What can we learn about payment choice from seeing all retail transactions?[1]

Tamás Ilyés[2] – Lóránt Varga[3]

Draft version

Abstract

In our study we examined payment card acceptance and card usage in the Hungarian retail sector based on a receipt-level detailed database derived from online cash registers. The main objectives of our research were to identify the primary explanatory variables to test conventional card acceptance hypotheses and to verify the main findings of the international literature using this broad-based database and to identify features specific to Hungary. For the purposes of our analysis, we relied on anonymised online cash register data provided by the National Tax and Customs Administration (NTCA) for the years 2015 and 2016. Covering an extremely broad section of the Hungarian retail sector, with nearly 7.5 billion data points the database provides a basis for complex and robust analyses. Based on the robust results of our research, we found that store size can be considered to be the most important explanatory variable behind card acceptance decisions; however, the correlation is not linear. The marginal effect of size is negligible among small and large-sized stores, but there is a strong positive correlation among mid-sized stores. We also analysed the effect of the store's customer base and other attributes, and although numerous effects proved to be statistically significant, even they wielded negligible influence in card acceptance decisions. On the other hand, being open on Sundays – a subjective variable that was used as a proxy for store ownership – had a significant negative effect on card acceptance decisions. In payment instrument decisions payment value is the most important consideration, but the relationship is non-monotonic; above a value equivalent to 100 EUR it reverses course and the card usage rate begins to decline. As regards the proxy attributes, the result of questionnaire-based surveys can be confirmed, namely, that card usage is strongly influenced by income and education level. In this paper, we provide evidence that the ease of payment significantly influences card usage: when fewer denominations are needed and thus cash payment is faster, the card usage rate is significantly lower.

Journal of Economic Literature (JEL) codes: C44, D22; D12, G02, G20

Keywords: payment transactions, card usage, payment methods, logistic regressions, payment choice, card acceptance, big data

[1] Draft version, do not cite without author permission

[2] Economic Analyst, The Central Bank of Hungary email:ilyest@mnb.hu

[3] Head of department, The Central Bank of Hungary email: vargalor@mnb.hu

Extended abstract

In this paper we study payment card acceptance and payment card usage in Hungary's retail sector relying on a large, comprehensive database of the National Tax and Customs Administration (NTCA). Due to a lack of adequate direct data, most of the empirical studies in payment economies rely heavily on survey data collected from merchants and customers. Even if reliable data are available they usually cover only a small segment of the entire retail sector. However, the nature of survey studies makes it impossible to analyse small niches of the sector, e.g. very large or small transactions, small shops and the relatively few large merchants. We contribute to the studies on payment economies with a detailed empirical model-based description of the Hungarian retail sector and the empirical tests of the usual assumptions about payment behaviour.

From a payment system perspective, Hungary is paradoxical: more than 80% of households have access to payment services – payment accounts and, generally, debit cards. Acceptance among merchants is average compared with the European Union, but the use of electronic payment instruments is relatively low, with most of the retail sector heavily cash oriented. On the other hand, the use of payment cards continues to show a double-digit increase for years both in the value and volume of transactions. The aim of this paper is to find the main drivers of payment card acceptance and payment card usage. Aside from the usual drives, i.e. number and size of transactions, we also test whether behavioural variables have significant effect on payment method acceptance and choice. We test the assumption that different merchants by size face different dilemmas relating to payment acceptance, and the corresponding explanatory variables differ and also whether the time cost of transaction has any impact on the payment choice at the point-of-sale.

The basis of our analysis is the granulated, receipt-level database of the NTCA in Hungary. Since 2014, it has been obligatory to use online, NTCA-connected cashier machines across the retail sector in Hungary. We use data covering 2015 and 2016 from more than 170 000 merchants handling more than 7.5 billion retail transactions. Despite the anonymised database, we have information about the type of owner, location, primary branch of economy and network characteristics of the merchants.

In this paper we rely on logistic regressions estimating the likelihood of card acceptance per merchant and the Heckmann-selection corrected estimation of card use probability. The size of this database enables us to reliably estimate even small effects and complex functional forms of relationships on payment method acceptance and choice. We compare models for different sizes of merchants and have estimates for transactions that are usually excluded from payment survey – for example transaction over EUR 100.

We find that the most significant driver of payment card acceptance is the size of merchant, although the relationship is not linear. For merchants of different size the marginal effect of transaction volume differs significantly. It has the greatest importance for middle-sized shops, and the smallest for very small and very large merchants. We show that the average transaction value, the share of small and big transactions and the network size of the shop or franchise also have a statistically significant impact on acceptance in accordance with the usual assumptions. However, we find that although behavioural variables, e.g. if the shop is family owned, and location, customer base, are significant from a statistical perspective, they have a small impact on payment choice.

The 7.5 billion data points enable us to have adequate estimates for payment card usage even for very small and high-value retail transactions. We find that, contrary to the theoretical approach, payment card usage does not increase continuously with transaction value. Card usage stops increasing at around EUR 100 and declines for higher values.

Further research is required to determine if this is a unique Hungarian characteristic or it is true for most payment systems. Apart from the non-linear relationship, transaction size remains the most important determinant of payment choice, but we find strong evidence for behavioural and cost effects, e.g. the number of banknotes needed. Location and merchant size have a statistically significant effect, but a very small actual impact on payment choice.

Comparing the two years we find that the sharp increase in volume of card purchases comes mainly from relatively small transactions, but the sharpest increase in the percentage of payment card use comes from high-value transactions. The main reason for the difference is that the distribution of transaction size is asymmetrical, with most of the transaction being very small at 70% of all transactions below EUR 4.

This analysis is intended to contribute to the literature on payment economies by providing a comprehensive analysis of a retail payment system and testing the main hypotheses of the theoretical works on payment choice.

1. INTRODUCTION

In this paper we study payment card acceptance and payment card usage in Hungary's retail sector relying on a large, comprehensive database of the National Tax and Customs Administration (NTCA). Due to a lack of adequate direct data, most of the empirical studies in payment economies rely heavily on survey data collected from merchants and customers. Even if reliable data are available they usually cover only a small segment of the entire retail sector. However, the nature of survey studies makes it impossible to analyse small niches of the sector, e.g. very large or small transactions, small shops and the relatively few large merchants. We contribute to the studies on payment economies with a detailed empirical model-based description of the Hungarian retail sector and the empirical tests of the usual assumptions about payment behaviour.

From a payment system perspective, Hungary is paradoxical: more than 80% of households have access to payment services – payment accounts and, generally, debit cards. Acceptance among merchants is average compared with the European Union, but the use of electronic payment instruments is relatively low, with most of the retail sector heavily cash oriented. On the other hand, the use of payment cards for years continues to show a double-digit increase both in the value and volume of transactions. The aim of this paper is to find the main drivers of payment card acceptance and payment card usage. Aside from the usual drives, i.e. number and size of transactions, we also test whether behavioural variables have significant effect on payment method acceptance and choice. We test the assumption that different merchants by size face different dilemmas relating to payment acceptance, and the corresponding explanatory variables differ and also whether the time cost of transaction has any impact on the payment choice at the point-of-sale.

The basis of our analysis is the granulated, receipt-level database of the NTCA in Hungary. Since 2014, it has been obligatory to use online, NTCA-connected cashier machines across the retail sector in Hungary. We use data covering 2015 and 2016 from more than 170 000 merchants handling more than 7.5 billion retail transactions. Despite the anonymised database, we have information about the type of owner, location, primary branch of economy and network characteristics of the merchants.

In the first part of our paper we give a detailed description of the OCR database and show the main characteristics of data. The second section details the model to estimate the likelihood of card acceptance and main results. In the last section we estimate card use propability on the receipt level.

1.1. Data source: online cash register (OCR) database

Under Decree No. 2013/48 (XI. 15.) NGM, the Ministry for National Economy mandated the installation of online cash registers directly linked to the tax authority. The replacement of cash registers was implemented as part of a gradual process at the end of 2014; subject to certain conditions, taxpayers were permitted to use traditional cash registers until 1 January 2015. The scope of the online cash register system has been expanded significantly since the adoption of the Decree. Initially, the regulation covered retail trade turnover primarily; from 1 January 2017, however, its provisions became applicable to a substantial part of the services sector (e.g. taxi services, hospitality/catering, automotive repair services).

The online cash registers provide the NTCA with itemised data on all receipts issued. For the purposes of our analysis, we used an anonymised database of receipt-level aggregate data. Pursuant to legislation currently in force, retail outlets are not required to issue itemised receipts for each product; they only need to separate products according to collective VAT rate categories. As a result, the itemised breakdown of the database cannot be used for a comprehensive analysis. Besides aggregate data – value, VAT content, payment method, store information – data on the number of items listed on the receipt are also available.

Store information is displayed anonymously through randomly generated identifiers; the only known information about the physical location is the county, while the activity is only marked by the primary, four-digit TEÁOR'08[4] code. Since merchants are not required to request their respective TEÁOR codes based on their main activity, differences cannot be ruled out completely; however, certain specific activities can be identified with a high degree of reliability, such as retail sale of automotive fuel.

1.2. Aggregate characteristics of the database

The online cashier database (OCR) contains the retail data of two years and a total of 7.3 billion unique transactions. The vast majority of transactions are cash; the average value corresponds to the average of the Magyar Nemzeti Bank's (The Central Bank of Hungary – MNB) questionnaire-based payment surveys (Table 1.). There is no accurate statistical data on the amount of cash transactions available, but in 2011 the MNB estimated the approximate number of transactions. Turján et al. (2011) estimated between 2 and 2.8 billion cash payments. The OCR database contains significantly more cash transactions, 3.3 billion for each year. The 704 million card transaction in the database, however, falls short of the 888 million transactions reported to the Central Bank of Hungary. The reason is that the online cashier machines are not required for all retail outlets; a significant part of the service sector is excluded from the legislation. The total value of turnover is about half of the consumption expenditure of households in the national accounts. The remaining half of the expenditure side is mainly related to the service sector and housing costs.

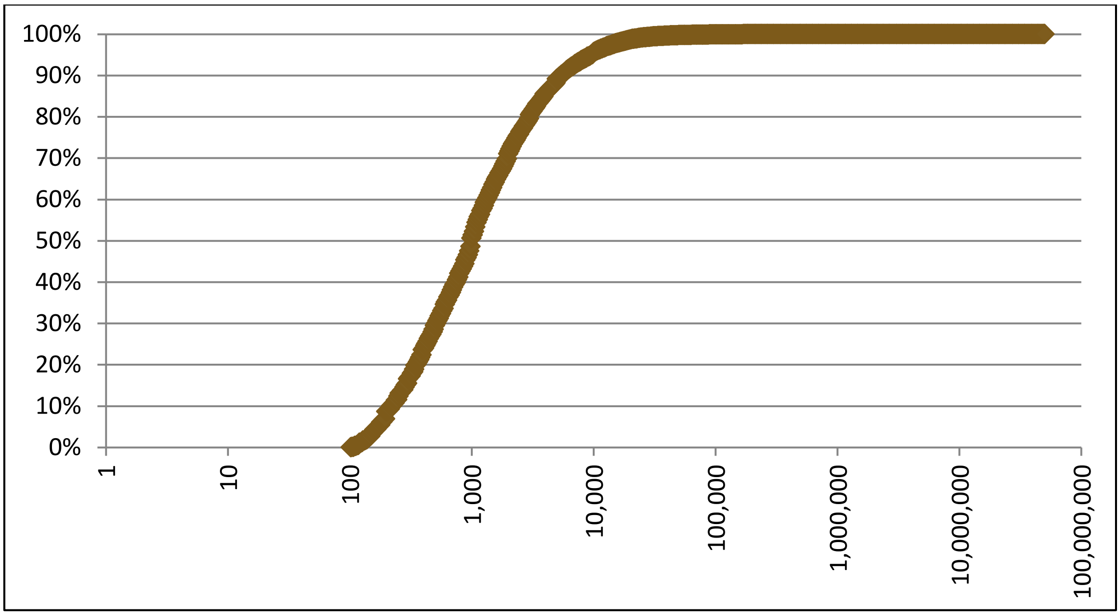
Table 1: Aggregated information for the OCR database 2015-2016

Number of receipts	**7.3 billion**
Gross value	18 915 billion
Net value	15 422 billion
Cash value	13 741 billion
Number of cash transactions	6.5 billion
Card value	4301 billion
Number of card transactions	704 million
Other payment method value	838 billion
Other payment method number of transactions	27 million

Despite the size of the database, it is not suitable for direct comparions with the national accounts. Despite the significant differences in definitions, it can be stated that data is of a magnitude-acceptable nature. The granular, account-level nature of the sample, on the other hand, gives an opportunity to examine the expenditure structure in greater detail.

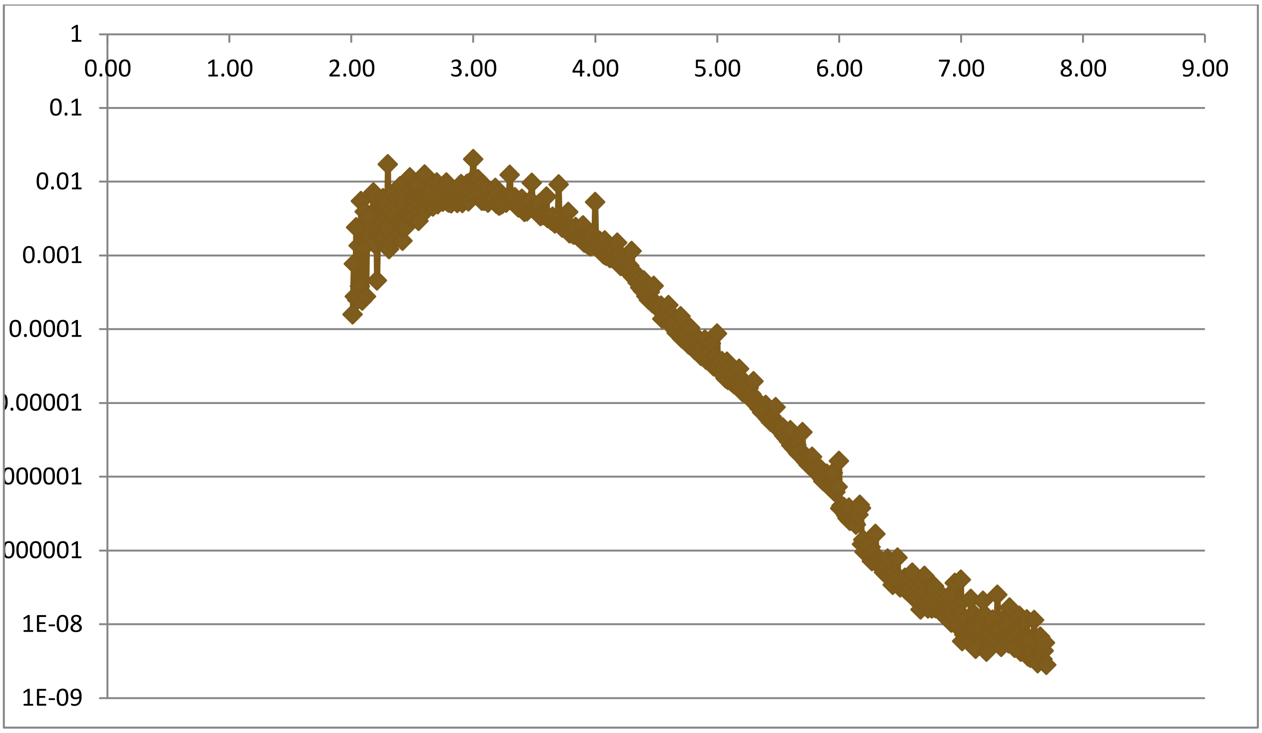
[4] The Hungarian TEÁOR 2008 codes correspond to the European classifications of NACE rev. 2.

