Predicting Patterns of Customer Usage, with Niftecash

Bargary, N. and Berkery, E.M. and Cellai, D. and Dobrzynski, M. and Faqeeh, A. and Krzyzanowski, G. and Melnik, S. and Nyamundanda, G. and O'Callaghan, R. and Sheikhi, A. and Sikora, M. and Vylegzhanin, E. and Wardell, O. and Wilson, R. E.

June 2013

Predicting Patterns of Customer Usage, with Niftecash

N. Bargary[1], E. M. Berkery[2], D. Cellai (davide.cellai@ul.ie)[2], M. Dobrzynski[1], A. Faqeeh[2], G. Krzyzanowski[3], S. Melnik (sergey.melnik@ul.ie)[2], G. Nyamundanda[2], R. O'Callaghan[4], A. Sheikhi[2], M. Sikora[3], E. Vylegzhanin[5], O. Wardell[6], R. E. Wilson[7]

[1]University College Dublin
[2]University of Limerick
[3]Wroclaw University of Technology
[4]Cornell University
[5]Heriott-Watt University
[6]The High School Rathgar, Dublin
[7]University of Bristol

1. Introduction

Niftecash (NC) is a company that issues pre-paid cards, for example gift cards. They have several products in operation, and launching a new product is a difficult task because little is known about customer behaviour and how it affects revenue generation. The group was asked to analyse historical data over about 5 years with the aim of understanding how customer behaviour affects revenue generation.

The revenue is generated from the following three sources. The first source is the transaction (interchange) fee: this is typically a small percentage of the transaction amount that NC receives from each transaction. This fee is paid by the retailer, not the card user. Moreover, there is a fixed 10c transaction fee imposed on NC for the use of the payment circuits. As a result, small transactions are actually not profitable. The second source is the card 'expiry' fee. These are monthly charges of €3 deducted directly from balances of cards older than 1 year. This was different before May 2011 where, in addition to a monthly fee, the entire balance was taken as a fee at the end of a card's life (3 years from the date of issue). Accounts opened since May 2011 do not expire but are run down gradually. Finally, the third source of income is the investment of balances on cards.

Broadly speaking, an ideal customer is someone who

- loads a large amount on the card, that is kept there for a long time before any of it is used
- makes mostly large transactions
- leaves a non-zero balance on the card after 1 year, that is then depleted due to monthly expiry fees.

2. Data set description

The data provided to the group covers the period from January 2007 to June 2013 and consists of Load data, Transaction (Tx) data, and Fees data.

Load data: This contains date and amount loaded on cards. Each card is only loaded once, at the beginning of its usage period. The dataset contains approximately 1.4M records of card loads (Date, Amount, AccNo).

Tx data: This contains details of the transactions, *i.e.* money spent by retailers when requested by customers. Transactions can take place in any currency, however they must be settled against the currency that the account is created in. Information is gathered on the location of the retailer and at the business they are in. Fees are gathered for each transaction. Each account may potentially have multiple entries in this dataset as it is used over its life cycle. An account can be charged at any time up until it has been depleted. The dataset contains approximately 2.8M records of transactions (Date, TxAmt, TxFee, AccNo, TxCountry, TxMCC).

Fees data: This contains the expiry fees applied to each account. For all recent accounts, this is a service fee which is applied monthly from month 13 since the card was loaded; these recent accounts do not expire, instead they are depleted slowly month by month (unless the outstanding balance is fully spent by the customer). Accounts that were opened before May 2011, in addition to a monthly fee, also expire at the end of their lifetime (3 years since the date of issue of the card) and the balance is taken as a fee (breakage fee). Accounts opened since May 2011 do not expire but are run down gradually. Each account may potentially have multiple fee entries on this dataset as it is depleted over its life cycle. Accounts only incur fees towards the end of their life cycle, so most accounts do not incur any fees. The dataset contains approximately 1.2M records of card fees for cards older than 1 year (Date, FeeAmt, AccNo).

Spending patterns

Figure 1 shows the total card values per year and the corresponding amount of aggregated transaction and expiry (including breakage) fees. It is interesting to note that the total loaded amount kept growing during the recession years, whereas the income from transaction fees remained stable and the one from expiry fees dropped slightly from 2011 to 2012. This may be interpreted as a higher awareness of customers, who were spending the card more efficiently in year 2011.

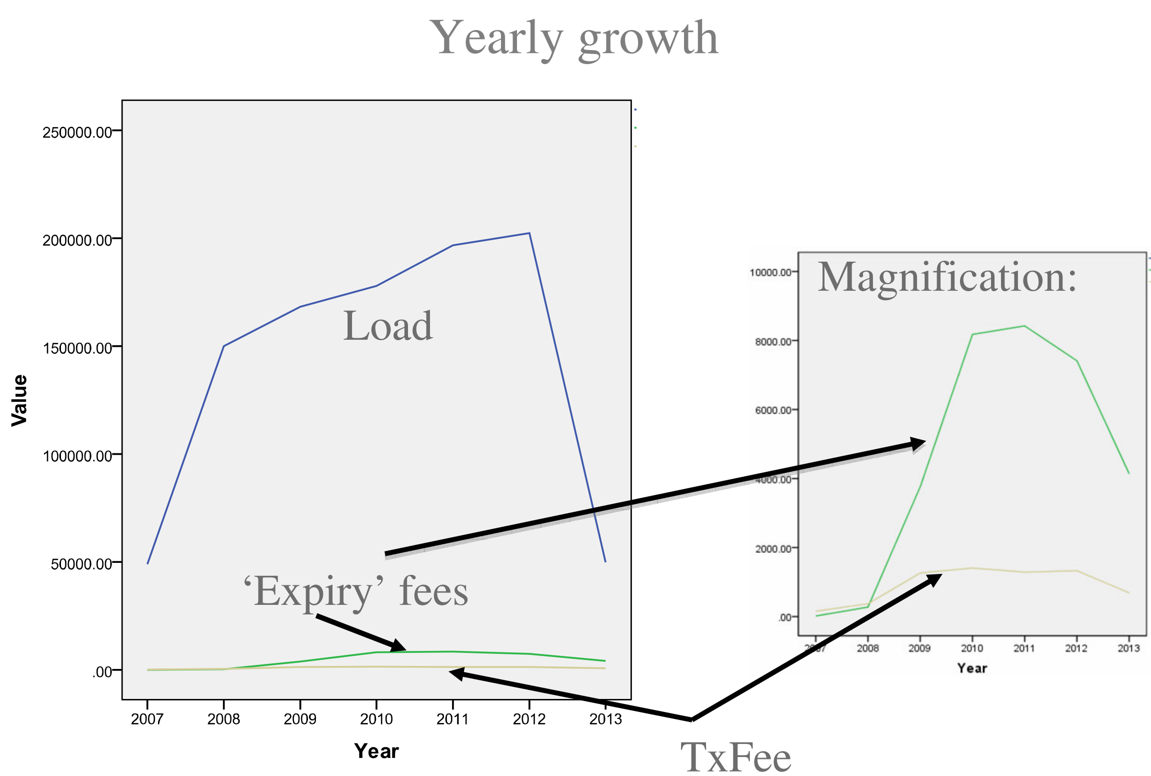


Figure 1. Yearly values of aggregate card loads, expiry fees and transaction fees (in euro). Year 2013 is incomplete, therefore it has lower values than year 2012. Most profit is made from expiry fees. Observe the positive correlation between expiry fees and transaction fees, which is probably due to the common dependence on the number of cards.

Figure 2 shows two examples of single customer behaviour. In the first panel, the cumulative earnings are plotted against time. It is easy to see that small transactions occurring around the end of 2011 represent a cost for NC (due to the 10 cent fee to the electronic service), which pushes the total earnings into the negative region. The second panel shows the time evolution of the outstanding balance for another customer. Here it appears that the customer tends to use most of the card balance within a few weeks in relatively big transactions. Then, smaller transactions occur and the tiny amount left on the card is wiped by the first expiry fee.

