data is split into an estimation (15 months) and a holdout period (7.5 months). The estimation data is then used to estimate the expected true and observed individual-level IPTs on category-level as well as the respective SOWs. The aggregated results for the original three-parameter model by Chen and Steckel (2012) and our simplified two-parameter model are presented in Table 1.

Table 1: Predictions of the hierarchical three- and hierarchical two-parameter model

Category	Average observed IPT	Two-parameter model		Three-parameter model	
		SOW	E(true IPT)	SOW	E(true IPT)
Capsule coffee	28.2	28.9%	6.1	37.5%	8.9
Cola	23.3	28.2%	4.6	37.6%	6.7
Detergent	38.4	27.6%	8.1	36.9%	13.3
Dish liquid	39.7	27.6%	8.6	37.1%	14.5
Fresh milk	22.6	31.6%	5.2	38.7%	7.2
Lemonade	22.8	25.6%	4.2	38.6%	6.3
Potato chips	31.1	24.4%	6.0	37.4%	9.3
Salty snacks	27.1	23.9%	5.4	37.0%	8.0
UHT milk	23.4	33.9%	5.7	38.8%	7.9
Average	28.5	28.0%	6.0	37.7%	9.1

Note: E(true IPT) = expected true inter-purchase time

[1] We plan to present the results of our outlined work-in-progress at the EMAC conference.

The average observed IPT (across all nine categories 28.5 days) is much higher than the mean expected true IPT resulting from the two- and three-parameter model (6 days and 9.1 days respectively). This relation also holds on the category-level and results from the assumption that customers visit multiple retailers. Thus, observed IPTs from a single retailer are inflated estimates of true (unobserved) IPTs.

In the three-parameter model, the SOWs show hardly any differences across categories. The average SOW ranges from 37.5% for coffee capsules to 38.8% for UHT milk, with an average of 37.7% across all nine categories. In the two-parameter model, there is greater variation. The average SOW ranges from 23.9% for salty snacks to 33.9% for UHT milk, with an average across all nine categories of 28.0%. For the mean expected true IPTs, the three-parameter model shows greater variation (mean expected true IPTs range from 6.3 days for lemonade to 14.5 days for dish liquid). The two-parameter model delivers estimates ranging from 4.2 days for lemonade to 8.6 days for dish liquid. The differences in the variation of SOW and expected true IPT across categories illustrate that both models explore the parameter space very differently. Consequently, the question arises which approach better describes customers' true purchase behavior.

4.2 Model validation

As we do not know the true IPTs on category-level and customers' SOWs, it is not possible to directly validate our results. However, we do know the observed IPTs in the estimation and the holdout period and use them for model validation. Following Chen and Steckel (2012), we use the estimations for the observed IPTs to predict the (observed) IPTs in the holdout period. We do this using two approaches: (1) the hierarchical Bayesian approach by Chen and Steckel (2012) (see section 2.4), and (2) a direct approach that estimates all parameters on an individual level (without the hierarchical layer)[2]. In addition, to the original three-parameter model by Chen and Steckel (2012) and our simplified two-parameter model, we use the observed IPTs of the estimation period as a benchmark. We compare the models using the root mean squared error (RMSE) and present the results in Table 2.

Table 2: Model validation using the root mean squared error (RMSE)

		Two-parameter model		Three-parameter model	
Category	Benchmark	Direct	Hierarchical	Direct	Hierarchical
Capsule coffee	19.33	18.71	17.60	18.77	17.58
Cola	17.37	16.89	14.80	16.85	14.79
Detergent	27.68	27.51	27.59	27.59	27.54
Dish liquid	30.09	30.37	29.97	30.57	30.02
Fresh milk	18.02	17.77	15.98	17.75	15.93
Lemonade	19.38	18.12	15.81	18.07	15.76
Potato chips	25.50	23.91	21.27	23.79	21.29
Salty snacks	22.75	22.17	19.23	22.09	19.31
UHT milk	20.01	18.98	18.55	18.94	18.42
Average	22.24	21.60	20.09	21.60	20.07

We find that the direct and hierarchical two- and three-parameter model provide lower RMSE and consequently better predictions than the benchmark (average RMSE across all nine categories is 22.24 days). The average RMSE across all nine categories is 21.60 days for the direct two- and three-parameter model. The introduction of the hierarchical layer further improves the average RMSE to 20.09 days in the two- and 20.07 days in the three-parameter

[2] Due to limited space, we decided to only present the hierarchical Bayesian approach in section 4.1.

model. A category-level analysis provides similar results. Thus, there is almost no difference between the two- and the three-parameter model with respect to RMSE.

However, there is a difference in runtime across all models. The direct two- and three-parameter models took 38 and 42 minutes respectively, parallelized on a server with 4 CPUs. The hierarchical models, however, took 17.9 hours in the two-parameter case and 28.9 hours in the three-parameter case. Thus, from a runtime perspective, our hierarchical two-parameter model has an advantage over the original three-parameter model by Chen and Steckel (2012).

5 Summary and Discussion

In this paper, we apply the modelling approach by Chen and Steckel (2012) to infer customers' true category-level IPT based on data from a single company. We use the model that was initially designed to forecast purchases in the grocery category based on credit card transactions to predict category-level grocery purchases based on transaction data from a single retailer. Both, the original model by Chen and Steckel (2012) and our proposed simplified model, predict observed purchases at the focal retailer equally well. However, the model results differ substantially with respect to category SOW predictions and predictions of true category-level IPTs. We will further explore these differences in the future and apply both models to our focal retailer's recommendation system. A successful application will enable us to (1) improve the timing of personalized price promotions based on customers' true category IPTs and to (2) further assess both models' performance and ability to describe customers' purchase behavior. The latter will also enable us to characterize the retailer's customer base with respect to cherry-pickers and systematic store switchers.

Finally, we can conclude that our model simplification provides a way to predict observed IPTs as accurate as the original model by Chen and Steckel (2012), while considerably speeding up the estimation process.

6 References

Chen, Y., & Steckel, J. H. (2012). Modeling credit card share of wallet: Solving the incomplete information problem. *Journal of Marketing Research*, *49*(5), 655-669.

Fox, E. J., & Hoch, S. J. (2005). Cherry-picking. *Journal of Marketing*, *69*(1), 46-62.

Gijsbrechts, E., Campo, K., & Nisol, P. (2008). Beyond promotion-based store switching: Antecedents and patterns of systematic multiple-store shopping. *International Journal of Research in Marketing*, *25*(1), 5-21.

Herniter, J. (1971). A probablistic market model of purchase timing and brand selection. *Management Science*, *18*(4-part-ii), P-102.

Ho, C., & Chen, J. (2015). *Point of view: The Modern Day Food Gatherer*. Retrieved from https://www.ipsos.com/sites/default/files/2017-03/IpsosMarketing_POV_ModernDayFoodGatherer.pdf. (Last accessed: December 5, 2017).

The Nielsen Company. (2015). *The future of grocery: e-commerce, digital technology and changing shopping preferences around the world*. Retrieved from https://www.nielsen.com/content/dam/nielsenglobal/vn/docs/Reports/2015/Nielsen%20Global%20E-Commerce%20and%20The%20New%20Retail%20Report%20APRIL%202015%20(Digital).pdf. (Last accessed: December 5, 2017).

Vuckovac, D., Wamsler, J., Ilic, A., & Natter, M. (2016). Getting the Timing Right: Leveraging Category Inter-purchase Times to Improve Recommender Systems. In S. Sen, W. Geyer, J. Freyne, and P. Castells (eds.), *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 277-280). New York, ACM.