Figure 1: Cumulated frequency of transaction by transaction value (HUF)



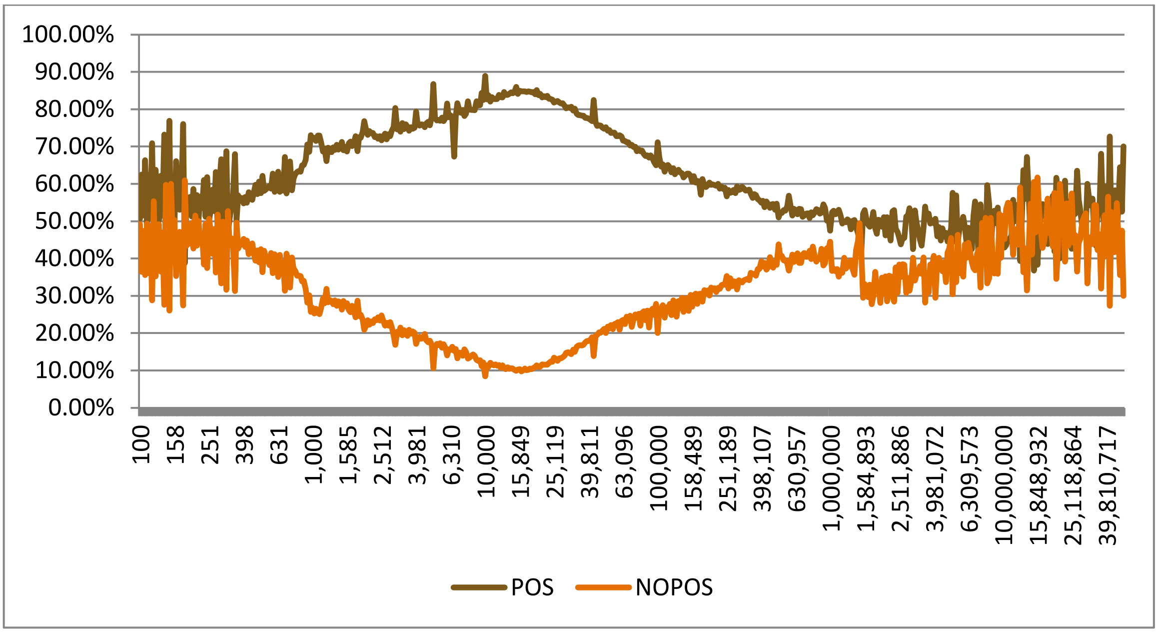
It is important to note that distribution is concentrated strongly towards the smallest values (Figure 1.). Half of the transactions are less than HUF 1,000 and 80% less than 2500 HUF. This result in the fact that in all statistics related to the average, the properties of small transactions dominate. The value distribution of transactions consists of a square and a linear section, represented in a log-log scale. The distribution of low-value transactions is basically lognormal, however, the number of the high-value transaction is significantly higher than in the lognormal distribution and can be better describes by the power rule distribution (Figure 2.). The literature recognizes these types of distributions as a double Pareto-lognormal distribution (Reed, Jorgensen [2004]). The purpose of our study is not to determine the exact mathematical distribution, so we did not have an accurate estimate of the structure of the distribution. However, it is an important feature of the distribution that it basically presents a natural distribution on a logarithmic scale, so the logarithms of the value variables should be included in analyses.

Figure 2: Empirical distribution function of transaction size (log-log scale)



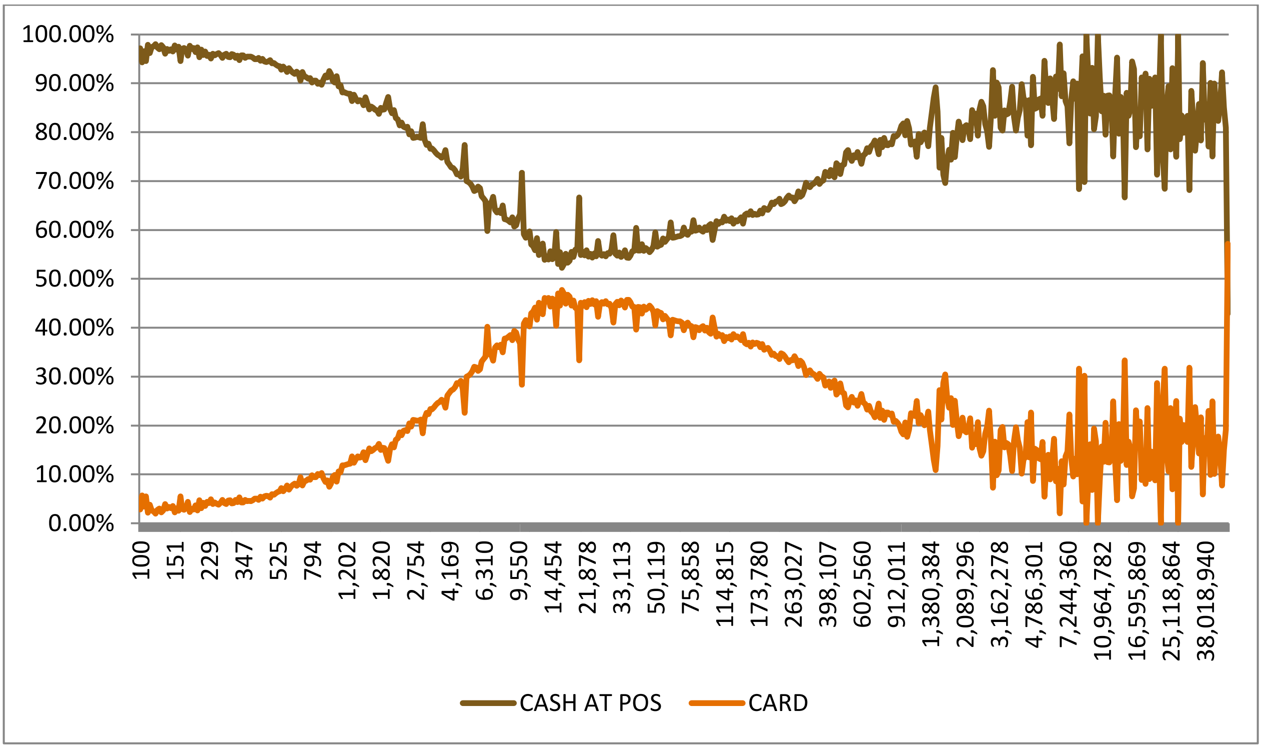
The proportion of card transactions is less than 10% in the total sample, but account for nearly 23% of the total value of payments. However, the actual relevant card usage rate is higher, since a significant part of the transactions are related to businessess where payment cards are not accepted. Card acceptance, however, is not constant for the value of the transactions. Approximately half of small and large transactions take place in a store where there is a card terminal (Figure 3.). For moderate values, however, this ratio may be 70-90%. This phenomenon indicates that there is a significant scope for the spread of card usage in the middle value category.

Figure 3: Share of card accepting store by transaction value (HUF)



This relation can be observed even if the actual card usage is used (Figure 4.). For transaction less than one thousand forints, card utilization is practically negligible, but gradually increases up to 16-17 thousand HUF transactions. Between 16,000 and 50,000 HUF, the card usage rate remains steadily above 40%. However, in transactions above this, the card usage rate is gradually decreasing as the value increases. Based on the data of the two years examined, the breakpoint appears to be at 32,000 forints, however, only for transaction bigger than 50,000 HUF is expected to decrease at a faster rate. It can be stated that this phenomenon can be identified not only at the aggregate level, but also at the level of indicidual stores. The main reason for low card usage in high-value transactions is the large amount of savings in cash that is concentrated in the hands of the household sector. In the domestic denomination structure, the 10 and 20 thousand denominations are highly over-represented, suggesting that a large part of the cash portfolio are part of savings. Because transactions over 50-100 thousand forints are more likely to be linked to the purchase of a high value item, it can be assumed that part of the population uses cash for savings purposes. However, the exact logic of the phenomenon can not be identified solely on the basis of the OCR database.

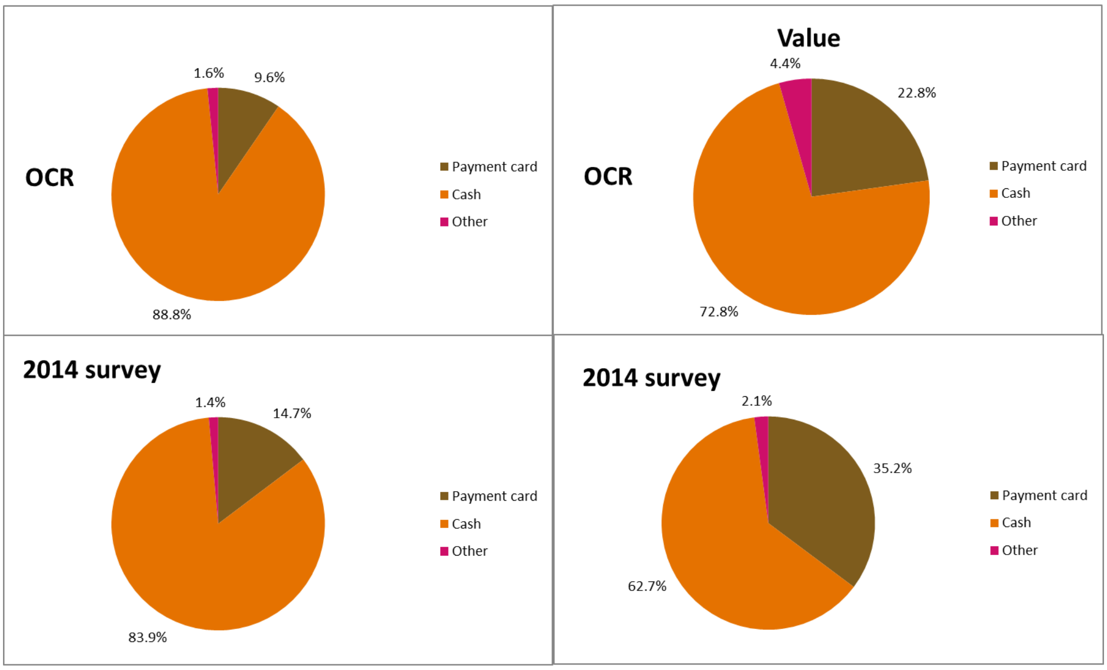
Figure 4: Card transaction share at card accepting stores (HUF)



Estimated data were compared with the results of the 2014 MNB survey (Ilyés-Varga [2015]). During the survey, it was not possible to distinguish the data between the merchant, so we can only compare the total spending rates. It can be stated that the questionnaire survey distorts to some extent both the volume and the value of the cash transaction shares. This is clearly due to the lack of small value transactions in self-assessment surveys. While the average cash payment in the survey was 2500 HUF, the actual value is 2100 HUF based on the OCR database (Figure 5.). The deviation was caused by transactions below HUF 1000, which were underrepresented in the survey. Based on these, the results from the questionnaire and the diary surveys should be treated cautiously and their evaluation should take into account that respondents may not fully remember their low value purchases.

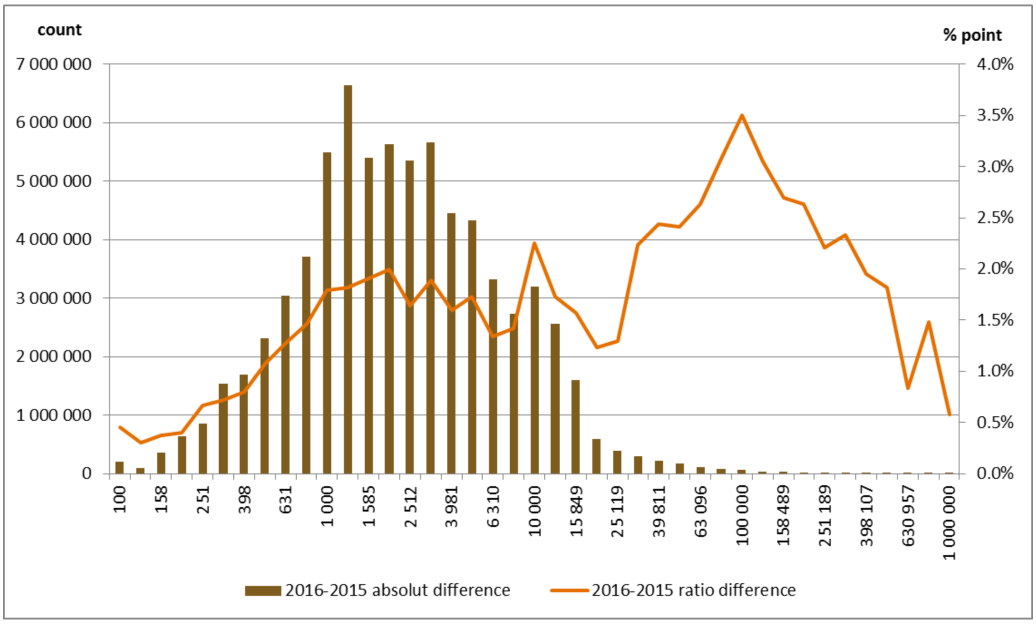
The average value of card payments - 6100 HUF - is lower than the average of HUF 6500-7000 in the MNB's payment statistics, suggesting that mercahnts without online cash have a higher transaction value on average.

Figure 5: Household payment habits based on the OCR and Ilyés-Varga [2015]



The payment patterns household have changed slightly over the two years, so we can explain the causes of the changes in detail. Based on macroeconomic statistics, the number and value of card purchases made at physical store locations over the two years increased by 21-22% (MNB 2017), which is in line with OCR's 23% growth. However, nearly 72 million new transactions are not distributed evenly (Figure 6.). 80% of the increase is given under purchases below 5,000 HUF. If we compare it to the use rate, it can be seen that above HUF 30,000 the percentage point increase was even higher.

Figure 6: Differences between 2015 and 2015 in volume and percentage increase in card usage (log scale)



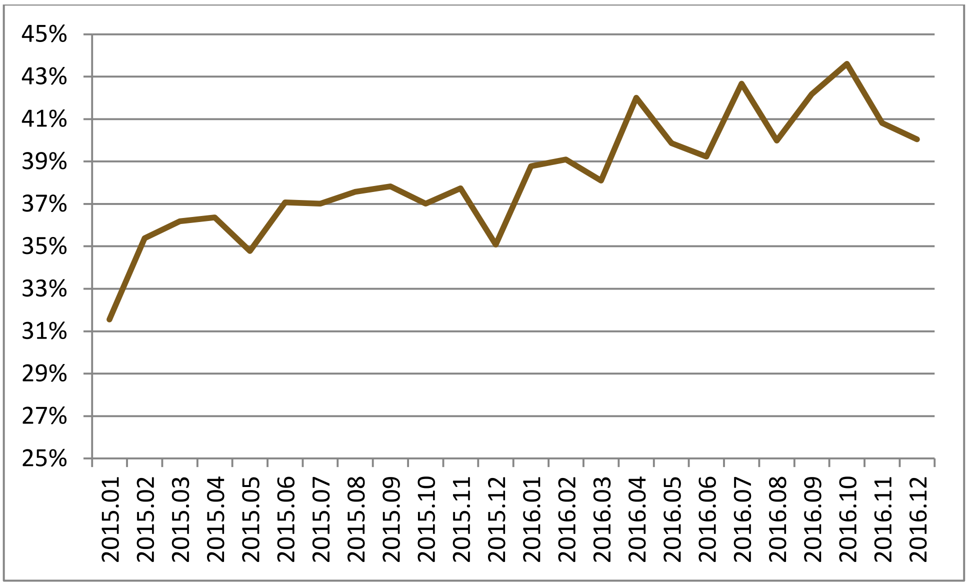
1.3. Store level characteristics

The full online cash register database contains two years of data, so it covers a number of administrative and business events where the number of reported stores has changed. In our sample, we distinguish between 160-180 thousand individual retail stores, but this number varies considerably over the months.

On average, 40% of retail stores accept payment card, however, due to higher rates of card acceptance in larger stores, the card accepting stroes cover 77% the total value of purchases. Based on these, the current Hungarian payment infrastructure would allow for three-quarters of transactions to be executed through card schemes. In reality, however, only 10% of the transaction is a card transaction, although this ratio is constantly increasing.