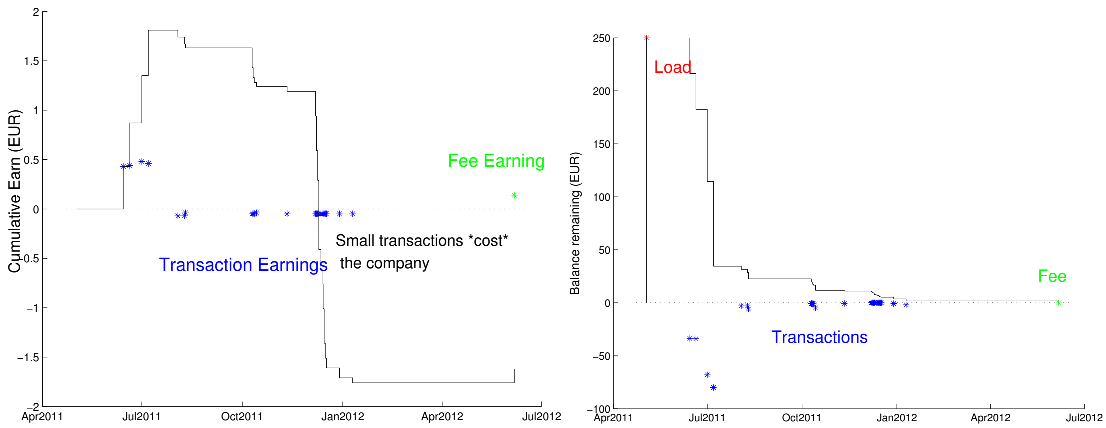


Figure 2. Time evolution of card balance (left panel) and cumulative earn (right panel) for two different accounts. Blue stars represent transactions, while green stars represent monthly expiry fees.

Figure 3 shows a comparison of two customer behaviours as the time evolution of card balance and NC earnings are plotted. In the left panels we can see that the customer makes a single (large) transaction in December 2008 and then leaves the remaining balance unspent, which is fully collected by the expiry fees starting in November 2009. In the right panels, the amount left on the card balance is much higher (about 50€, half of the loading amount), and it eventually becomes NC earnings due to the expiry fees. The unspent balance is much higher than the earnings from the transaction fees.

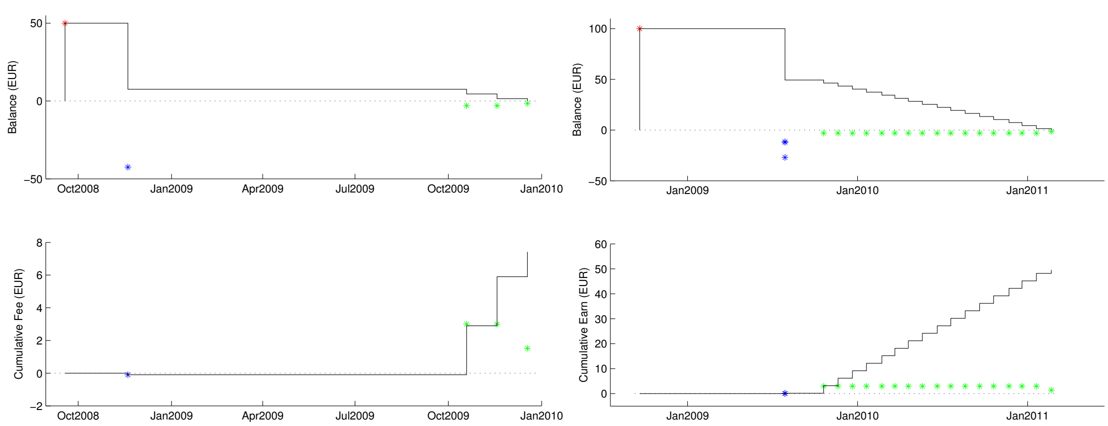
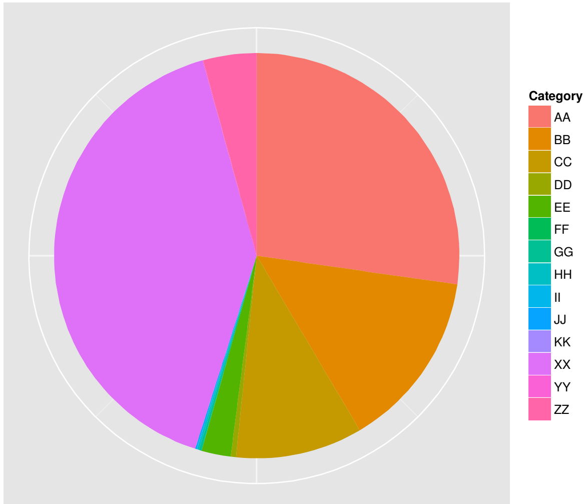


Figure 3. Two examples of individual customer behaviour. These figures show card balance and cumulative profit against time. Card load is represented by a red star, transactions are blue stars, and monthly expiry fees are green stars. The customer on the right made two transactions on the same day, and never used the card again, thereby generating about 50 euro profit from monthly fees.

Customer schemes

Customer behaviour can be captured by the different schemes created by NC. In Fig. 4, we show the breakdown of the different schemes in the card population. The largest proportion of cards is in the XX scheme with approximately 41% of all cards loaded. Scheme AA accounts for 27% of loaded cards, while 14% of loaded cards are in scheme BB and 10% of loaded cards are in scheme CC. The remaining 8% of loaded cards are distributed across the remaining schemes. We will investigate further if and how these schemes can help in differentiating between more or less profitable customers of NC.



Scheme	# Cards	Fraction
XX	575462	0.4077
AA	384839	0.2727
BB	201033	0.1424
CC	142874	0.1012
ZZ	60146	0.0426
EE	32956	0.0234
DD	5876	0.0042
GG	2816	0.0020
HH	2276	0.0016
JJ	1638	0.0012
YY	630	0.0005
II	477	0.0003
FF	270	0.0002
KK	135	0.0001
Total	1411428	1

Figure 4. The pie chart and the table show the fraction of customers per scheme. XX, AA, BB and CC are the four largest schemes.

Latency time

Further insight into the spending behaviour can be gained from an analysis of the latency time. We define latency time of a card as the time between the load date and the day of the first transaction. Figure 5 shows the distribution of latency times among all cards. We can see that most customers use the card within the first weeks, but there is also a long tail at longer times. The second week appears to be the one where most "first transactions" occur. After one year, one can distinguish a small peak, which suggests that some customers are motivated to start using their cards due to the upcoming monthly fees. Figure 6 shows that there is no substantial difference between cards loaded on weekends or on weekdays.

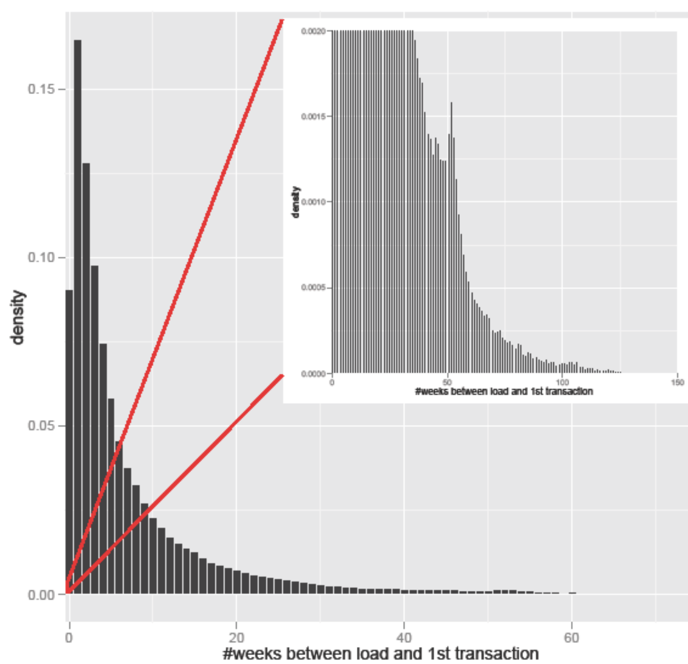


Figure 5. Latency time distribution among all cards showing the fraction of cards versus the number of weeks between the load and the first transaction. Most cards are first used within a few weeks. A few cards are first used after 2 years,

and these are probably the most profitable cards. In the zoom, one can distinguish a spike at one year, which suggests that some customers are motivated to start using their card due to the upcoming monthly fees.