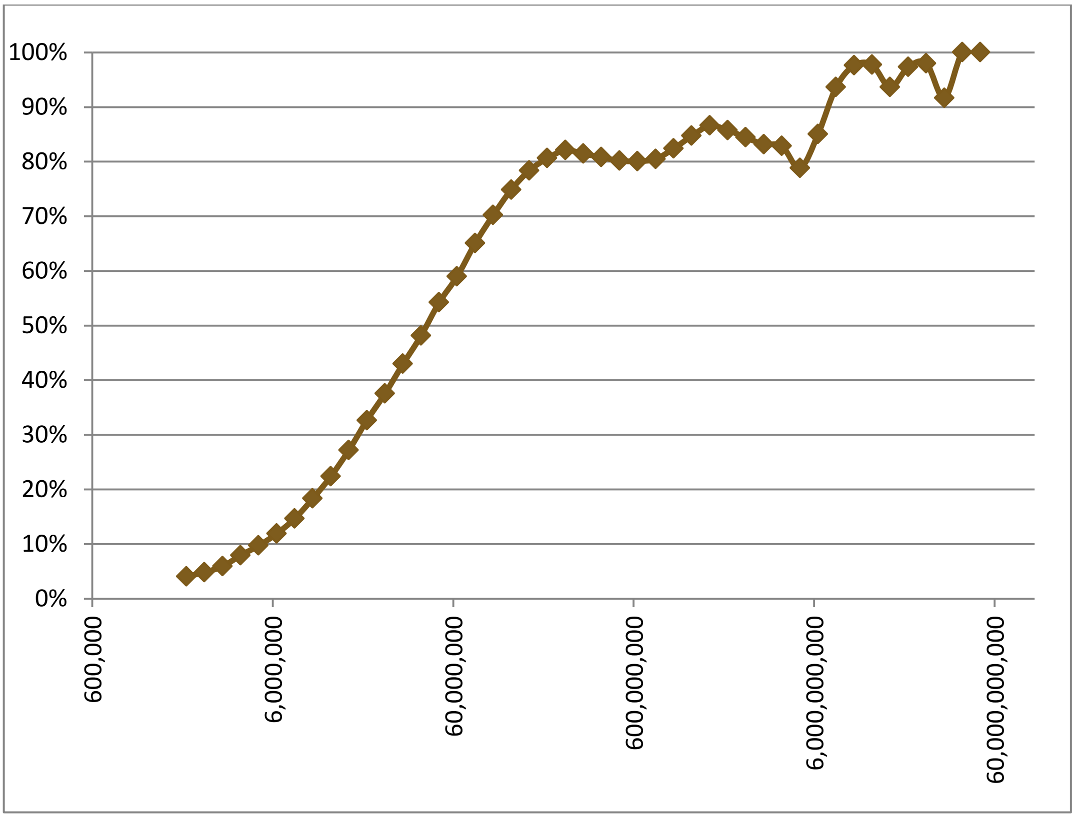
During the two years, the average card acceptance rate gradually increased in the number of stores. Between the start of 2015 and the end of 2016, the proportion of accepting shops increased by an average of 10 percentage points. We can distinguish between 60 and 70 thousand accepting stores in the database, while at the aggregate level there were 75 to 85,000 merchant accepting payment cards in the country during the period under review. The difference is mainly attributable to the service sector and other sectors not covered by online cashier machine legislation. However, the pace and dynamics of growth correspond to the MNB aggregate data (Figure 7.).

Figure 7: Share of card accepting stores in 2015-2016



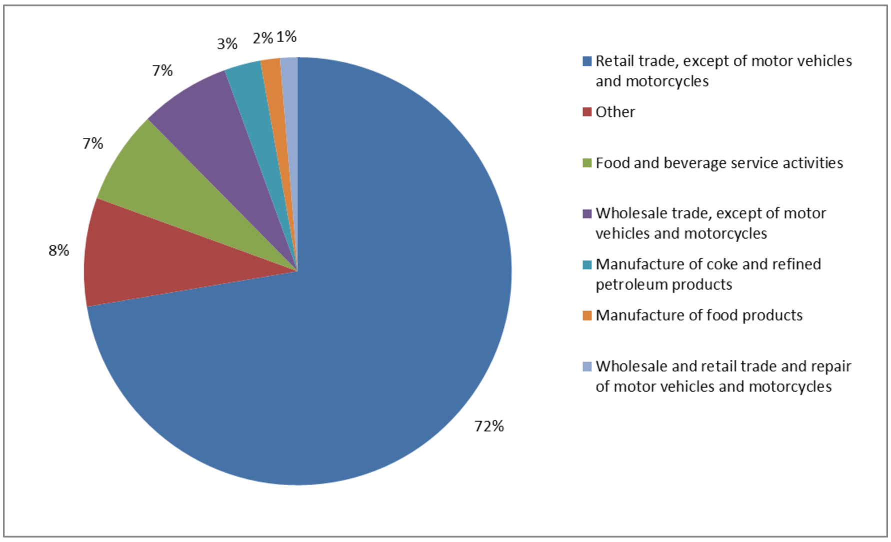
Acceptance rates between stores vary greatly depending on the size of the store (Figure 8.). In small shops, the ratio of accepting stores is negligible. Between the annual turnover of approximately 10 and 100 million forints, the unit percentage change in business size significantly increases the likelihood of card acceptance. However, in large stores, turnonver no longer has a big effecta and acceptance is stable around 80-90%, up to hypermarkets with a turnover of more than 10 billion HUF, where card acceptance is practically complete.

Figure 8: Card accepting ratio by annual revenue (log scale)



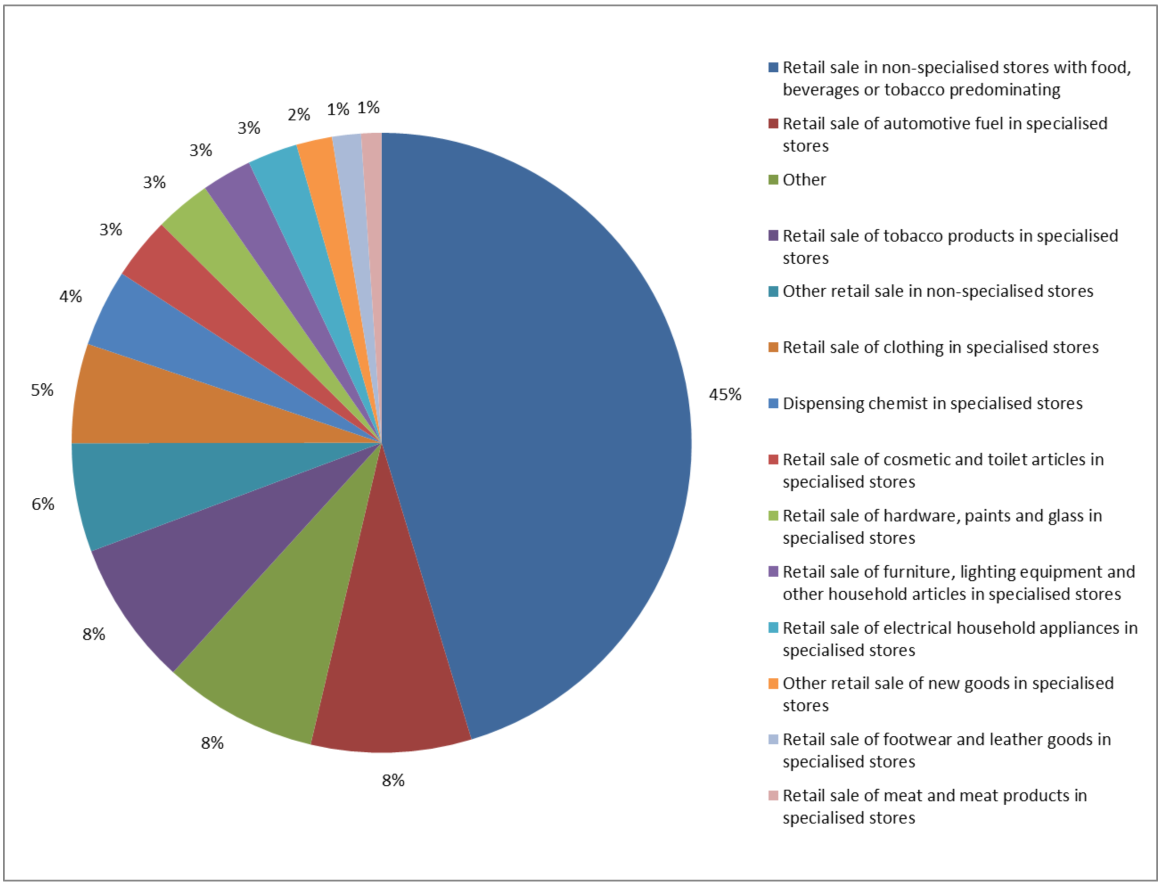
The available TEÁOR codes are incomplete and not exactly accurate, but they are able to show the sectoral distribution of purchases in the OCR database. 72% of turnover is related to the retail sector (Figure 9.), but presumably the actual rate is even higher. The problem of the primary TEÁOR codes based on self-declaration is well illustrated by the 3% share of coke and crude oil processing, which is most likely to be part of the retail fuel trade and not the direct coke purchase of the population.

Figure 9: Expenditure structure in the OCR database



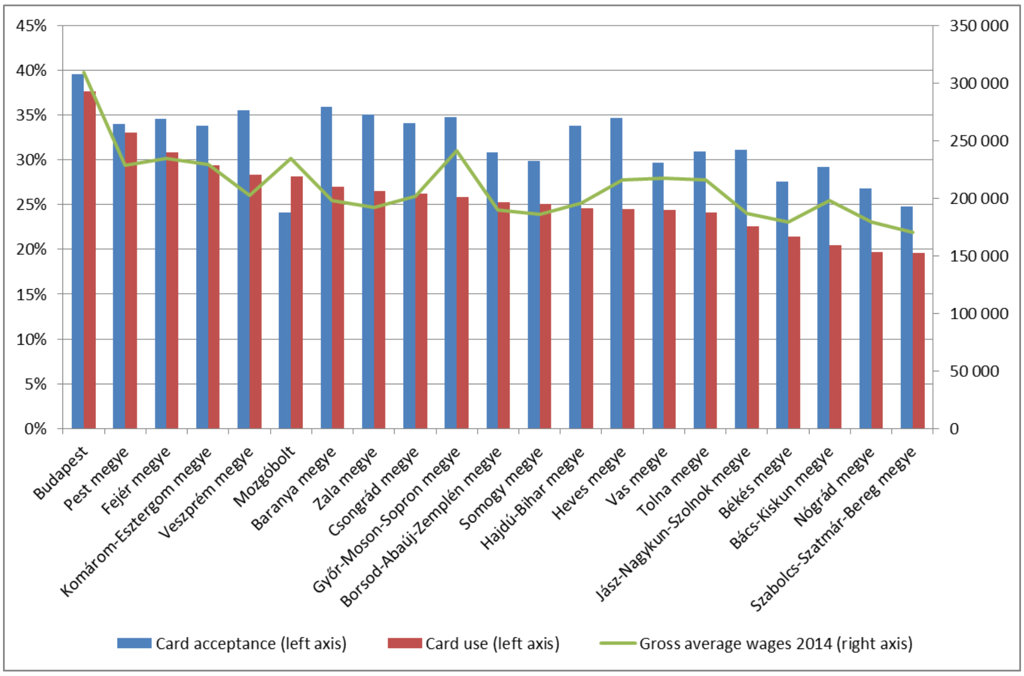
The breakdown of the retail main category is also heavily concentrated, with almost half of the turnover being attributable to the mixed retail trade (Figure 10.). The remainder is evenly distributed between a few larger and several smaller sub-categories. For the examination of the TEÁOR codes, it can be said that only a dozen of the 500 categories listed above exceed the one percent share.

Figure 10: Expenditure distribution in retail purchases



Among the counties there are large differences in the use of the card. In the more developed counties and the capital, card usage is typically higher, while in the east and south it is considerably lower. On the other hand, there is no such strong connection to card acceptance (Figure 11.). It is important to point out that the apparent correlation shown here does not necessarily represent the actual effect, and the separation of the county-level effect from other factors requires more complex regression analysis.

Figure 11: Card acceptance and card usage ratio by counties



(For stores without fixed location, the national average was included as the gross income Source: HCSO)

2. RETAIL CARD ACCEPTANCE

2.1. Objective

The objective of this section of our study was to explore the aspects considered in the card acceptance decisions of retail merchants and to provide an exact estimate of their discrete effects. Given the broad range of businesses, no analysis has been produced so far that examines the bases of card acceptance decisions across the entire retail sector. Since neither the payment service providers, nor card companies have a database that also covers "cash only" merchants, all previous analyses of this kind relied on questionnaire-based surveys. However, the Hungarian online cash register database made available to us by the National Tax and Customs Administration (NTCA) allowed us to inspect the entire retail business from the aspect of payment card acceptance. Thanks to the large sample size, we were able to reliably identify even narrow segments and negligible effects. Our key research questions are the following:

- The literature points out that the single most important factor motivating card acceptance is the yearly revenue of the merchant. Can we confirm this result in the Hungarian retail sector?
- Does the expected card usage share determine the card acceptance decision in Hungary?
- Do other factors have a significant effect on card acceptance decisions?
- Do small, mid-sized or large stores face a similar decision situation upon making their card acceptance decisions?
- Is there a significant difference between independent shopkeepers and chain merchants in terms of their card acceptance decisions? In network decisions, is it the size of the network or rather individual store sizes that affect card acceptance the most?

The secondary objective of our analysis was to set up reliable models that can strip out the selection effect in future data analyses designed to explain customers' card usage decisions. The models presented in our study provide the Heckman selection correction terms in card usage analys.

In the first section of our paper we review the relevant international literature on card acceptance, and describe, in addition to the online cash register database constituting the basis of our analysis, the data available. Some of the variables used in our analysis were derived directly from receipt-level data, while others were used as proxies.

In Section 3 we define the methodology of the analysis and the variables used, and expound on the method applied in determining our sub-samples. We proceed to provide a detailed description of our findings in Section 4, and check the robustness of the results from various perspectives.

2.2. Literature review

The card acceptance is primarily a theoretical field in payment studies. Most studies focus on the effect of the interchange fee on card acceptance, and the calculation of the equilibrium, competitive fee level on the oligopolistic market of card issuing. In one of the first analysis in this field, Baxter (1983) argues in favour of the interchange fees to achieve a higher level of card acceptance and use. However his model received criticism form Rochet and Tirol (2003) and Wright (2003), who significantly updated the model, but still concluded that without surcharge the interchange fee has a neutral effect on the market. In Rochet and Tirol (2007) they created an empirical test, called the tourist test, to calculate an equilibrium fee level. Based on this test (Keszy-Harmath et el. (2012)) concluded that in the Hungarian market the fee is above desired levels, which resulted in an legislative cap in 2013. These theoretical studies however provide little guidance to analys card acceptance in cross samples, because in the abstract and simplified models the merchants usually only differ in unit acceptance costs.

In line with the theoretical studies considerable part of the empirical literature focuses also on the costs of card acceptance (*Humphrey et al.* 2003) and (Turján et al. (2010)). Our empirical study primarily draws from the results of questionnaire-based surveys. *Jonker* (2011) explored card acceptance and surcharging using survey data collected among 1,008 Dutch merchants. The results of the author's regression analysis revealed that, while the merchant's revenue and the number of employees are significant explanatory variables, the cost of card payments also influences card acceptance. *Arango and Taylor* (2008) investigated card acceptance decisions in the Canadian market primarily focusing on merchant perceptions, whereas *Polasik and Fiszeder* (2014) studied the payment method acceptance decisions of online shops. The lion's share of empirical studies, however, concentrates on consumers' card usage rather than the supply side [*Bolt* (2008), *Bolt* (2010) and *Borzekowski* (2006)].

In our study we aim to tests the empirical results of the literature, first of all the effect of the size of the merchant – yearly revenue in our analysis – and other factors on card acceptance. Unlike in surveys we also have a significant sample size to test the card acceptance decision on different subsamples.

2.3. Methodology

2.3.1. Data source: online cash register (OCR) database

In the part of our paper we use the OCR database described in the first section aggregated to get store level data. Owing to the annuality of the database, the group of merchants under review changed during the period; some stores switched ownership, while others operated on a temporary basis. On several occasions, the taxpayer's activity was modified. This, combined with potential data errors, prevented us from identifying some cases in the database where the operation of the store remained the same even though changes had been reported in the relevant administrative data. As a result, the number of stores included in the analyses exceeds that of the online cash registers installed in Hungary and the database in its current form cannot be used for panel econometric purposes.

This problem occurs on a monthly basis; within the month, however, both the actual number of stores and the links within a network can be identified with a high degree of certainty. We corrected this anomaly by segregating the database monthly and created subsamples for every month. In this way every store has 24 versions in the database. If the identification anomaly correlates with store size, this approach makes sure that in the final database the distribution does not change. In any other case, for example if the bigger stores would be easier to follow between months, the raw dataset would have a higher percentage of smaller stores than in reality.

With the monthly subsamples we have two options to estimate the regressions. In the first case we estimate different regressions for every month. With this method we have 24 regressions, but for practical purposes we only put a representative month to the coefficient tables in the paper. In every case there is no significant difference between the monthly regressions. The second option is to estimate one model on the entire database. We differentiate the different subsamples with a dummy variable, but the marginal effects of the predictors will be the same. In our analysis we primarily follow the first approach, however for robustness checks we also compare the results with the aggregated regression.

2.3.2. Steps for developing our model

The objectives of our analysis can be separeted to three independent parts. The first research question is the effect of the different predictor variables on card acceptance propability. The second is, if the card acceptance decision is significantly different for small, middle of big stores, and third is, if it is different when the decision to accept or not accept card is made on network level.

In the first case, as described in the previous chapter, two models can be estimated by the fact that the monthly data is included in a joint or in a separate regression. In our analysis, we present a model with only annual revenue and its orthogonal polynomials, and a complex model with all significant explanatory variables.

For the second question, we first introduced the interaction term between the main explanatory variables and the annual revenue. In the second step, the sample was split into three size categories based on the annual revenue (Annex 1). The cutting points were determined by a simple decision tree model in an endogenous manner. The smallest group includes shops with annual sales of less than 15 million HUF. The revenue of these small stores can only cover one or two people's labor costs. For medium-sized businesses we consider the annual revenues of HUF 15 to 150 million. Based on cross-sectional data, the marginal impact of annual revenue on card acceptance is the highest in this category. For larger stores, card acceptance is steadily high and does not depend on the size of the business based on the descriptive analysis.

Three subcategories were distinguished for network level decisions. Individual stores were separated. The networks of stores – stores sharing a tax identifier - were regarded as network-level decision for card acceptance, if card acceptance or non-acceptance affects more than 95 percent of stores. In the remaining stores, it is presumably a unique decision to accept the card regardless of whether they belong to a network. The three subcategories were included as dummy variables in the main regression, but in the second step we examined whether, if we estimate the three subcategories regression differently, different results are obtained.

In our study we analysed the following regression models:

- A model with only yearly revenue as a predictor
    - Aggregated on two years of data
    - In monthly subsamples
- All significant predictors, full model
    - Aggregated on two years of data
    - In monthly subsamples
- Full model with cross variables
    - Aggregated on two years of data
    - In monthly subsamples
- Subsamples created using network-independent characteristics with all predictors
    - Independent stores
    - Network stores
        - Decision-independent stores
        - Network scale decision stores
- Subsamples created using size variables with all predictors
    - Small merchants
    - Middle sized merchants
    - Big merchants

Due to the complexity of the models and the high multicollinearity between the variables, the direct comparison is not obvious in the case of a binary dependent variable. In our analysis two models are considered similar if their explanatory power and the generated classificiation are the same on the other sub-samples as well. For example, the regressions estimated on the sample of small and large stores may differ in coefficients, but if one model achieves the same good explanation for the other sample, it can be assumed that the differences are due to multicollinearity between the predictors.

2.3.3. Variables in the model

Dependent variable

In line with our research question, the primary dependent variable is card acceptance. A merchant or a store is considered to be a point of sale when payment card transactions are linked to it in the database. Since payment information is often entered manually in the cash register, some transactions might be erroneous. For the purposes of our analyses, we selected 0.5% as the lower margin of error.

Company size

In our analysis, store size is the most important and most decisive explanatory variable. As we have no external information on the store, annual turnover is derived from the sum of the relevant receipts. Although this rough time series has good mathematical attributes – a lognormal-exponential distribution –, owing to the identification problems mentioned above it may cause bias. Since in some cases a single business may be included more than once (due to store information modifications), it would appear in the database as several, small-turnover stores.

Therefore, we use annualised turnover calculated on the basis of actual turnover and opening days. The review period – 2015–2016 – includes the mandatory Sunday closure as well as the period following the revocation of the regulation (the provision on the repeal was announced on 15 April 2016). The projection base, therefore, is not identical in the two years concerned; we define the proportion in such a way that the modes of the two size distributions overlap.

There is a strong correlation between store size and card acceptance but it is non-linear; therefore, complex functional forms are required to ensure good explanatory power. In the models we include the log of the yearly revenue and several of its higher degree orthogonal polinomials. In the case of the models segmented by store size, the sample selection itself increases the complexity of the functional form further.

Value categories

As we could see at the descriptive attributes, the willingness to accept payment cards strongly depends on payment value. Presumably, therefore, in the case of stores with the same annual turnover actual card usage is likely to be higher in businesses where the majority of transactions fall into the appropriate value category as opposed to the stores whose turnover, for the most part, comprises mainly very small-value or very large-value transactions. With that in mind, turnover was broken down according to value categories as follows:

Table 2: Card usage by value category

Value category	Average card usage in 2015-2016
transactions below HUF 1,000	5.0%
transactions of HUF 1,000 – HUF 5,000	15.1%
transactions of HUF 5,000 – HUF 10,000	27.7%
transactions of HUF 10,000 – HUF 20,000	37.0%
transactions above HUF 20,000	29.6%

As regards the turnover structure, we can examine absolute and relative turnover separately in each individual category. In the case of ratios, the benchmark category is always the highest value category. Due to the nature of the relationship, given the limited number of explanatory variables, the final models include the turnover's log and its square.

Temporal attributes of the store

Not only the annualised turnover of the stores, but also the turnover's monthly and weekly distribution can be established based on the dates indicated on the receipts. Accordingly, in our analysis we also studied the effect of the weekly turnover structure on card acceptance. For the most part during the two years under review, the decree on Sunday store closure was in effect in the retail sector. Family-owned stores represented the main exceptions. Consequently, Sunday opening hours can be used as a proxy for ownership. Since the correspondence is imperfect, this variable is included in conjunction with the TEÁOR variable in the models. In this way, we can separate the effects of individual sectorial exceptions from the attributes of the owner.

Since the store's closure on Mondays and Tuesdays proved to have a significant explanatory power in our analysis, this serves as the control variable in the rest of the models. These attributes are linked to special stores – e.g. museum gift shops, sample stores – where the business is not considered to be an independent financial unit.

Network attributes

A large part of the retail sector operates in the form of a network; in other words, numerous outlets are operated by a single legal entity. According to our hypothesis, the fact that the store is part of a chain affects card acceptance decisions in two ways. In networks were each member of the network belongs to the same category – it accepts or does not accept card payments – card acceptance is presumably based on a network-level decision; therefore, the decision situation itself may differ from that of independent stores. By contrast, in networks where, according to the observations, card acceptance is based on the independent decision of the store, the decision situation is determined by the store's unique characteristics. Therefore, our models were also designed to examine the effect of decomposing the sample into three parts – independent store, independent decision, network decision –; moreover, in the case of network stores, we also included the network's total turnover and the number of stores included in the network. According to the cross-sectional analyses, the correlation is non-linear; therefore, we also include the squared terms in the regressions.

Activity

The NTCA database includes the four-digit TEÁOR identifier of the stores' primary activity. Due to the nature of the sample, nearly three thirds of the stores belong to the narrowly interpreted retail sector. In several cases during the modelling, estimating the detailed breakdown is problematic and cannot even be performed completely – for example, where certain secondary activities only involve stores accepting or not accepting cards – or the large number of dummy variables poses obstacles to the estimation of the model. Because of this, we only use the first three digits of the identifier for the majority of our models. The only exceptions are decision tree and decision forest models, where this phenomenon does not present a technical problem.

County code

To ensure the anonymity of the stores, the explanatory variables do not include the precise physical location, only the county identifier. Unfortunately, this restricts the examination of stores that have a different customer base significantly, as we could only distinguish between 21 different types. In consideration of this, the models do not include customer base information, only the dummy variables of the county codes and the capital city. In the second step, we explain the coefficients of the dummy variables with the aggregate data of the given county.

Item number

The database includes the number of products purchased under each receipt. This allowed us, on the one hand, to use the total item number of the store as another approach to the size variable and to introduce average and maximum item numbers. The average and the maximum item number presumably correlates strongly with the payment time and as such, it is used as the proxy variable of the latter. We used average payment value as the control variable in several cases; however, this variable correlates extremely strongly with the decomposition of the turnover by value and with the proportions of the ranges.

The main statistical characteristics of the predictor variables can be found in the appendix of our paper (Annex 2.).

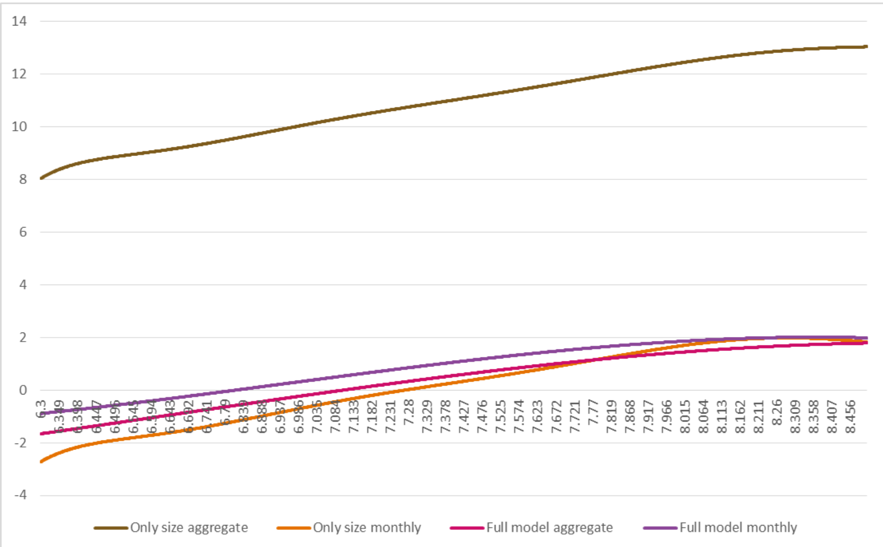
2.4. Results

As stated in the objective, our first question is how mcuh the annual turnover explains card acceptance in Hungary. According to international literature, the size of the store is the most important explanatory variable for card acceptance decisions. Then we examine if the different valued transactions have a significantly different marginal. In the second step, we analyze the impact of other factors - industry, location – and difference between the models estimated on the subsamples of size and network decision

2.4.1. Store size effects

Based on the results (Annex 3), it can be stated that only on the basis of annual turnover the card accepting and non-accepting stores can be distinguished quite well from each other. That is, in the introduction of card acceptance is the most important factor in the decision is the store's annual revenue. The high significance of even the higher degree polynomials suggests that the logistic function is not able to explain the functional relation well. The main reason for this is that the size effect is not linear (Figure 12). For stores over 150 million annual revenue, the marginal impact of the yearly revenue is significantly reduced. This result confirms the phenomenon observed in the descriptive characteristics.

Figure 12: The effect of yearly revenue (log) on card acceptance



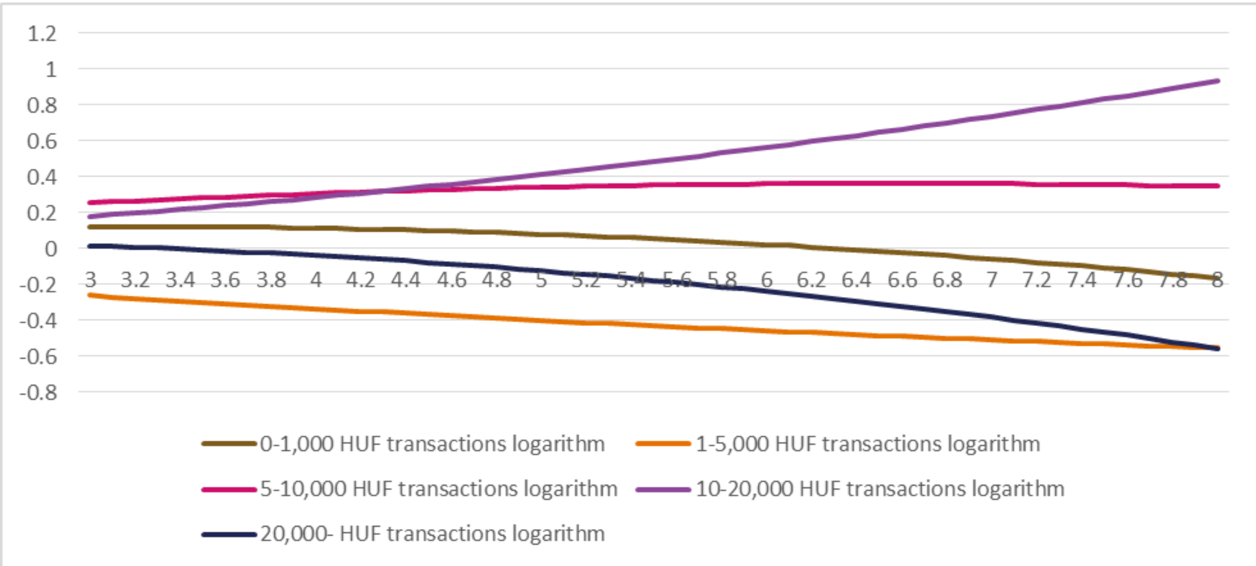
(In the case of logistic regressions, the effect of coefficients can be interpreted directly in two ways: the individual effect on the scale of explanatory variables before logistic transformation and the odd-ratio multiplier on the scale ofdependent variabl. For easier interpretation of polynomial function forms, the

results are presented in the first version, so they can be compared in the order of magnitude by the different variables.)

(Vertical axis: calculated effect on the scale of explanatory variables, horizontal axis: logarithm of annual turnover)

The effect of annual turnover with different value categories is not substantially altered, the explanatory power of the model does not significantly improve (Annex 4). However, the different categories have a slightly different impact on card acceptance. Shops where cards are accepted are more likely to have transactions with a higher card-use ratio. Transactions between HUF 10 and HUF 20,000 have the strongest effect, while transactions with a small proportion of card usage can slightly lower the chance of card acceptance (Figure 13). However, the overall effect in magnitude is less than the impact of annual total turnover. Based on these, it can be concluded that although the breakdown by value category is statistically significant their impact is not significant; the total annual turnover is a good explanatory variable for card acceptance in Hungary.

Figure 13: The effect of value categories in the regression



(Vertical axis: calculated effect on the scale of explanatory variables, horizontal axis: logarithm of annual turnover of the given transaction values)

2.4.2. Effect of other factors

Logistic regression has been extended with a number of other explanatory variables, but the explanatory power of the model increases only slightly. Due to the large number of records in our sample, in our analysis, we can identify small effects that are statistically significant but do not significantly affect the card acceptance decision. Three of the dummy variables can be highlighted, which have a relatively greater effect on the dependent variable. If the store is open on Sunday it considerably lowers the likelihood of card acceptance, while detailed billing increases it considerably. The other variables have a small impact similarly to the value categories, and their marginal value corresponds to the expectations. The size of the network and the number of stores increases card acceptance but decreases in marginal effect. By contrast, dummy variables in network-resolved stores are of the lowest value. From this we can conclude that in Hungary several large networks have decided centrally to discard card acceptance. Overall, it can be stated that the effect of other factors is small.

For the purposes of our analyses we only had the county codes of the store locations available; therefore, instead of including county information directly in the models we only used a county dummy. In the second step we examined the extent to which the coefficients of the dummy variables exhibited a co-movement with the available county-level statistics. Including the capital city, there are 20 county codes in total; therefore, in view of the small sample size we did not estimate a regression and only examined the linear correlation.