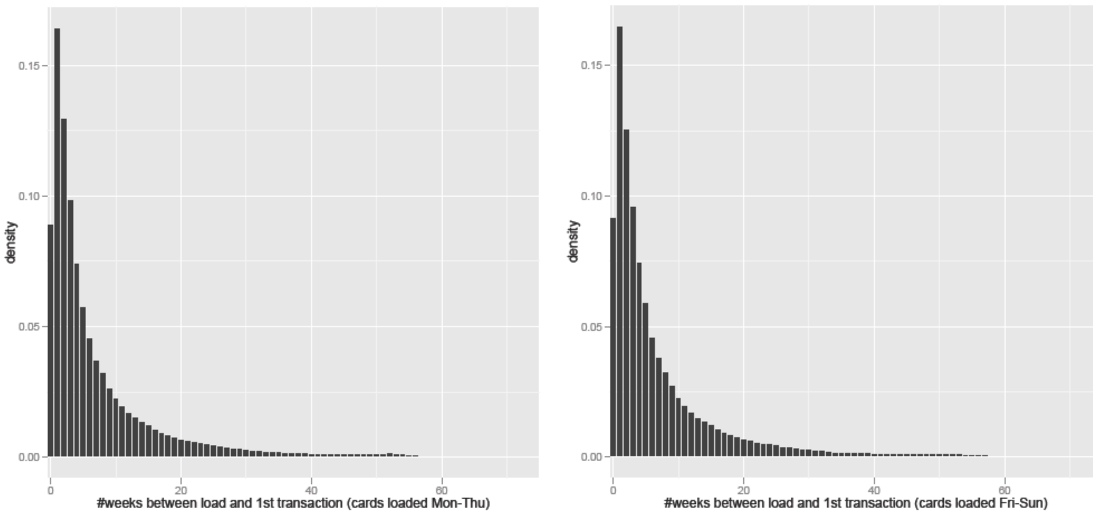


Figure 6. Same as in Fig. 5, but separately for cards loaded on Mon-Thu and Fri-Sun.

Figure 7 shows instead the distribution of latency times at different initial load amounts. While most cards are spent within the first weeks, we can also see that the tail of the distribution at high latency time extends quite significantly. The most interesting customers are the ones with the latency time larger than about 55 weeks (13 months) as they are the ones who most substantially contribute to the earnings through expiry fees. Among those, the ones with a larger total initial amount are the most profitable ones. This figure seems to suggest, even if as a qualitative remark, that the distribution of latency time does not shrink dramatically at higher initial loads. In other words, it seems that the smaller tails are due to the smaller number of points than a more peaked distribution around the mean. This is an interesting observation, as it implies that increasing the number of cards with high initial load could lead to an increase in earnings from expiry fees.

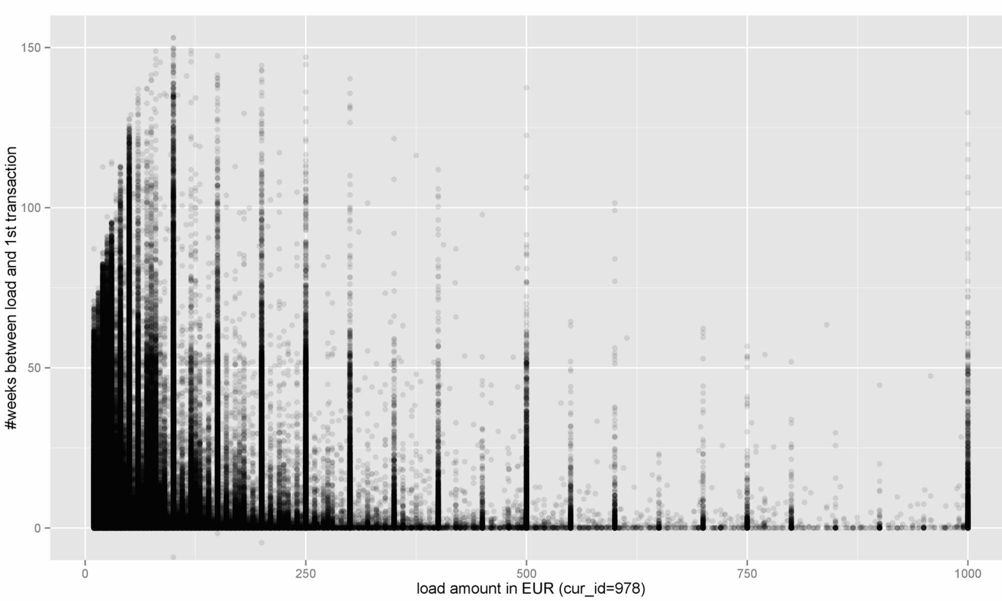


Figure 7. Distribution of latency times between load date and first transaction for different initial load amounts.

Merchant codes

The dataset also allows us to analyse the spending volumes in different shops. Figure 8 shows the seasonal patterns in the number of transactions for different merchant codes. There is a clear peak in the number of transactions at Christmas time. It is interesting to note different patterns for different merchant categories (higher family clothing volumes in 2012, higher music purchases in 2008, etc.

The Music Store example indicates the change in people spending patterns in relation to how they buy music: they are no longer buying their music in shops with many people presumably moving to online purchases. In this example the majority of transactions have occurred in 2008 with obvious decline in the subsequent years.

In the case of Ryanair the seasonal pattern seems to be less pronounced. In this example, years 2010, 2011 and early 2012 look busier than other years which is perhaps related to the absence of Ryanair card fees for purchasing tickets during these years.

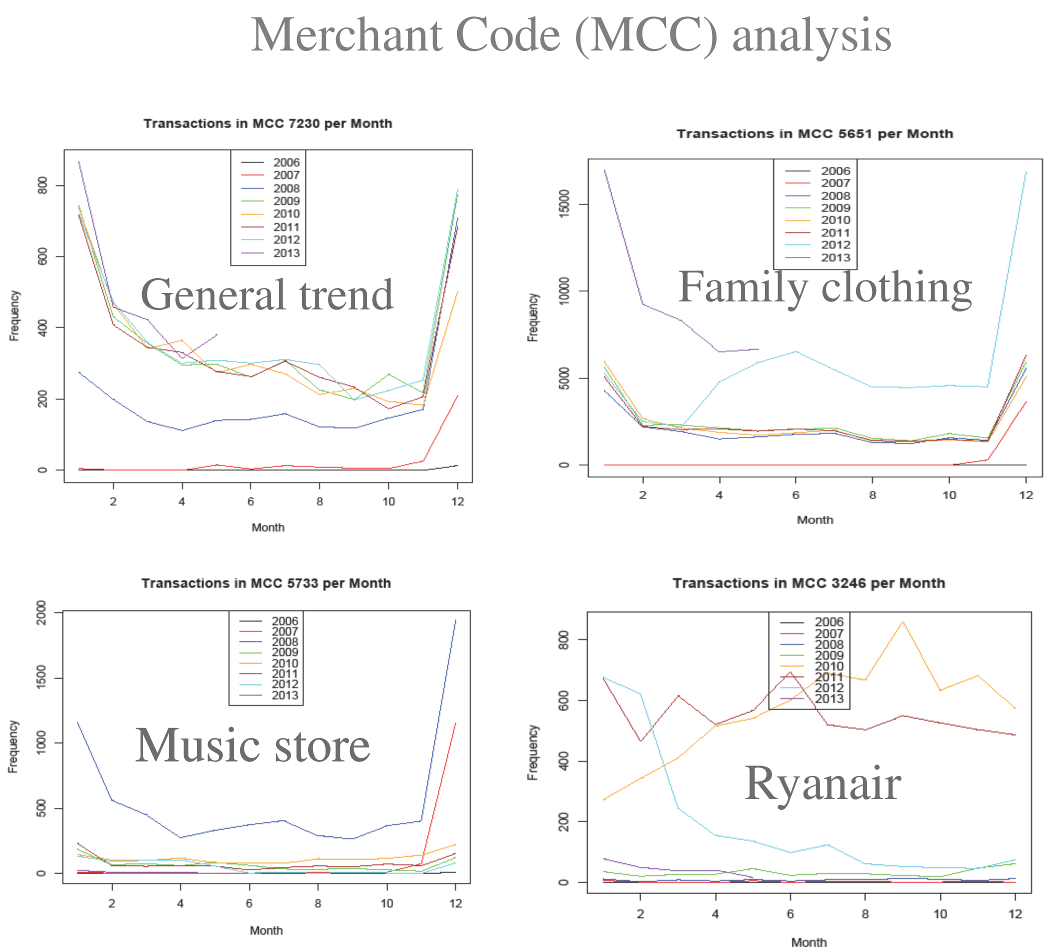


Figure 8. Yearly spending patterns for different years for different merchant categories. In general, there is a high volume of transactions during the Christmas period and a much quieter period during the rest of the year. Notably there are some exceptions, such as the Ryanair example, where the seasonal pattern seems to be less pronounced.