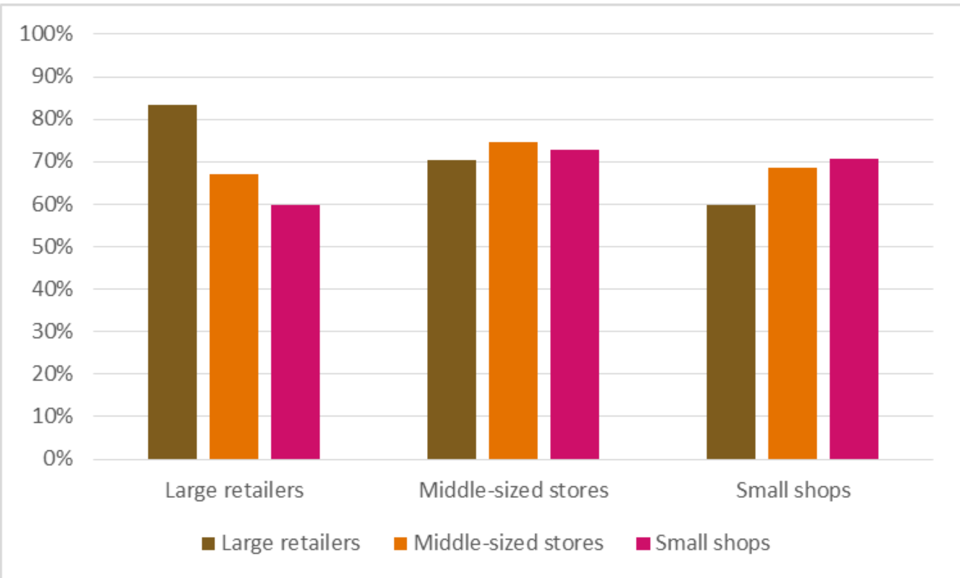
Among the variables under review, the percentage of the working age population, the number of municipalities and the number of residents per shopping centre indicate a medium-strong correlation. Contrary to expectations, development and income variables did not prove to be significant at any level. The correlation does not improve even after the omission of the capital city's outstanding, outlier data point. We must conclude that the composition of the customer base has no affect on card acceptance at the county level. If there is any correlation at all, only a deeper – presumably municipality-level – segmentation would be able to identify it.

2.4.3. Effect of subsamples

In the regression we included the interaction term between the main factors and annual turnover, which, without exception, became statistically significant. Based on these, the marginal effect differs considerably between shops of different sizes. The dummies formed according to the network decision are also very significant in the model. To further investigate our research question, we cut the sample into small, medium and large stores and reestimated regressions as described in the previous chapter (Figure 14 and Annex 5).

The three estimated regression performs moderately well in their sub-samples, but the explanatory power is significantly deteriorating in the case of large stores. That is, the estimated regression in small and medium-sized businesses can not be used for large stores. Based on the analysis of the parameters it can be stated that mainly the effects of the value categories are different. While small to mid-sized stores, broken down by value categories, has a small impact as described in the previous chapter, the breakdown greatly improves the accuracy of the model in the case of large stores. In summary, it can be stated that, by size breakdown, the largest categories of stores are clearly to be treated separately. For high revenue shops, card acceptance can be explained by other observed factors.

Figure 14: Fitting characteristics by size subsamples (AUC)

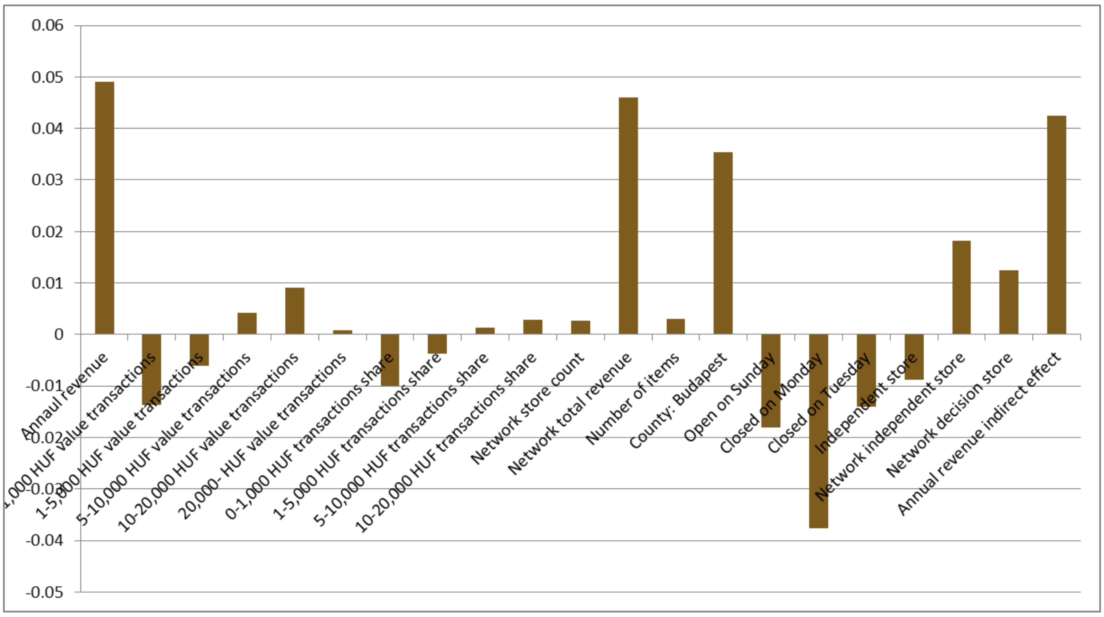


The sample can be split according to the type of network decision as well (Figure 15 and Annex 5). According to this grouping, significantly greater differences can be observed in the explanatory power of the models. It can be clearly stated that in non-networked stores, card accepting and non-accepting stores can be distinguished with other factors than in stores which are part of a network. Based on the models, it can be stated that the size of the network significantly explains card acceptance at the appropriate stores and network stores cannot be estimated in the same model as individual stores.

Figure 15: Fitting characteristics by network decision subsamples (AUC)

The breadth of the parameters of the logistic regression cannot be interpreted directly. Therefore, a simulation was used to analyse the effect of the different variables. We prepared new estimates based on the model used to run the model in such a way that we increased the value of each variable one by one by a total of 10%, with all other variables remaining unchanged. In the case of the dummy variables we replaced the variables with the higher value and the county variable was Budapest for all stores. The results of each simulation are shown by Figure 16. Turnover had the greatest impact both directly and indirectly through the cross-products. The coefficients of network turnover were comparable in magnitude. Of the dummy variables, Sunday and Monday closure significantly reduced the probability of card acceptance, while among the county variables, the effect of Budapest was outstanding.

Figure 16: Effect of the explanatory variables in the simulation



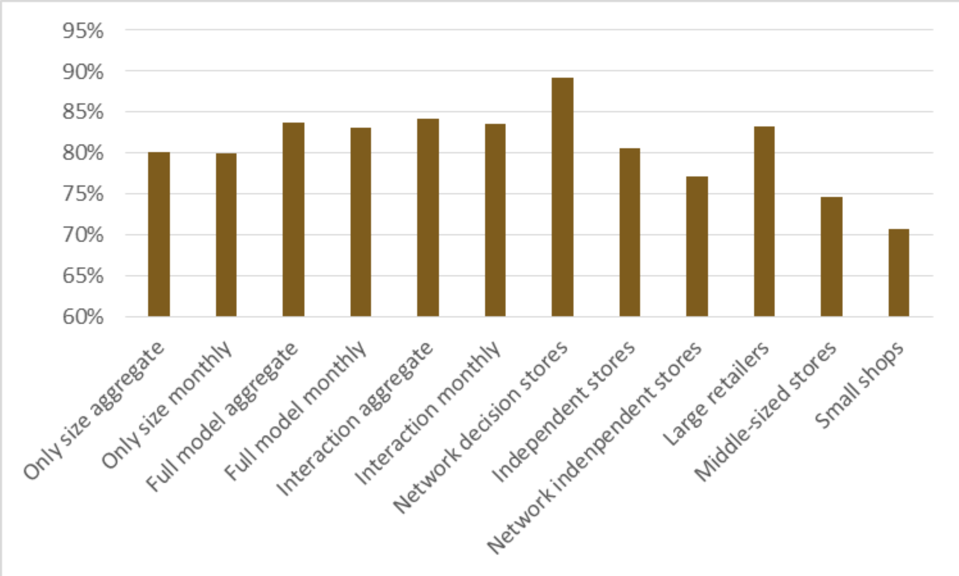
2.5. Robustness analysis

In view of the extremely large sample size, the findings described in the previous section can be considered reliable; nevertheless, we tested the statistical robustness of the findings with several methods.

Since the high degree of multicollinearity between the variables renders the parameter estimates uncertain, we examined the significance and sign of the variable groups in various combinations. The main problem is caused by the fact that although most construed variables exhibit a strong correlation with store size, in some cases this correlation exists by definition – such as turnover broken down by value categories –, while it is empirical in others. To eliminate this discrepancy, we also ran the regressions without the size variables and studied the explanatory power of the rest of the variables. The results confirm that the proportion of the item numbers and value categories strongly correlates with size and takes over the role of size in the restricted model. The explanatory power of the model declines significantly without the direct use of the size variables. This suggests that the remaining variables are unable to take over a significant part of the explanatory power directly.

The explanatory power of the eight tested models can be considered good (Figure 17), even the simplest model can also distinguish card-accepting and non-accepting businesses with great reliability. The inclusion of all variables slightly improves the explanatory power of the model, but it is not improved by splitting into sub-samples. The subdivision breakdown in both cases shows different results in the groups with the smaller number of stores, so the explanatory power of the entire model is not affected significantly.

Figure 17: The fitting of the estimated models (AUC)



The models were re-estimated on 80 percent randomly selectred subsamples, however the results always showed that parameter estimates are stable. We may conclude, overall, that the findings of our analyses are robust; there is no indication of an inadequately chosen functional form or an overfit procedure for the sample.

2.6. Summary

In our study we examined payment card acceptance in the domestic retail sector based on a receipt-level, detailed dataset derived from online cash registers. The main objective of our research was to identify the primary explanatory variables and to test conventional card acceptance hypotheses.

For the purposes of our analysis, we relied on anonymised online cash register data provided by the National Tax and Customs Administration (NTCA) for the years 2015 and 2016. Covering an extremely broad section of the Hungarian retail sector, with nearly 7.5 billion data points the database provides a basis for complex and robust analyses. We tested store-level monthly aggregate data with county and network attributes.

According to our analysis, the introduction of payment card acceptance can be mainly attributed to store size, which was approximated, in our case, with annual revenue. Revenue affects the card acceptance of mid-sized stores the most; the marginal effect is far lower in the case of small-sized and large stores. The results confirmed that the logistic functional form aptly demonstrates the correlation between size and card acceptance.

In addition to its value, even the structure of the revenue influences, albeit to a lesser degree, card acceptance. A store is more likely to accept payment cards if the bulk of its turnover is composed of transactions that have a higher expected card usage rate. It is explicitly among medium-sized and large stores where the effect of the value structure is significant.

The opening hours of the business and, indirectly, its ownership structure, wield a similarly strong influence in card acceptance decisions; in the case of owner-operated stores card acceptance is significantly lower. We found that the income of the customer base does not correlate significantly with card acceptance; however, for a more in-depth analysis of this issue the dataset should be broken down further than the county level.

The decomposition of the retail sector by network type and size does not improve the predictive power of the models significantly, but it has a moderate effect on all other significant variables. Consequently, the hypothesis that different store types face considerably different decision situations does not hold true.

By presenting an analysis that robustly identifies basic correlations across a large-scale sample, we wish to contribute to the existing literature on the choice of payment instruments.

3. RETAIL CARD USAGE

3.1. Objective

The purpose of this section of our study is to examine Hungarian payment habits in retail payment situations. The basis of our analysis is the online cash register (OCR) data. With the assistance of this extremely detailed database that covers a substantial part of the Hungarian retail sector, we can identify the effect of individual factors precisely and gain a deeper insight into domestic retail payment transactions.

Hungarian payment transactions can be considered cash-oriented by European standards: the level of cash in circulation is higher than the European average and the share of electronic transactions in retail payment situations is fairly low. This notwithstanding, Hungarian households have good access to electronic infrastructure; 82.7 per cent of households have a payment account and 80.1 per cent have a payment card. Despite a 15–20 per cent increase in electronic payments over the last few years, the vast majority of transactions are conducted in cash.

In the field of payment research, numerous empirical and theoretical studies have analysed the choice between cash and card payments. However, most of them are questionnaire-based surveys, using payment log data, which gives rise to the problem that respondents may forget about some of their transactions. And the few surveys where receipt-level data collected from merchants are available cover only a limited number of stores. With this paper, we wish to contribute to international payment research by reproducing the results of these surveys on the unique database of online cash registers, and supplement them with several additional variables that could not be examined previously.

Our study is an empirical analysis, in which we assess the extent to which the main theoretical and empirical results of the international payment literature can be confirmed on this data source and explore the reasons behind the features specific to Hungary.

Our key research questions are the following:

- Can we confirm the international literature’s finding, namely, that the main driver of retail consumers’ payment choices is the payment value of the transaction?
- Cross-sectional analyses point to a reversal in the positive relationship between card usage and payment value even in the case of relatively low-value transactions. Does this non-monotonic relationship hold even after having been adjusted for the effects of additional observed variables?
- Since cash payments are still prominent in Hungary, it is important to identify the attributes that encourage the use of cash. In our analysis, we seek to determine the extent to which the effect of the ease of payment can be captured in card usage decisions, for example, through the number of denominations needed.
- Aside from international surveys, Hungarian questionnaire-based analyses also suggest that income and education level are particularly important determinants of payment choices. Can this relationship be established using this database as well, based on OCR data?
- What kind of seasonal, monthly and weekly fluctuations can be observed based in our dataset?

As data collection was primarily driven by tax audit considerations, most of the data contain store information, while apart from the payment instrument, no information is available on the payers’ attributes. As a result of the anonymisation, the physical location of the store is only indicated by county codes. Therefore, this database allows us to examine the effect of the payer’s attributes on card usage to a limited degree only, through socio-demographic characteristics and the differences across counties.

3.2. Literature review

In our research, the online cash register (OCR) database enables us to analyse turnover across a large-scale sample covering a substantial segment of the retail sector. Previous payment studies were typically rooted in questionnaire-based surveys, and the literature offers few examples of payment analyses that cover such a significant volume of data as ours.

The focus of questionnaire-based surveys is the relationship between respondents' socio-demographic characteristics and their payment choices. At the European level, *Cruijsen and Plooij* (2015) compared the results of two Dutch questionnaire-based surveys over a decade-long horizon. Although the use of electronic payment methods is far more intense in the Netherlands, education and age proved to be similarly important explanatory variables. The authors emphasised the role of subjective perceptions – speed and safety – in payment choices. Although the non-linear and non-monotonic relationship described in the cross-sectional analysis of the online cash register database between payment value and card usage intensity was not observed in the Dutch survey, it is important to note that the highest category selected by the authors – above EUR 60 – is still below the Hungarian maximum. Similarly, using US household panel data *Cohen* and *Rysman* (2013) identified transaction size as the most important determinant of payment choice. The study by *Bagnall et al.* (2014) is an important cross-country comparison harmonising questionnaire-based surveys from seven countries: Canada, the United States, Austria, Germany, the Netherlands, France and Australia. The authors' main conclusions are consistent with the results of the Hungarian surveys: card usage increases with higher income and education; the most significant variable is transaction value, while subjective factors also play an important role in all countries considered.

Takács (2010) used data from a 1,000-person questionnaire-based survey to examine Hungarian payment habits. The author found that payment account and card coverage was primarily driven by education and income level. Also based on a 1,000-person questionnaire survey and on payment diary data, *Ilyés* and *Varga* (2015) arrived at similar conclusions; the relationship between socio-demographic variables and card usage habits showed no difference in the two surveys.

Beside questionnaire-based surveys, over the past decade only two surveys have provided an opportunity for the analysis of a large volume of receipt-level data. The first one is a survey conducted by *Klee* (2008) analysing the transaction data of US households. In her survey, the author matched the receipts of 99 retail outlets with demographic information on the local environment of the stores concerned. The main finding of the study is that transaction costs and opportunity costs influence the choice of payment instruments significantly, with transaction value being the most important explanatory variable. *Wolman and Wang* (2014) used transaction-level data from a large US discount chain covering the transactions of a three-year period. In their research paper, the authors presented a detailed analysis of the marginal effects of the individual variables and, with the assistance of the three-year time horizon; they were able to forecast the long-term trends of future card usage. Wolman and Wang (2014) analysed more than two billion transactions in their research. Based on the results presented, neither Klee's (2008), nor Wolman and Wang's (2014) database shows a non-monotonic relationship between cash use and transaction value on the values examined by the authors.