In Fig. 8.1(a), we plot the total number of transactions for each merchant code sorted in descending order by the number of transactions. We keep the same order of merchant codes in Figs. 8.1 (b) and (c) where we show the aggregate bill and fee amounts for each merchant code. The majority of transactions are distributed over just a few merchant categories: Men/Women clothing stores, Department stores, Family clothing stores and Women’s ready to wear stores. For these major categories, the aggregate bill amounts follow in the same descending order as the number of transactions, while the transaction fees for Family clothing stores exceed that for Women’s ready to wear stores, which may indicate that the former category has larger average transaction amount. Also in Fig 8.1(c), some fee amount points are negative because of the 10c cost associated with each transaction. Excluding these merchant categories or encouraging people to spend more in each transaction may help solve this problem.

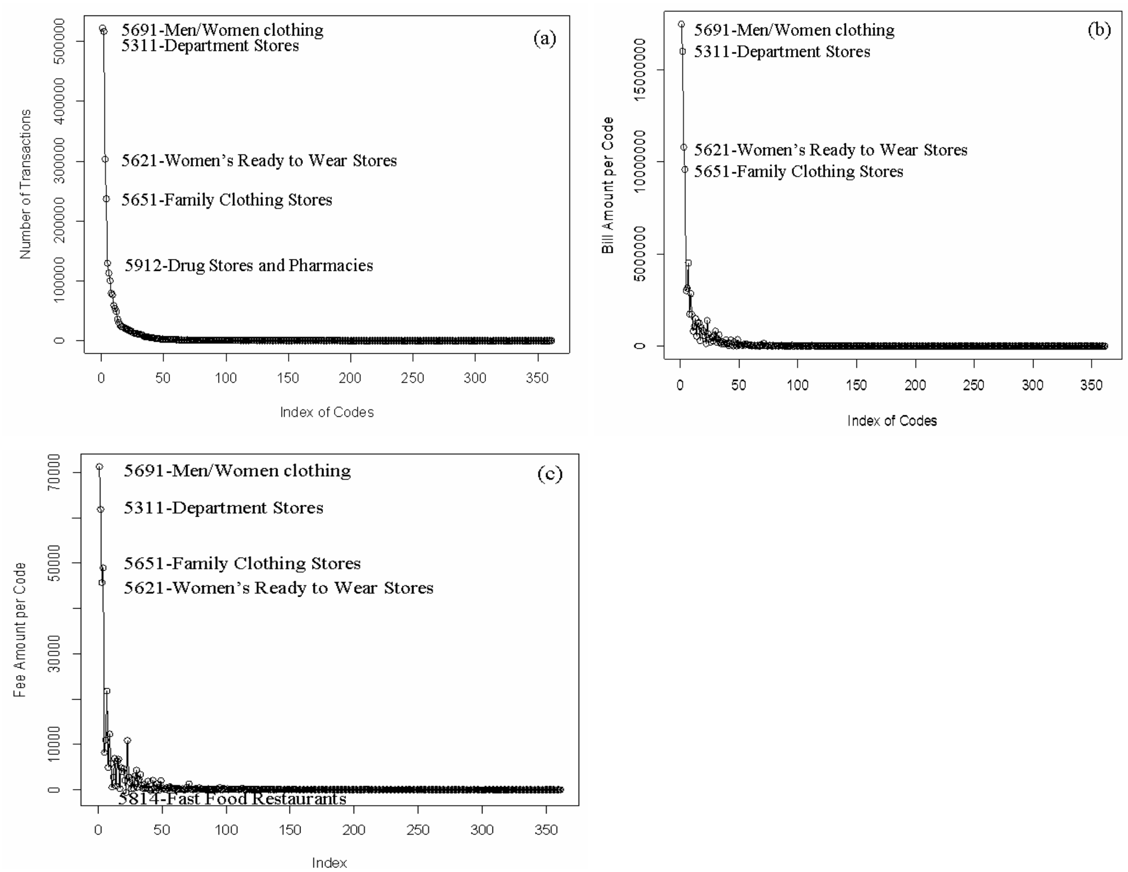


Figure 8.1. (a) The number of transactions for each merchant code sorted in descending order by the number of transactions. (b) The aggregate bill amount for all transactions within each code (codes appear in same order as in (a)). (c) The aggregate fee amount for all transactions within each code (codes appear in same order as in (a)). Some fee amount points are negative as there is a 10c cost associated with each transaction.

3. Results

Transaction Fee analysis

General information

We now evaluate the income exclusively generated by transaction fees. In the following, 1,412,548 (unique) cards are loaded, corresponding to the cards where we have a full history. 1,120 cards have a load value of 0 and have therefore been removed from this analysis. 97,557 of the cards have load data but not a single spend transaction. The histogram in Fig. 9 shows the distribution of cards in the different schemes.

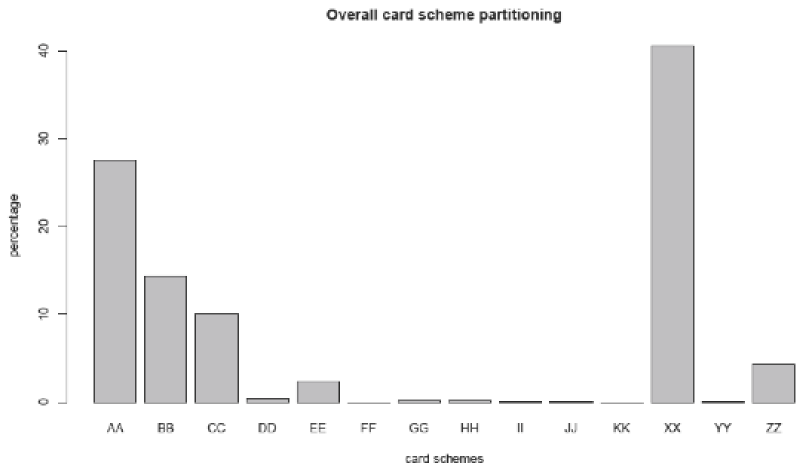


Figure 9. Histogram of the card distribution across schemes.

Load size by scheme

To determine if there are any differences between the initial load amounts across schemes, we categorise the initial load values into 7 groups (euro):

<30
30-50
50-75
75-100
100-250
250-500
>500

We calculate the proportion of cards loaded with each amount in each scheme. Figure 10 shows this breakdown.

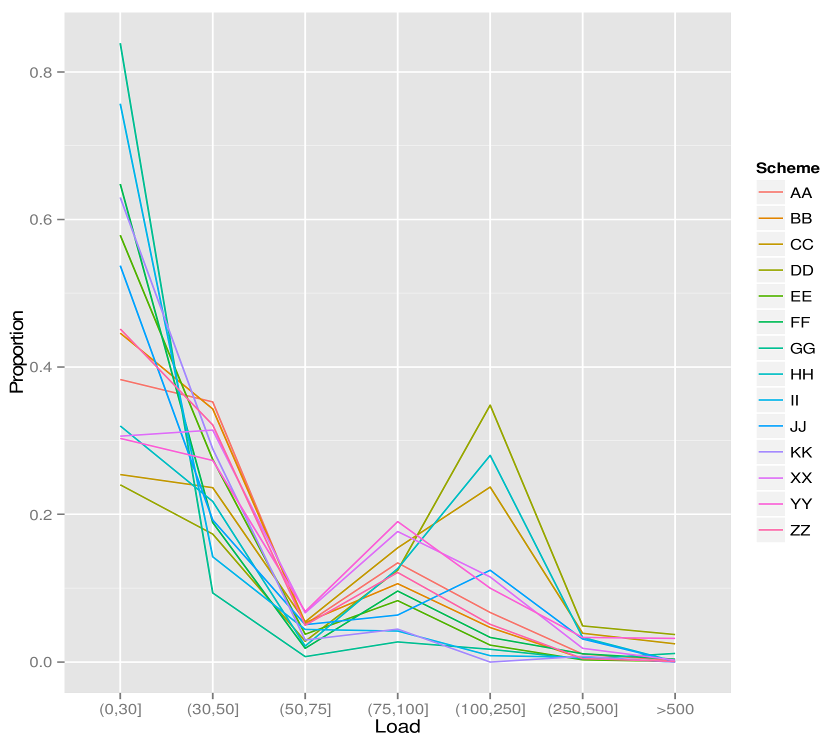


Figure 10. Distribution of initial card load for different schemes.