Empirical results show that several theoretical models have been constructed to explain the relationship between transaction value and the card usage rate. *Briglevics and Schuh* (2014) used US payment diary data, while *Huynh et al.* (2014) relied on Canadian and Austrian data to construct their respective decision models. According to transaction value, both models estimate monotonic and concave card usage patterns. While Briglevics and Schuh (2014) described payment instrument choice as a dynamic optimisation problem, Huynh et al. (2014) supplement the Baumol–Tobin model.

Despite the use of receipt-level data, our database differs significantly from the two studies analysing transaction data and from the surveys built on payment diaries in several regards. The database of online cash registers provides national coverage and the vast majority of merchants are subject to the relevant regulation. Accordingly, compared to the studies mentioned above, we were able to distinguish between far more merchants both in terms of size and type. On the other hand, due to the anonymisation, we had little data on the customers of the stores. County identifiers were of limited use as there is scant variance across the counties with regard to the main demographical aspects; consequently, as opposed to Klee (2008), there is no sufficient variance to add a consumer characteristics proxy. However, as opposed to the payment diaries, there is significantly more information available on payment location; moreover, due to the statutory obligations, the reliability of the data is presumably better.

3.3. Data source and methodology

The basis of our analysis is the transaction database of online cash registers. The database contains data for 2015 and 2016 and its records have been processed fully anonymised. We dismissed negative transactions and those exceeding HUF 50 million, but did not apply any filters regarding store size.

The primary objective of the analysis is to explain card usage based on the variables available. We used logit regressions for our estimates and performed the calculations with the assistance of the RevoScaleR module of the R programming language. Due to technical constraints, instead of including geographic data directly in the model, we analysed the coefficients of the county dummy variables in a separate regression.

In order to control for the Heckman selection bias, in the regression we included the inverse Mills ratio computed in accordance with the probit version of the final card acceptance model presented in the previous chapter.

3.3.1. Variables used

Card payment

The main outcome variable of the analysis is the binary variable of card payment. Unlike in theoretical models, in practice payers may use cash and payment cards simultaneously. In the database, the share of cards was 100 per cent in 98 per cent of the card transactions. For the rest of the transactions, the limit of card payment has been defined at a share of 10 per cent.

Transaction value

The database contains the receipt’s gross and net value and its breakdown according to the five VAT rates. Gross value is considered to be the main value of the transaction and in view of the high multicollinearity, we do not use the net value. The share of VAT content is included as a separate variable, excluding the size effect. Since transaction values roughly follow a log-normal distribution, the log of the gross value was also included. In addition, because of the decreasing card usage rate observed for high payment values, we doubled all size variables into values above and below HUF 32,000, which allows the originally monotonic functional form to have an up-sloped and a down-sloped section.

Item number

The number of items purchased was also indicated on the receipts, and the model includes this information as an explanatory variable. Since we do not have direct information on the exact number of items, item number became a proxy variable of purchase size. Based on the non-linear relationship observed by the cross-sectional analysis, we also included the square of the item number in the model.

Ease of payment

The granulated nature of the database provides the means for using such computed variables in the model that can be generated only with a low degree of reliability based on questionnaire and diary based surveys. We approximate the ease of payment by using three variable groups. Firstly, the value variable of all banknote denominations – i.e. the cases where the transaction can be paid directly by one denomination – is used as a dummy variable. Secondly, we also use the number of banknotes and coins handled in the ideal case – as calculated by the descriptive study – as a dummy variable, up to a value of 10. The third group includes total amounts divisible by ten, hundred and thousand. These three groups of variables capture the ease of cash payment, which presumably correlates with payment time and as such, it can be considered to be a cost variable.

Store attributes

Although the model constructed for card acceptance contained numerous variables, due to space limitations, we can only include the most important ones in this study. As regards store attributes, most models include the log and square of annualised turnover and the aggregate form of the activity. Additional variables – opening hours, other size variables – were included in the model only on an ad-hoc basis as a robustness check, but they are excluded from the final model.

Stores that had recently adopted card technology were separated as special cases. In their case, the model uses two additional variables: the date of adoption and the length of time since the adoption.

County data

In the card acceptance model, county effects did not correlate significantly with the county's level of development, but a correlation can be observed during card usage on raw data. We estimated county codes in two steps: the main regression includes only the county dummy variables, while in the second step we focus on the correlation between the coefficients and the main socio-demographic data of individual counties.

Temporal data

The database contains data for a two-year period, which reflect significant monthly and weekly seasonality. Since a sufficient amount of data was available, we included yearly and monthly dummy variables and dummies pertaining to the days of the month and the days of the week.

Inverse Mills ratio

As card usage and card acceptance mutually affect each other, the model calculated by us reflects a significant degree of selection bias. In order to remove the bias, we also included the inverse Mills ratio computed from the probit version of the model constructed for card acceptance. The Heckman selection thus performed reduces estimation uncertainty, especially in the case of the affiliated store data.

The main statistical characteristics of the predictor variables can be found in the appendix of our paper (Annex 6.).

3.4. Results

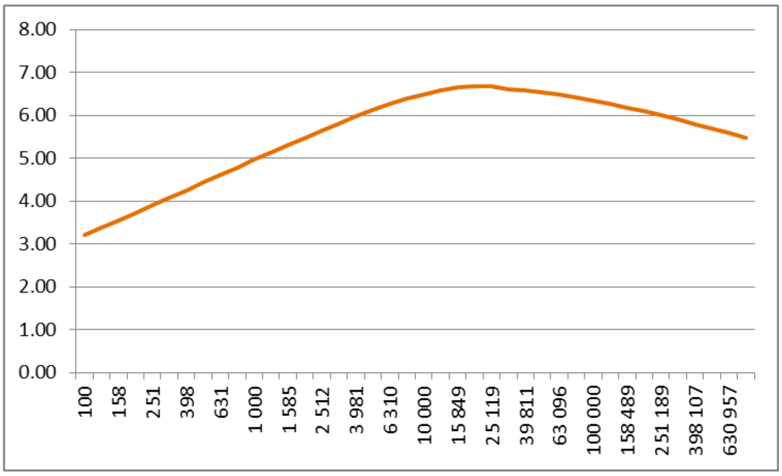
As described in the objective, our first research question is whether the OCR database is consistent with the main results of the international literature. Based on the literature, the main driver of payment choices is transaction value. According to international experience, the relationship is monotonic and concave. Since based on the descriptive characteristics the relationship reverses course above a relatively low value, we examine whether the effect holds even after having been appropriately adjusted. We also analyse the effect of the ease of payment; in other words, whether situations where cash payment requires a fewer number of denominations perceivably influence the decision on card payment. Owing to the limited number of observations, we can only verify the findings of questionnaire surveys indirectly; however, we examine the effect of income and education at the aggregate county level. In the last step, we present the results of the seasonality check.

3.4.1. Effect of the size variables

Our own calculations confirm that, of all the variables, information on transaction value has the greatest explanatory power. The orthogonal polynomials of the logged terms examined are strongly significant in the regression equation (Annex 7). The coefficient of the variables separated above the payment value of HUF 32,000 is significantly different, which suggests that the nature of the relationship is different for low-value and high-value payments. Figure 18 indicates that the marginal effect of the value increases up to HUF 10,000–20,000, whereupon it falls back to the vicinity of zero. In the case of high-value transactions, the share of card payments progressively decreases with the decline in value. Consequently, the downward trend that can be observed with high values is not a selection or model specification problem; it holds even after the proper adjustments have been made.

The final model includes several, higher-degree polynomials of the transaction value concurrently. The joint presence of multiple transformations indicate that, mainly in the case of low and medium-value transactions, the logistic function does not fit perfectly. That notwithstanding, of all functional forms tested, the logit function is still the most suitable functional form to describe the relationship.

Figure 18: Aggregate effect of size variables (HUF log scale)



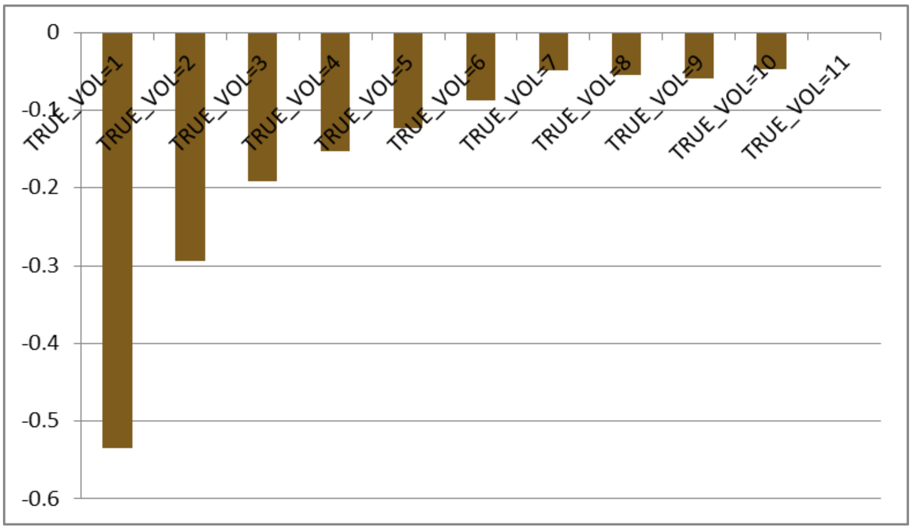
(Vertical axis: transaction size effect on the scale of the predictor variables, horizontal axis: transaction value)

The annual turnover of the store and the number of items purchased also have a significant effect on the card usage rate. As regards store size, the marginal effect increases convexly with the logged annual turnover. By contrast, the relationship between item number and card usage is concave; the card usage rate increases with the item number, but it does so to a diminishing degree. Accordingly, consumers are more likely to use their cards for larger purchases even beyond the value-based ratio.

3.4.2. Effect of the ease of payment

Among the subjective variables, we examined exact denominations, divisibility and ease of payment (Annex 8). The cross-sectional analysis of the database demonstrates that the payer is more likely to pay in cash if the payment value is precisely identical with a denomination value. In the final model, however, the main explanatory variable is the number of denominations required for the payment (Figure 19); banknotes diverge from this trend only slightly, 5,000 in the positive direction and 20,000 in the negative direction. Hence, even beyond the direct ease of payment, a total amount of 20,000 was far more likely to be paid in cash.

Figure 19: Effect of the denominations required for the payment on card usage



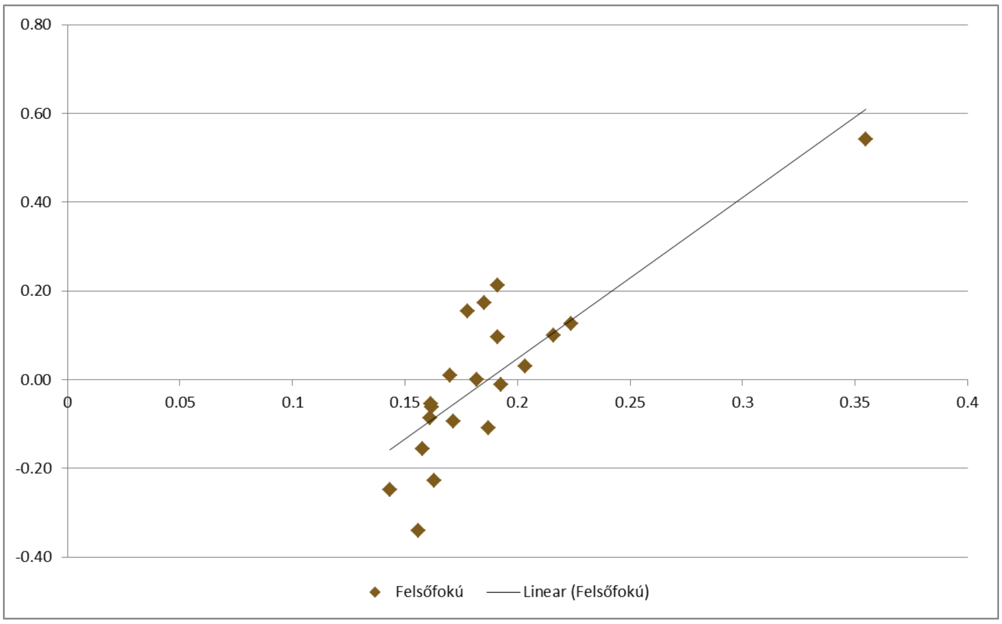
(Vertical axis: denomiation effect on the scale of the predictor variables, horizontal axis: minimum number of denomiations need)

In the case of the divisibility dummy variables, the effect is statistically significant, but its level is only a fraction of the rest of the variables. While the denominations of 10 and 100 increase the likelihood of cash payment, the 1,000 denomination does not influence the payment method. Identifying the causal relationship is not easy, but presumably, a round total amount may point to certain store types simultaneously, which cannot be directly identified by the TEÁOR (Nace Rev. 2) code.

3.4.3. Effects of the payer's characteristics

Unlike in the card acceptance regression, the coefficients of the dummy variables of county effects move closely together with the level of development indicators of the given county (Annex 9). The correlation coefficient of the relationship between gross average earnings and the marginal effect is 0.6, but excluding the outlier value of Budapest, it is 0.8. It can be clearly established, therefore, that an income/demand effect is at play in card payment decisions. Based on Hungarian questionnaire-based surveys, besides income, the explanatory power of education is also considerable. We approximated this with the share of consumers with higher education qualifications (Figure 20). The relationship is even stronger in this case, which is consistent with the result of the questionnaire surveys. 21 observations, however, are not sufficient alone to separate the effects of the various variables – a database aggregated at this level is not suitable for such an exercise. However, there is evidence that the correlations of questionnaire-based surveys also appear in the database of online cash registers.

Figure 20: Relationship between county effects and the share of consumers with higher education qualifications



(Vertical axis: country effects on the scale of the predictor variables, horizontal axis: share of inhabitants with higher education)

Source: own calculation and HCSO

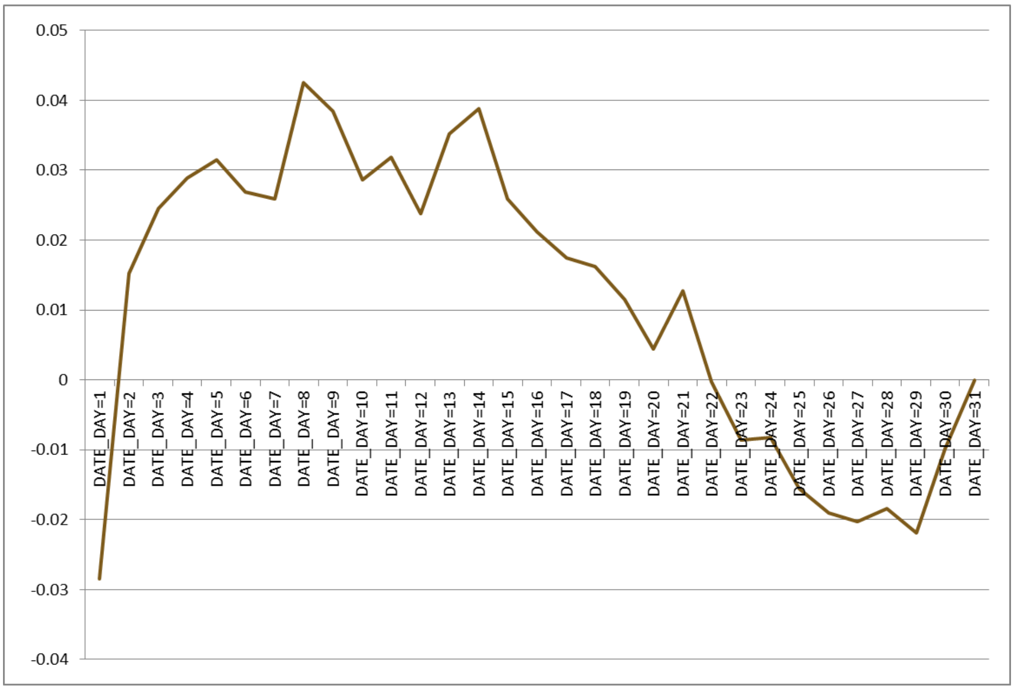
3.4.4. Seasonality

Time variables reveal distinct trends regarding the seasonality of card payments. The difference in level between the two years is significantly positive, testifying to an average increase in card usage (Annex 10). This is consistent with the trend of aggregate official statistical data. The monthly time profile is higher in the autumn and winter periods, whereas spring and summer months exhibit lower card usage levels. While the lower cash usage rate in November and December can be attributed to Christmas shopping, the actual maximum is observed in January, when the same explanation no longer holds. The high cash usage rate of the summer months may be related to the seasonality of tourism. Although we do not have direct data sources available, we assume that foreign tourists are less likely to use cards than domestic consumers.