Within the AA scheme, 38% of cards are loaded with amounts under €30 and a further 35% are loaded with amounts between €30 and €50. Similar results can be seen in the BB scheme, with 45% of cards loaded with amounts under €30 and 34% of cards loaded with amounts in the range €30-€50. In the CC scheme there are approximately equal proportions of cards loaded with under €30, €30-€50 and €100-€250 (25%, 24% and 24% respectively). The majority of cards in the DD scheme are loaded with €100-€250 (35%) with a further 24% loaded with under €30. The majority of cards in the EE, FF, GG, II, JJ and KK schemes are loaded with under €30 (58%, 65%, 84%, 76%, 54% and 63% respectively). In the HH scheme 32% of cards are loaded with under €30, 28% of cards were loaded with €100-250 and a further 22% of cards are loaded with €30-€50. In the XX scheme 31% of cards are loaded with under €30 and a further 31% are loaded with €30-€50. 18% of cards in this scheme are loaded with €75-€100. 30%, 27% and 19% of cards in the YY scheme are loaded with under €30, €30-€50 and €75-€100 respectively. Finally, 45% of cards in the ZZ scheme are loaded with under €30, 32% are loaded with €30-€50 and 12% of cards are loaded with €75-€100.
Broadly speaking, therefore, most cards are loaded with small amounts in the main schemes. However, it is remarkable to notice that there are three schemes (CC, DD, HH) where the fraction of initial loads in the €100-€250 section is higher than 20%. This signals a probable peculiarity in the customer composition of these schemes.

Income from transaction fees

The most profitable transactions for NC are the larger ones. Broadly speaking, it is quite intuitive that cards with higher initial load are used for higher transactions. This is clearly shown in Fig. 11, where the distributions of transaction fees are displayed in each load category.

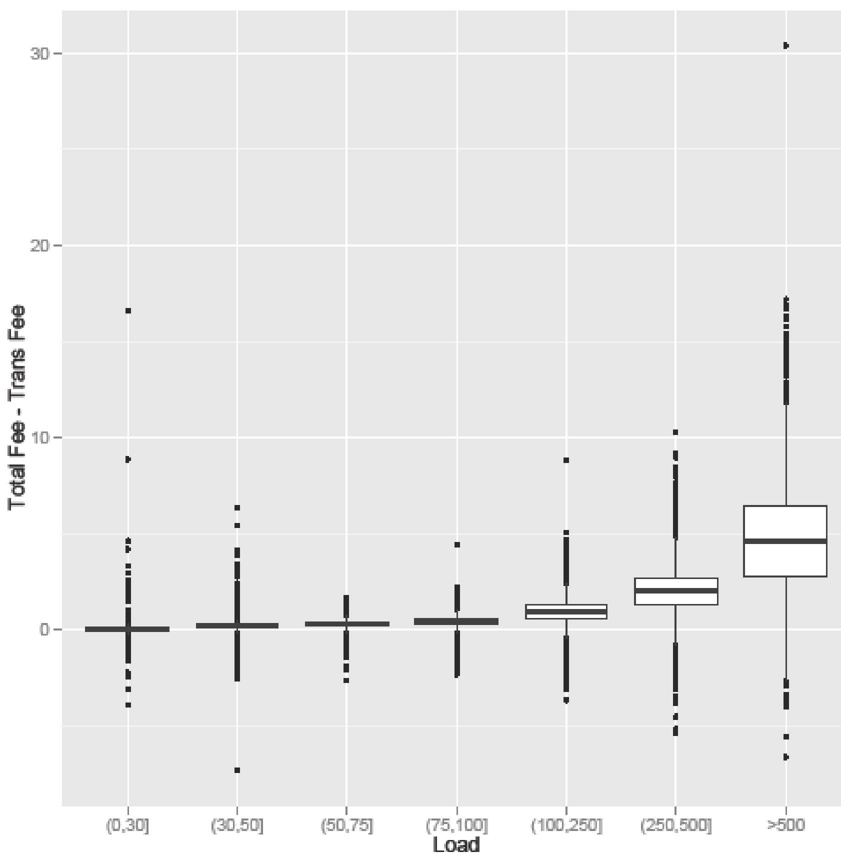


Figure 11. Distribution of earnings from transaction fees for different initial card loads.

The total transaction fee income is calculated by summing all the transaction fees per card and subtracting 10 cent for each transaction carried out. Figure 12 shows the distribution of income across schemes.

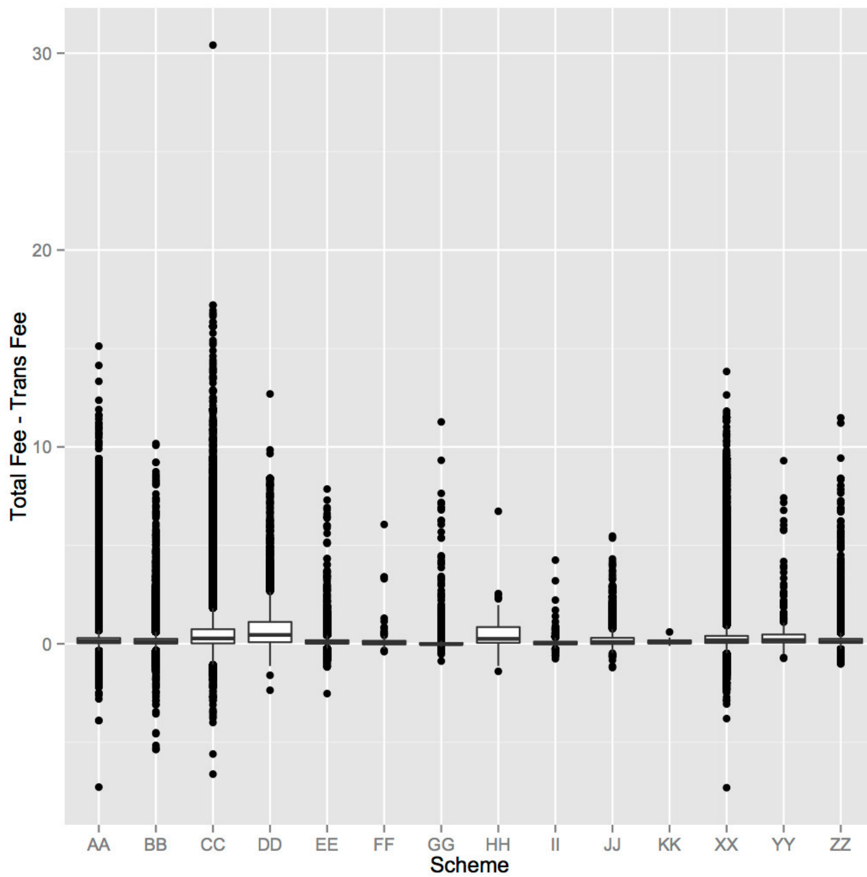


Figure 12. Distribution of earnings from transaction fees for different customer schemes.

Scheme DD appears to generate the most income. The very large outlier in the CC group is a customer who spent the card on a Caribbean cruise ship. The income per card per scheme was calculated by taking the total income per scheme and dividing by the number of cards in each scheme. Especially in the most popular schemes, it is noticeable that the tail of the distribution extends into negative income. This is due to the fact that transactions of low amount are not profitable as the costs of the used circuit of payment are higher than the transaction fee.

Figure 13 shows a bar chart representing this data. Again, scheme DD appears to have the highest income per card. According to the features of scheme DD, mean transaction amount, initial load amount and income per card appear to be related. However, the small number of cards in this scheme does not allow a definitive conclusion. Among the four most popular schemes, CC is the most profitable one, while AA and XX are affected by the long tail in the negative income region. Setting a minimum transaction amount would obviously improve this outcome.