According to the monthly time profile, households are more likely to use the income transferred to their accounts at the beginning of the months, and then gradually start dipping into their cash reserves (Figure 21). Accordingly, if households draw their income both in the form of bank account transfers and in the form of cash, first and foremost, they deplete their electronic money. The spending structure indicates that the primary form of saving for a significant portion of households is cash, rather than electronic money.

The weekly change in card usage suggests that consumers are more likely to use their cards during the week than at the weekend. The maximum falls on Fridays, whereas the card usage rate on Saturdays is lower than on weekdays on average, and Sundays – presumably also due to the Sunday store closure during the review period – are strongly cash-oriented.

Figure 21: Coefficients of monthly time variables



(Vertical axis: intra month effect on the scale of the predictor variables, horizontal axis: day of the month)

3.5. Robustness analysis

Due to the size of the database and technical constraints, we were only able to check the robustness of the estimated logistic regression to a limited degree. We could only check the value of the AUC indicator[5] used to assess the accuracy of the classification on randomly generated 100 million subsamples. Based on the indicator, the model has medium explanatory power; it cannot separate card and cash transactions to an outstanding degree. A significant degree of variance remains in the model even in case of the inclusion of all variables. Accordingly, these variables alone cannot fully describe the card payment decision.

This finding is consistent with the fundamental problem deriving from the structure of the database, namely, that only indirect, aggregate county data is available with respect to the payer. In light of this, the main finding of the model is that the payment location and transaction characteristics alone cannot fully predict card usage but they do still have a significant explanatory power.

Given the size of the database, variable selection could only be performed in groups. Annex 10 shows the individual variable groups under review and the fit of the relevant models. It is clear that the main explanatory group is transaction value; the rest of the variables make only a limited contribution to the explanatory power of the model. This is because the variables affect only a few transactions in certain cases, which dampens their average explanatory power at an aggregate level – for instance, the divisibility dummy variables cover only a few tens of millions of transactions. Without transaction value, the fit of the model deteriorates significantly and the coefficients of the remaining variables are distorted.

[5] The area under the Receiver Operator Curve

In summary, we found that based on the information on payment location and the nature of the transaction alone, transactions can be classified medium well, and the remaining uncertainty is caused by the lack of sufficient information about the payer.

3.6. Summary

In our study, we analysed the payment characteristics of the transaction-level database of online cash registers. Unique by international standards, our database allowed us to verify, on a robust database, the findings of previous research that typically relies on questionnaire-based data, and to gain deeper insight into the structure of Hungarian payment transactions.

We found that the main findings of the international literature also apply to Hungary. The main explanatory variable in card payment decisions is transaction value. Card usage is negligible in low-value transactions, and the card usage rate gradually increases in line with value up to 100 EUR. Above this threshold, however, the relationship between the card usage rate and transaction value reverses course and embarks on a downward trend. The breaking point of the relationship is too high to be comparable with other international results, as the latter typically involves lower-value transactions. It can be still clearly stated that the downward trend holds even after having been controlled for the main variables. This database is insufficient in itself to explore the reasons, but it can be assumed that these purchases involve a substantial amount of the extensive cash savings typical of the current Hungarian economy.

In our research, we also examined the effect of the ease of payment on card usage and found that payment instrument choices are significantly influenced by the ease of payment. If the value of a transaction can be paid in fewer denominations (in other words, the use of cash is fast and easy), consumers will be significantly less likely to use payment cards. We can conclude, therefore, that convenience is an important consideration in consumers' payment instrument choices.

Even though the county-level aggregation of the data hinders the use of payer information, there is still evidence of the positive effect of income and education on card usage. Accordingly, the results of previous Hungarian questionnaire-based surveys in this regard can be considered robust.

The two-year horizon of the database enabled us to examine the changes in payment habits over a longer term. Our analysis points to strong monthly and intra-month fluctuations. Consumers are most likely to use payment cards during the autumn and winter months, while the card usage rate is higher at the beginning of the month and gradually declines from then on. Based on this, we can conclude that cash is the main form of savings in a substantial part of households, and electronic money is largely used to cover current expenditures.

4. CONCLUSION

The aim of our research was to explore the OCR database from payment perspective and identify the main drivers of payment card accetance and use in Hungary. The database provides us with a uniquely wide and deep insight into the retail payment landscape on a national level.

Our analysis confirms the main findings of the empirical payment literature. In card acceptance and card usage decisions the single most important factor is size. The bigger the annual revenue is, the more likely it is that a merchant accepts payment cards. However this relationship is not linear, for small stores and large retailers the marginal effect is negligible, but for middle-sized merchants the annual revenue has the biggest marginal effect. We find several other factors to be statistically significant but on overall they are not comparable to the effect of store size. Breaking the database into several subsamples, we find that there are significant differences between merchants of different size and chain and independent stores. The merchants of different size categories and types cannot be treated in a same model.

Considering payment choice decision our analysis confirms the previous results of the literature. The transaction size is the most important factor determining payment choice. However we show that the relationship is not monoton, for moderately large transactions it reverses course and card usage start to decline. We cannot provide a clear explanation for the phenomeon solely based on the OCR database, but it is a significant effect even when controlled for other factors. More research is needed to verifiy if this phenomeon is unique to Hungary or is present is other payment systems as well.

Apart from transaction size we find that payment choice is also heavily influenced by the characteristics of the payer. In accordance with the several survey results, we show that the effect of education level and income is present even on the county level. The OCR database also proves that ease of payment is an important factor in payment choice.

Our analysis is only a first look of the main characteristics of payment choice with the help of the OCR database. In the future we can answer several questions related to payment economics. The next step can be to build a dynamic model of payment card acceptance introduction and examine turnover increase and card usage after introduction.

5. REFERENCES

Arango, C., - **Taylor**, V. (2008). *Merchant acceptance, costs, and perceptions of retail payments: a Canadian survey* (No. 2008-12). Bank of Canada Discussion Paper.

John **Bagnall**, David **Bounie**, Kim P. **Huynh**, Anneke **Kosse**, Tobias **Schmidt**, Scott **Schuh**, Helmut **Stix** (2014): *Consumer Cash Usage A Cross-Country Comparison with Payment Diary Survey Data* ECB Working Paper Series No. 1685 / june 2014

Baxter, William P. (1983), *Bank Interchange of Transactional Paper. Legal and Economic Perspectives*. Journal of Law and Economics, vol. 26 no. 3, pp. 541−588.

Wilko **Bolt** (2008): *Consumer Choice and Merchant Acceptance of Payment Media*. DNB Working Papers No. 197 December

Wilko **Bolt**, Nicole **Jonker**, Corry van **Renselaar** (2010): *Incentives at the counter: An empirical analysis of surcharging card payments and payment behaviour in the Netherlands* Journal of Banking & Finance 34 1738–1744

Ron **Borzekowski**, Elizabeth K. **Kiser**, Shaista **Ahmed** (2006): *Consumers' Use of Debit Cards: Patterns, Preferences, and Price Response* Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C. 2006-16

Tamás **Briglevics**, Scott **Schuh** (2014): *This Is What's in Your Wallet... and How You Use It* Working Paper Series no 1684 / june 2014 European Central Bank

Michael **Cohen**, Marc **Rysman** (2013): *Payment Choice with Consumer Panel Data* Federal Reserve Bank of Boston Working Papers No. 13-6

Carin van der **Cruijsen,** Mirjam **Plooij** (2015): *Changing payment patterns at point-of-sale: their drivers* DNB Working Paper No. 471 / April 2015

Kim P. **Huynh**, Philipp **Schmidt-Dengler** and Helmut **Stix** (2014): *The Role of Card Acceptance in the Transaction Demand for Money* Bank of Canada Working Paper/Document de travail 2014-44

D., M. **Humphrey** – T. **Willesson**, Lindblom, and G. **Bergendahl** (2003): *What Does it Cost to Make a Payment*? Review of Network Economics (2): 159–74.

Ilyés, Tamás – **Varga**, Lóránt (2015): *Show me how you pay, I tell you who you are – Sociodemographic factors determinig payment choice*, Hitelintézeti Szemle, 14/2. 2015. June, 26–61. o.

Nicole **Jonker** (2011): *Card acceptance and surcharging: the role of costs and competition*. DNB Working Paper No. 300 May

Elizabeth **Klee** (2008): *How people pay: Evidence from grocery store data* Journal of Monetary Economics 55 526–541, 17 February 2008

Éva **Keszy-Harmath**−Gergely **Kóczán**−Surd **Kováts**−Boris **Martinovic**−Kristóf **Takács** (2012): The role of the interchange fee in card payment systems MNB Occasional Papers 96.2012

Magyar Nemzeti Bank (MNB) (2017): *Payment Statistics*. http://www.mnb.hu/statisztika/statisztikai-adatok-informaciok/adatok-idosorok/xiv-penzforgalmi-adatok/penzforgalmi-tablakeszlet

William J. **Reed** – Murray **Jorgensen** (2004): *The Double Pareto-Lognormal Distribution – A New Parametric Model for Size Distributions*. Communications in Statistics - Theory and Methods 33/8

Michal **Polasik**, Piotr **Fiszeder** (2014): *Factors determining the acceptance of payment methods by online shops in Poland*

Rochet, Jean-Charles – Jean **Tirole** (2003), 'An Economic Analysis of the Determination of Interchange

Fees in Payment Card Systems', Review of Network Economics, vol. 2 no. 2, pp. 69−79.
Rochet, Jean-Charles – Jean **Tirole** (2007), '*Must-take cards and the Tourist Test'*, DNB Working Paper, no. 127, January, De Nederlandsche Bank.
Takács, Kristóf (2011): *A magyar háztartások fizetési szokásai* MNB-Occasional Papers 98. 2011
Turján, Anikó - Éva **Divéki** - Éva **Keszy-Harmath**, Gergely **Kóczán** , Kristóf **Takács** (2011), *Nothing is free − A survey of the social cost of the main payment instruments in Hungary*, MNB Occasional Papers, vol. 93.
Zhu **Wang**– Alexander L. **Wolman** (2014): *Payment Choice and the Future of Currency: Insights from Two Billion Retail Transactions* Federal Reserve Bank of Richmond Working Paper No. 14-09R October
Wright, Julian (2003), *Optimal Card Payment Systems*, European Economic Review, vol. 47 no. 4, pp. 587−612.

6. ANNEXES

Annex 1: Average number of stores in subsamples

Size based subsamples		Type based subsamples	
Small stores	80 373	Network decision stores	67 984
Middle sized stores	94 646	Independent stores	93 598
Large retailers	21 055	Network stores with independent decision	34 493
Sum	196 074	Sum	196 074

Annex 2: The main descriptive characteristics of the variables in the card acceptance model

	Average	Standard deviation	Minimum	Maximum
Annual revenue 1th degree orthogonal polinom	-5,1E-03	6,3E-03	-1,6E-02	2,9E-02
Annual revenue 2nd degree orthogonal polinom	-1,1E-02	5,2E-03	-1,6E-02	4,8E-02
Annual revenue 3rd degree orthogonal polinom	8,2E-03	9,0E-03	-1,6E-02	7,4E-02
0-1 thousand HUF category logarith	6,3E+00	3,0E+00	-2,7E-03	1,6E+01
1-5 thousand HUF category logarith	6,4E+00	2,4E+00	-2,7E-03	1,6E+01
5-10 thousand HUF category logarith	3,9E+00	2,6E+00	-2,7E-03	1,4E+01
10-20 thousand HUF category logarith	2,7E+00	2,6E+00	-2,7E-03	1,3E+01
20- thousand HUF category logarith	2,0E+00	2,4E+00	-2,7E-03	1,3E+01

0-1 thousand HUF category logarith squared	4,9E+01	3,4E+01	0,0E+00	2,5E+02
1-5 thousand HUF category logarith squared	4,7E+01	2,9E+01	0,0E+00	2,6E+02
5-10 thousand HUF category logarith squared	2,2E+01	2,2E+01	0,0E+00	1,9E+02
10-20 thousand HUF category logarith squared	1,4E+01	1,9E+01	0,0E+00	1,8E+02
20- thousand HUF category logarith squared	9,6E+00	1,5E+01	0,0E+00	1,8E+02
Average number of items	1,9E+00	1,5E+00	0,0E+00	3,5E+02
Average transaction value	6,7E+03	5,8E+04	1,0E+00	4,9E+07
0-1 thousand HUF category share	4,3E-01	3,2E-01	0,0E+00	1,0E+00
1-5 thousand HUF category share	3,6E-01	2,2E-01	0,0E+00	1,0E+00
5-10 thousand HUF category share	9,0E-02	1,2E-01	0,0E+00	1,0E+00
10-20 thousand HUF category share	5,6E-02	1,1E-01	0,0E+00	1,0E+00
0-1 thousand HUF category share squared	2,9E-01	3,1E-01	0,0E+00	1,0E+00
1-5 thousand HUF category share squared	1,8E-01	1,8E-01	0,0E+00	1,0E+00
5-10 thousand HUF category share squared	2,3E-02	6,5E-02	0,0E+00	1,0E+00
10-20 thousand HUF category share squared	1,5E-02	6,0E-02	0,0E+00	1,0E+00
Network total revenue	1,8E+01	2,4E+00	1,5E+01	2,8E+01
Network total revenue squared	3,3E+02	9,6E+01	2,1E+02	7,7E+02
Network store count	5,8E+01	3,4E+02	1,0E+00	3,0E+03
Network store count squared	1,2E+05	8,9E+05	1,0E+00	8,7E+06
Closed on Monday	1,8E-01	3,9E-01	0,0E+00	1,0E+00
Closed on Tuesday	1,4E-01	3,5E-01	0,0E+00	1,0E+00
Open on Sunday	6,6E-01	4,7E-01	0,0E+00	1,0E+00
Interaction Closed on Monday	3,1E+00	6,5E+00	0,0E+00	2,4E+01
Interaction Closed on Tuesday	2,4E+00	6,0E+00	0,0E+00	2,4E+01
Interaction Open on Sunday	1,1E+01	8,1E+00	0,0E+00	2,5E+01
Interaction Network store count	9,5E+02	5,4E+03	1,5E+01	5,8E+04
Interaction Network total revenue	3,1E+02	6,5E+01	2,1E+02	6,7E+02
Annual revenue 4th degree orthogonal polinom	2,7E-03	1,1E-02	-1,8E-02	1,1E-01
Annual revenue 5th degree orthogonal polinom	-6,0E-03	9,3E-03	-1,8E-02	1,4E-01
Annual revenue 6th degree orthogonal polinom	2,0E-03	1,1E-02	-2,0E-02	1,7E-01

Annual revenue 7th degree orthogonal polinom	1,4E-03	1,1E-02	-2,4E-02	2,0E-01
Annual revenue 8th degree orthogonal polinom	-1,5E-03	1,1E-02	-3,1E-02	2,2E-01
Annual revenue 9th degree orthogonal polinom	7,7E-04	1,1E-02	-4,0E-02	2,5E-01
Annual revenue 10th degree orthogonal polinom	-4,1E-04	1,1E-02	-4,8E-02	2,7E-01

Annex 3: The coefficients and t-statistics for the main models

(For monthly models, the regression parameters are estimated for each month, but they show little dispersion. Due to limitations, the table shows only one representative month (March 2016), which is a good indicator of the results of a given model group.)