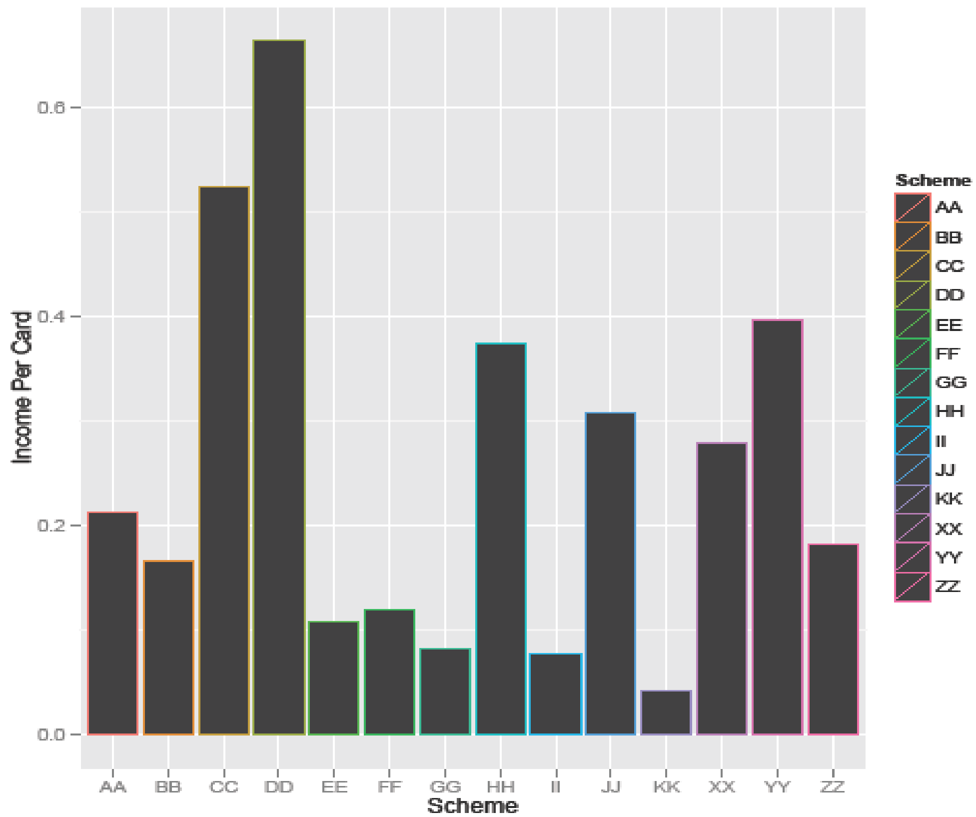


Figure 13. Distribution of earnings from transaction fees for different customer schemes.

In Fig. 13.2, we plot transaction fees versus transaction amounts. While the observed dependence is mostly linear, there are about a dozen different slopes which vary by an order of magnitude. We can conclude from this figure that transaction fee may be related the country where a transaction took place, though it also depends on some other factors. The exact relation between the transaction amount and the incurred transaction fee is unknown to us and requires further consideration.

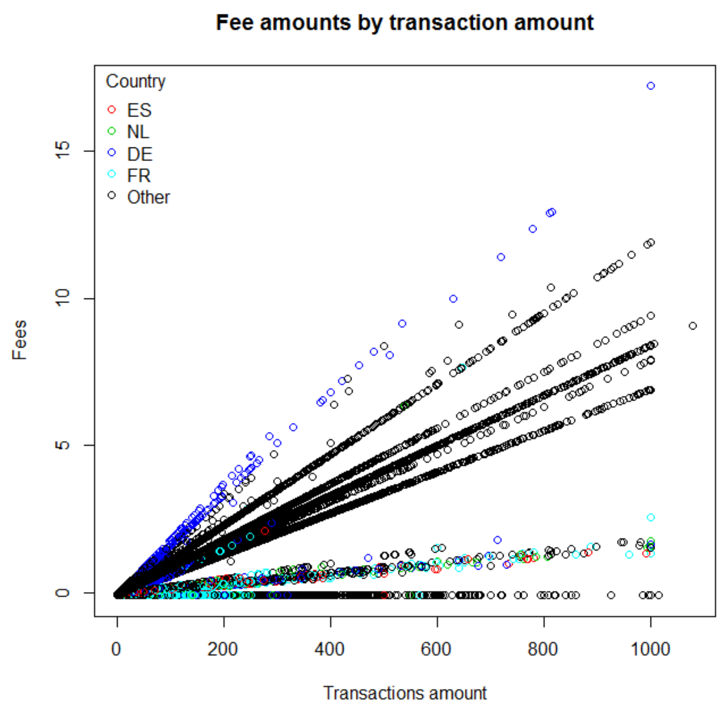


Figure 13.2. Plot showing linear dependence of transaction fee on transaction amount with several clearly visible bands.

The plots in Fig. 14 show the latency time (time occurred between load date and first purchase) distribution as a function of initial card load and customer scheme, respectively. It is interesting to note that while the behaviour of the curves is quite similar across different load amounts, a remarkable difference can be observed for the customer schemes. For example, customers in scheme JJ use the card much more quickly than average. On the other hand, characteristic peaks are observed in the scheme FF at about 21 weeks and 11 months. While the peak at 11 months could be related to the awareness that the card is close to the expiry fees, it is not easy to understand what is the cause of the peak at 21 weeks. This feature seems to be strongly related to the characteristics of the scheme, as it is not observed in the other ones.

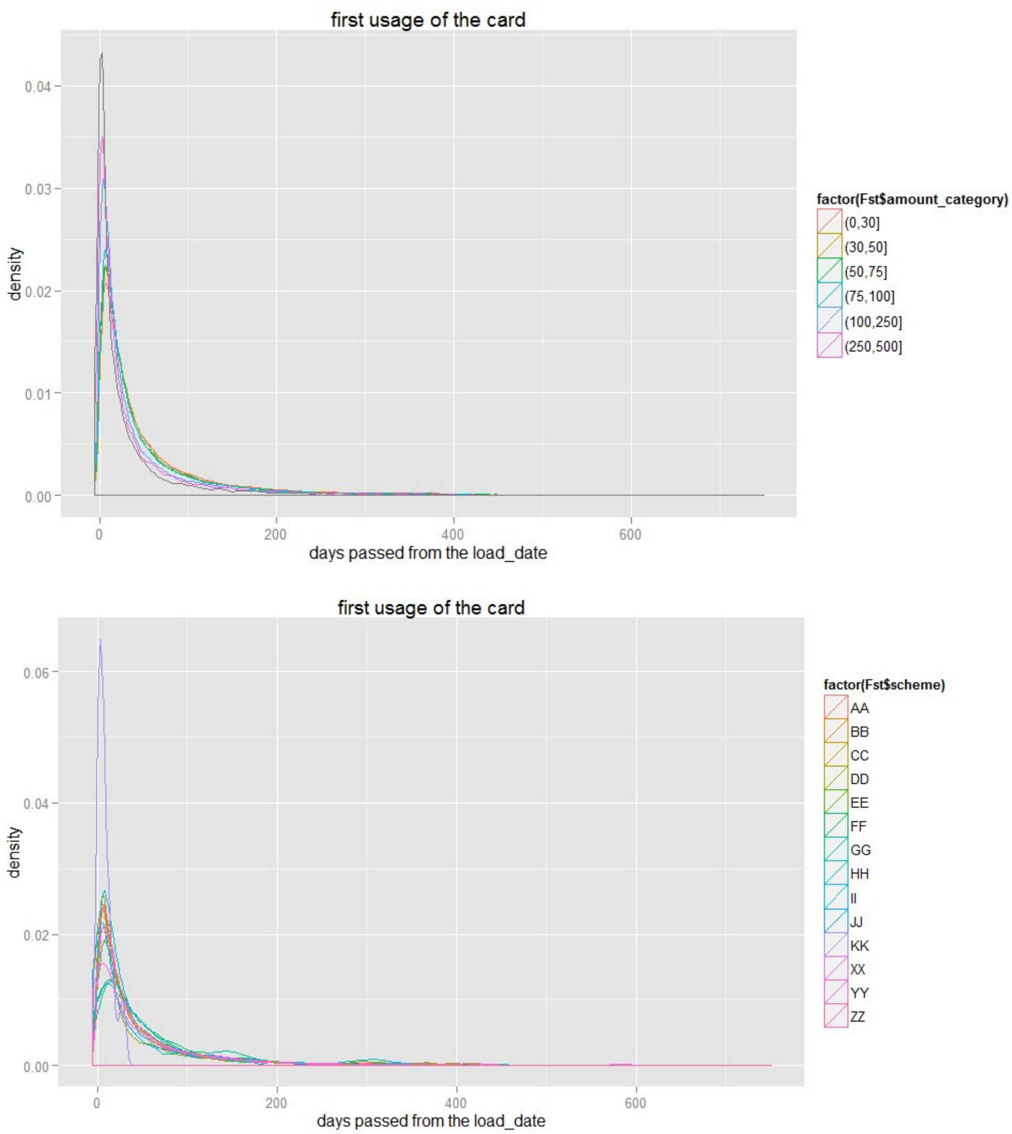


Figure 14. Latency time distribution for cards for various load amounts and various schemes. Apparently the time of first usage depends more on the scheme rather the value of a card. (Note that the curves shown are smoothed with a Gaussian function.)

Expiry Fee Analysis

We now examine the income generated from the expiry fees. Such fees consist of two types:

1. The monthly fees which are applied to the outstanding balance of the card after a year since the card was loaded.
2. The breakage fees that were charged for a period of time during the study three years after a card was issued.

First, let us examine the comparison between the transaction and expiry fees incomes over time. Figure 15 shows the total income generated from expiry fees (red) and transaction fees (blue), as a function of the card age (in days). Card age is defined as the difference between the date of the last transaction (the one which sets the balance to zero) and load date. Here we only consider all the cards in the dataset which have a full history.

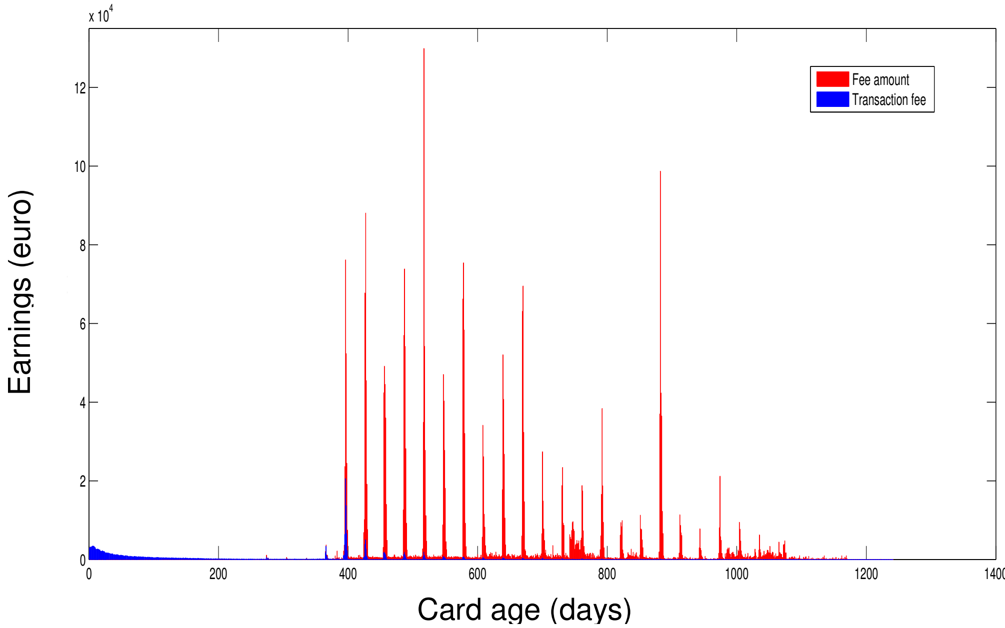


Figure 15. Comparison between income generated by transaction (blue) and expiry fees (red) as a function of card age.

The income generated from the transaction fee is clearly much smaller than the one from the expiry fee. As a general trend, we observe a peak in transaction fee at the beginning of the card's life and a few peaks around one year, probably due to the customer decision of using up the card because the start of the expiry fees. Starting from month 13, we can see the peaks of the expiry fees at the end of each month. The peaks are much higher than the ones due to transactions and do not decrease steadily, probably because of the breakage fees.

To further characterize the composition of expiry fee, we plot in Fig. 16 the number of cards that contribute a given amount of expiry fee (in euro). Most cards contribute 1€ or less, about 40,000 cards contribute 1-2€, etc… This distribution is not monotonically decreasing as there are peaks at multiples of 5€. This suggests that a sizeable number of cards that contribute to the expiry fee income are not used at all, as most cards are loaded with an amount which is a multiple of 5€.

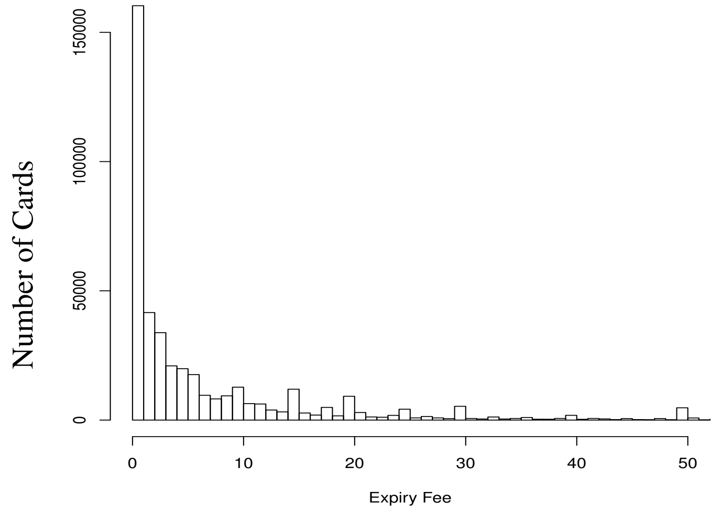


Figure 16. Histogram reporting the number of cards that contribute to a certain amount of expiry fee earnings (in euro).

Therefore, it is interesting to investigate the cards that have never been used, as those are the ones generating only expiry fees. Figure 17 shows the income generated by these cards.

In the period of study, about 1.4 million cards were loaded for a total amount of about 99 million euro. Among those, only 1.46% of those cards start to be used after one year of the card lifetime. This means that only a small fraction of customers start using the card after the expiry fees kick in. This generates a tiny income by transaction fees of about 4,600 euro.

On the other hand, it is very interesting to look at the cards that have never been used after one year before the end of the period of study. Those cards are about 3.5% of the total number of cards, they hold a capital of about 2.4 million euro, but they have generated an income of 1.1 million euro in expiry fees. Moreover, the histogram shows that the distribution of "forgetful customers" in each scheme is approximately the same as the one of the total population (cf. Fig. 9). In other words, none of the different card schemes capture the section of customers who are most profitable from the NC point of view.

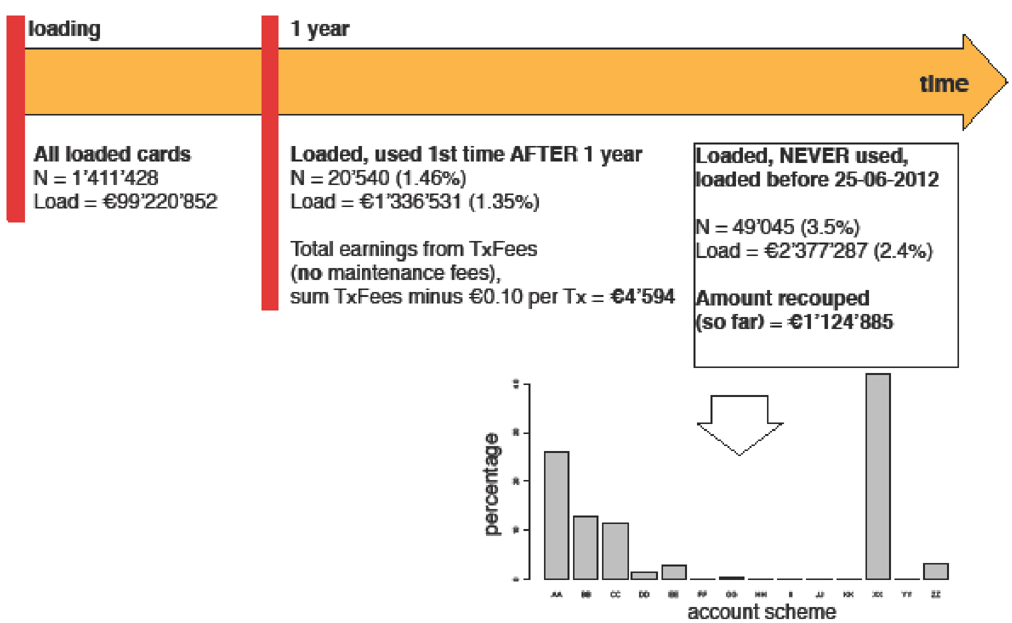


Figure 17. Illustration of the proportion and impact of "forgetful customers", *i.e.* customers who have never used their card. The histogram shows distribution of forgetful customers in each scheme and it is quite similar to the general distribution of customers shown in Fig. 9.

The relative homogeneity of the customer schemes is also shown by Fig. 18, where the mean fraction of outstanding balance is plotted over time for the most popular schemes. In all schemes we observe a similar exponential-like decay, with just a small difference in the intermediate regime.

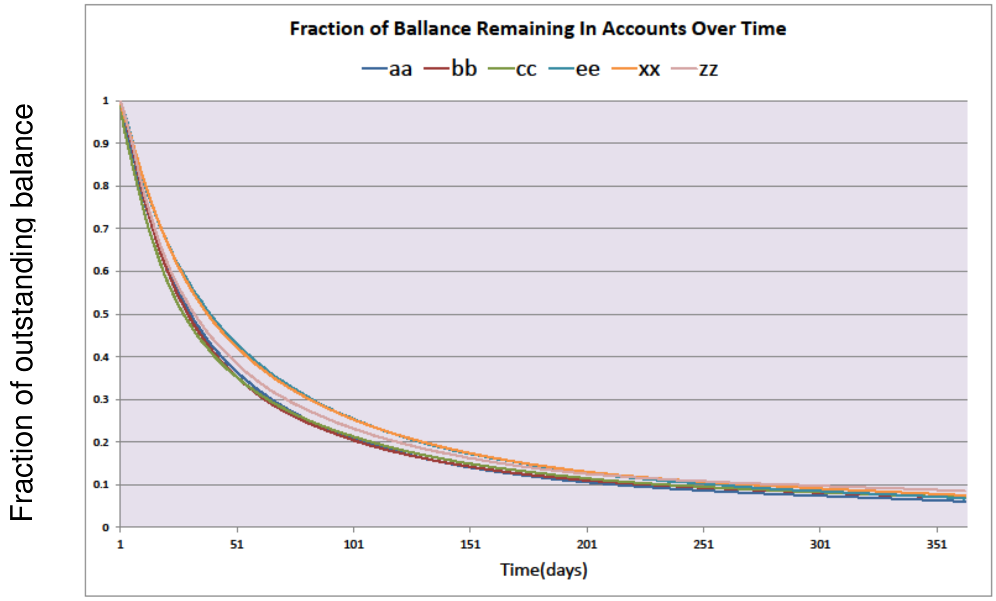


Figure 18. Time evolution of card outstanding balance (normalized by the initial load amount) versus card age.

4. Conclusions and directions for further analysis

Our data analysis shows a few interesting features:

1. Most of the income is generated by the expiry fees.
2. The most profitable customers are the "forgetful" ones, characterized by zero card usage (3.5% of the total population). These customers have generated 1.1 million euro in the period of study.
3. The fraction of forgetful customers is equally distributed in all schemes, so there is scope for improving customer characterization to capture higher fraction of forgetful customers.

Besides these main remarks, a set of interesting (and profitable) open questions appear to be worth exploring.

Regarding the transaction fees, we have shown that the scheme DD is the most profitable one (see Fig. 12). We have also seen that this scheme, together with CC and HH, is characterized by a large proportion of cards with high initial load (see Fig. 10). Therefore, it appears that the typical initial load amount could provide useful information on customer behaviour which can discriminate between cards used for large or small transactions. The difference in the profits from schemes DD and CC, then, seems to be due to the fact that the latter contains a larger fraction of small value cards that partially compensate for the earnings from the large value cards.

Regarding customer behaviour, it could also be interesting to characterize spending patterns from the behaviour of a customer in the first few weeks. Figure 19 displays the scatter plot of earnings per card from transaction fees, versus the time it took to spend 85% of the total card value. The scatter plots appear to be different in different schemes. This may be a proxy for further analysis.

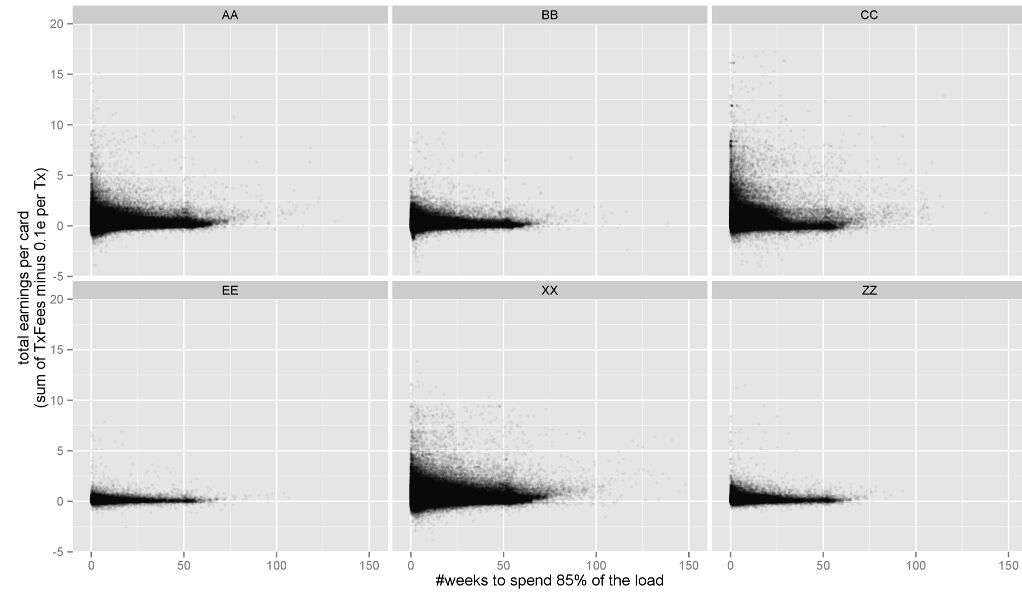


Figure 19. Scatter plots of total transaction fee earnings per card across the time used to spend 85% of the initial load value. The six panels report the most popular schemes.

One way to approach the problem in a practical way is to reduce the time series of each account balance to a set of 3-4 balances at fixed card ages (as shown in Fig. 20), and to study the correlations between these numbers and other characteristics such as initial load amount, fraction spent in the first two weeks, etc.

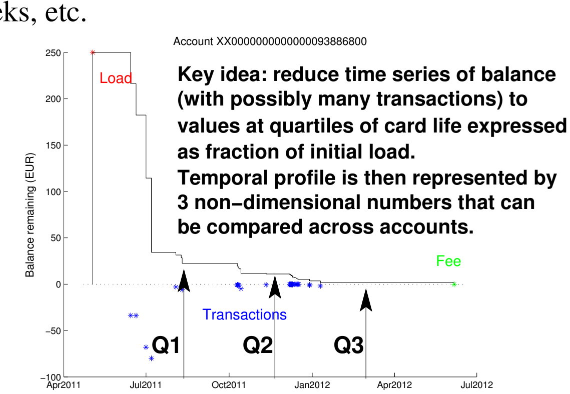


Figure 20. Example of schematic characterization of a customer spending behaviour.

Finally, there may also be potential in increasing the popularity of higher initial loads and even in increasing the maximum card load. The direct effect would be an increase in the potential earnings from forgetful customers. We have seen in Fig. 7 that the width of the distribution of latency times does not seem to shrink too fast with increasing initial load. This should be investigated further. From the available analysis, though, it seems that a high initial load does not reduce to zero the probability that the card goes unspent. Therefore, increasing the popularity of high load cards should have a benefit in terms of expiry fees earnings. Moreover, we can also see from Fig. 7 that the high number of cards at 1000€ may indicate a saturation, *i.e.* the market could be able to support cards with higher load value.

Acknowledgements

We acknowledge support of the Mathematics Applications Consortium for Science and Industry (MACSI) funded by the Science Foundation Ireland (SFI) Investigator Award 12/IA/1683, and the SFI conference workshop grant 13/CW/12578 and the Enterprise Ireland Innovation voucher scheme.