	Only size aggregate	Only size monthly	Full model aggregate	Full model monthly	Interaction terms aggregate	Interaction terms monthly
Constant	-11,481 (-22,474)	-0,850 (-0,159)	-2,742 (-4,952)	-22,033 (-0,017)	-8,640 (-13,816)	-33,877 (-0,026)
Annual revenue 1th degree orthogonal polinom	1 434,354 (24,750)	2 101,801 (5,725)	62,694 (59,217)	-7,652 (-1,273)	-613,569 (-36,806)	-687,525 (-7,083)
Annual revenue 2nd degree orthogonal polinom	-1 452,703 (-22,729)	272,596 (0,373)	-83,019 (-131,717)	-112,240 (-32,349)	-149,889 (-139,832)	-155,076 (-25,734)
Annual revenue 3rd degree orthogonal polinom	1 372,369 (22,260)	2 635,542 (7,153)	-27,195 (-53,301)	-49,776 (-17,439)	-14,471 (-17,605)	-26,731 (-6,094)
0-1 thousand HUF category logarith			0,077 (34,514)	0,125 (10,557)	0,104 (45,941)	0,112 (9,383)
1-5 thousand HUF category logarith			-0,098 (-30,877)	-0,114 (-7,077)	-0,071 (-22,483)	-0,112 (-6,937)
5-10 thousand HUF category logarith			0,110 (47,372)	0,106 (8,641)	0,107 (45,656)	0,063 (5,077)
10-20 thousand HUF category logarith			0,026 (12,003)	-0,019 (-1,702)	-0,010 (-4,569)	-0,061 (-5,345)
20- thousand HUF category logarith			0,050 (25,592)	0,054 (5,437)	0,061 (29,817)	0,072 (7,020)
0-1 thousand HUF category logarith squared			-0,012 (-49,728)	-0,019 (-15,266)	-0,016 (-64,825)	-0,019 (-14,500)
1-5 thousand HUF category logarith squared			0,004 (10,213)	0,005 (2,568)	0,003 (8,127)	0,004 (2,095)

5-10 thousand HUF category logarith squared			-0,008 (-22,492)	-0,014 (-7,250)	-0,009 (-23,208)	-0,008 (-4,187)
10-20 thousand HUF category logarith squared			0,011 (26,853)	0,025 (11,319)	0,019 (43,497)	0,032 (14,387)
20- thousand HUF category logarith squared			-0,015 (-39,931)	-0,019 (-9,649)	-0,018 (-45,057)	-0,023 (-11,169)
Average number of items			-0,081 (-53,316)	-0,168 (-20,793)	-0,067 (-39,907)	-0,124 (-14,086)
Average transaction value			0,000 (-46,180)	0,000 (-6,288)	0,000 (-34,580)	0,000 (-5,515)
0-1 thousand HUF category share			0,933 (23,625)	1,134 (5,686)	0,840 (21,409)	1,290 (6,499)
1-5 thousand HUF category share			0,783 (17,867)	0,355 (1,593)	0,446 (10,239)	0,255 (1,153)
5-10 thousand HUF category share			-0,018 (-0,391)	0,998 (4,123)	-0,024 (-0,511)	1,148 (4,739)
10-20 thousand HUF category share			2,386 (48,211)	2,045 (7,972)	2,060 (42,220)	1,874 (7,401)
0-1 thousand HUF category share squared			-0,427 (-12,581)	-0,655 (-3,795)	-0,423 (-12,497)	-0,906 (-5,256)
1-5 thousand HUF category share squared			-0,400 (-9,497)	-0,028 (-0,130)	-0,296 (-7,092)	0,035 (0,163)
5-10 thousand HUF category share squared			-0,203 (-3,173)	-1,618 (-4,739)	-0,515 (-8,035)	-2,137 (-6,172)
10-20 thousand HUF category share squared			-2,363 (-38,484)	-2,232 (-6,844)	-2,297 (-37,879)	-2,294 (-7,100)
Detailed receipt			0,184 (92,513)	0,218 (22,010)	-2,167 (-106,012)	-1,322 (-12,586)
Network total revenue			-0,233 (-16,605)	0,438 (5,625)	-3,024 (-23,624)	-2,091 (-2,592)
Network total revenue squared			0,014 (38,512)	-0,004 (-1,938)	0,036 (10,567)	0,009 (0,397)
Network store count			-0,009 (-161,213)	-0,002 (-2,749)	0,079 (137,386)	0,050 (8,458)
Network store count squared			0,000 (155,676)	0,000 (-8,091)	0,000 (-140,840)	0,000 (7,484)
Closed on Monday			-0,266 (-65,195)	-0,389 (-16,968)	0,007 (0,123)	1,256 (3,886)
Closed on Tuesday			-0,006 (-1,272)	-0,063 (-2,202)	-0,527 (-8,022)	-1,679 (-4,310)

Open on Sunday			-0,339 (-115,088)	-0,424 (-31,380)	-0,711 (-18,186)	-0,415 (-2,196)
County: Mozgóbolt			-0,806 (-68,875)	-0,484 (-7,030)	-0,826 (-69,670)	-0,500 (-7,185)
County: Bács-Kiskun			0,071 (10,062)	0,125 (3,323)	0,076 (10,744)	0,121 (3,176)
County: Baranya			-0,039 (-4,408)	-0,088 (-1,829)	-0,037 (-4,109)	-0,084 (-1,740)
County: Békés			-0,288 (-34,592)	-0,248 (-5,515)	-0,296 (-35,285)	-0,245 (-5,414)
County: Borsod-Abaúj-Zemplén			-0,354 (-38,755)	-0,362 (-7,289)	-0,359 (-38,943)	-0,353 (-7,070)
County: Budapest			-0,152 (-18,408)	-0,128 (-2,868)	-0,161 (-19,408)	-0,134 (-3,000)
County: Csongrád			-0,104 (-12,057)	-0,083 (-1,778)	-0,109 (-12,542)	-0,073 (-1,560)
County: Fejér			-0,081 (-8,994)	-0,022 (-0,452)	-0,078 (-8,652)	-0,018 (-0,378)
County: Győr-Moson-Sopron			-0,243 (-28,951)	-0,181 (-4,002)	-0,242 (-28,631)	-0,170 (-3,749)
County: Hajdú-Bihar			-0,108 (-12,926)	-0,134 (-2,953)	-0,109 (-13,053)	-0,120 (-2,636)
County: Heves			0,034 (3,613)	0,047 (0,917)	0,030 (3,195)	0,052 (1,013)
County: Jász-Nagykun-Szolnok			0,052 (5,514)	0,008 (0,149)	0,059 (6,218)	0,022 (0,434)
County: Komárom-Esztergom			-0,188 (-16,520)	-0,128 (-2,099)	-0,194 (-16,859)	-0,121 (-1,967)
County: Nógrád			-0,064 (-8,537)	0,005 (0,135)	-0,070 (-9,268)	0,001 (0,029)
County: Pest			-0,200 (-21,718)	-0,080 (-1,583)	-0,208 (-22,338)	-0,089 (-1,742)
County: Somogy			-0,431 (-50,527)	-0,396 (-8,913)	-0,438 (-50,879)	-0,401 (-8,967)
County: Szabolcs-Szatmár-Bereg			-0,199 (-22,190)	-0,130 (-2,714)	-0,207 (-22,837)	-0,140 (-2,890)
County: Tolna			-0,242 (-23,867)	-0,273 (-4,988)	-0,247 (-24,112)	-0,273 (-4,944)
County: Vas			-0,559 (-56,715)	-0,590 (-10,943)	-0,577 (-57,740)	-0,591 (-10,852)
County: Veszprém			-0,023 (-2,557)	0,040 (0,837)	-0,020 (-2,192)	0,051 (1,054)

County: Zala					
Dummy variable 2015 1.	-0,237 (-33,022)		-0,026 (-3,251)		-0,041 (-5,098)	
Dummy variable 2015 2.	-0,088 (-10,296)		-0,160 (-17,690)		-0,170 (-18,595)	
Dummy variable 2015 3.	-0,074 (-8,620)		-0,155 (-16,981)		-0,172 (-18,683)	
Dummy variable 2015 4.	-0,271 (-35,330)		-0,353 (-43,085)		-0,365 (-43,929)	
Dummy variable 2015 5.	-0,148 (-19,030)		-0,145 (-17,360)		-0,153 (-18,160)	
Dummy variable 2015 6.	-0,173 (-21,040)		-0,252 (-28,732)		-0,244 (-27,507)	
Dummy variable 2015 7.	-0,147 (-17,076)		-0,230 (-25,240)		-0,225 (-24,391)	
Dummy variable 2015 8.	-0,181 (-21,806)		-0,240 (-27,098)		-0,247 (-27,580)	
Dummy variable 2015 9.	-0,107 (-12,619)		-0,217 (-24,006)		-0,233 (-25,494)	
Dummy variable 2015 10.	-0,106 (-12,567)		-0,203 (-22,525)		-0,212 (-23,302)	
Dummy variable 2015 11.	-0,079 (-9,238)		-0,187 (-20,395)		-0,208 (-22,499)	
Dummy variable 2015 12.	-0,044 (-5,197)		-0,133 (-14,582)		-0,149 (-16,168)	
Dummy variable 2016 1.	-0,050 (-6,736)		-0,034 (-4,379)		-0,041 (-5,207)	
Dummy variable 2016 2.	0,092 (11,757)		0,139 (16,768)		0,122 (14,466)	
Dummy variable 2016 3.	0,019 (2,248)		0,022 (2,465)		0,026 (2,829)	
Dummy variable 2016 4.	0,000 (0,000)		0,000 (0,000)		0,000 (0,000)	
Dummy variable 2016 5.	0,037 (4,290)		-0,087 (-9,602)		-0,083 (-9,021)	
Dummy variable 2016 6.	0,020 (2,407)		-0,060 (-6,666)		-0,063 (-6,971)	
Dummy variable 2016 7.	0,058 (7,448)		0,102 (12,244)		0,094 (11,152)	
Dummy variable 2016 8.	0,009 (1,027)		-0,024 (-2,689)		-0,022 (-2,396)	
Dummy variable 2016 9.	0,008 (1,001)		0,021 (2,355)		0,016 (1,765)	

Dummy variable 2016 10.	0,063 (7,594)		0,094 (10,833)		0,097 (11,077)	
Dummy variable 2016 11.	0,041 (4,805)		0,037 (4,154)		0,028 (3,132)	
Dummy variable 2016 12.	0,092 (10,896)		0,097 (10,980)		0,088 (9,783)	
Dummy variable network decision						
Dummy variable independent store	0,000 (0,000)	0,000 (0,000)	0,230 (63,840)	0,236 (12,744)	0,257 (61,647)	0,238 (11,177)
Dummy variable network independent decision	0,000 (0,000)	0,000 (0,000)	0,535 (155,361)	0,581 (32,652)	0,555 (157,752)	0,560 (31,106)
Interaction detailed receipt					0,135 (113,655)	0,087 (14,299)
Interaction Closed on Monday					-0,016 (-4,582)	-0,097 (-5,146)
Interaction Closed on Tuesday					0,030 (7,844)	0,091 (4,044)
Interaction Open on Sunday					0,021 (9,469)	0,000 (-0,015)
Interaction Network store count					-0,005 (-152,060)	-0,003 (-7,345)
Interaction Network total revenue					0,174 (23,170)	0,201 (4,352)
Annual revenue 4th degree orthogonal polinom	-1 135,276 (-22,008)	928,262 (1,298)				
Annual revenue 5th degree orthogonal polinom	913,584 (22,784)	2 236,127 (9,100)				
Annual revenue 6th degree orthogonal polinom	-633,510 (-22,732)	824,614 (1,792)				
Annual revenue 7th degree orthogonal polinom	390,951 (23,344)	1 092,196 (9,977)				
Annual revenue 8th degree orthogonal polinom	-218,289 (-22,179)	352,757 (1,977)				
Annual revenue 9th degree orthogonal polinom	91,791 (26,253)	262,319 (10,880)				
Annual revenue 10th degree orthogonal	-40,567 (-20,566)	53,806 (1,575)				

polinom						

Annex 4: The effect of the share of purchases with different transaction values

(Vertical axis: calculated effect on the scale of explanatory variables, horizontal axis: proportion of the given category)

Annex 5: The coefficients and t-statistics for the subsample models

	Network decision stores	Independent stores	Network independent stores	Small stores	Middle sized	Large retailers
Constant	-12,702 (-9,820)	-419,759 (-0,178)	42,560 (17,920)	-19,222 (-0,026)	-48,492 (-10,674)	42,560 (17,920)
Annual revenue 1th degree orthogonal polinom	-70,250 (-8,334)	-8 260,101 (-1,698)	59,917 (3,332)	181,105 (1,195)	543,140 (2,161)	59,917 (3,332)
Annual revenue 2nd degree orthogonal polinom	-124,414 (-25,701)	-640,691 (-0,144)	-141,303 (-14,731)	-40,092 (-0,933)	-299,240 (-1,976)	-141,303 (-14,731)
Annual revenue 3rd degree orthogonal polinom	-34,418 (-8,014)	-36,719 (-5,221)	-64,748 (-7,512)	33,333 (0,490)	177,927 (2,534)	-64,748 (-7,512)
0-1 thousand HUF category logarith	0,094 (5,259)	0,112 (5,369)	-0,023 (-0,767)	0,085 (5,324)	-0,079 (-1,754)	-0,023 (-0,767)

1-5 thousand HUF category logarith	-0,107 (-4,058)	-0,106 (-3,961)	-0,098 (-2,573)	-0,129 (-5,587)	-0,049 (-0,656)	-0,098 (-2,573)
5-10 thousand HUF category logarith	0,154 (7,669)	0,039 (1,841)	-0,089 (-3,229)	0,051 (2,976)	0,185 (2,505)	-0,089 (-3,229)
10-20 thousand HUF category logarith	-0,082 (-4,659)	0,004 (0,209)	-0,020 (-0,756)	-0,048 (-3,154)	-0,227 (-3,985)	-0,020 (-0,756)
20- thousand HUF category logarith	0,121 (8,040)	0,087 (4,623)	0,053 (2,190)	0,064 (4,590)	0,024 (0,602)	0,053 (2,190)
0-1 thousand HUF category logarith squared	-0,014 (-7,030)	-0,016 (-7,100)	-0,006 (-2,043)	-0,011 (-6,395)	0,002 (0,422)	-0,006 (-2,043)
1-5 thousand HUF category logarith squared	0,003 (0,965)	-0,001 (-0,200)	0,001 (0,145)	0,004 (1,806)	-0,016 (-2,203)	0,001 (0,145)
5-10 thousand HUF category logarith squared	-0,026 (-8,541)	-0,002 (-0,623)	0,017 (3,720)	-0,010 (-4,026)	-0,013 (-1,638)	0,017 (3,720)
10-20 thousand HUF category logarith squared	0,042 (12,589)	0,023 (5,564)	0,004 (0,671)	0,029 (9,997)	0,051 (6,586)	0,004 (0,671)
20- thousand HUF category logarith squared	-0,029 (-10,203)	-0,032 (-8,530)	-0,013 (-2,569)	-0,021 (-7,563)	-0,021 (-3,670)	-0,013 (-2,569)
Average number of items	-0,178 (-15,223)	-0,068 (-4,231)	-0,053 (-2,567)	-0,092 (-7,718)	-0,108 (-5,111)	-0,053 (-2,567)
Average transaction value	0,000 (-1,685)	0,000 (-9,375)	0,000 (-2,097)	0,000 (-8,484)	0,000 (-1,176)	0,000 (-2,097)
0-1 thousand HUF category share	1,717 (5,495)	0,443 (1,340)	2,409 (4,704)	1,353 (5,024)	7,464 (8,561)	2,409 (4,704)
1-5 thousand HUF category share	-0,188 (-0,537)	0,188 (0,507)	1,313 (2,315)	0,174 (0,569)	-0,378 (-0,372)	1,313 (2,315)
5-10 thousand HUF category share	1,646 (4,126)	1,058 (2,707)	0,552 (0,912)	1,043 (3,136)	0,713 (0,550)	0,552 (0,912)
10-20 thousand HUF category share	2,140 (5,307)	0,769 (1,846)	2,167 (3,323)	2,253 (6,413)	5,078 (3,910)	2,167 (3,323)
0-1 thousand HUF category share squared	-1,893 (-6,768)	-0,218 (-0,785)	-1,929 (-4,417)	-1,672 (-7,032)	-6,711 (-7,929)	-1,929 (-4,417)
1-5 thousand HUF category share squared	0,496 (1,475)	-0,063 (-0,178)	-0,936 (-1,694)	-0,160 (-0,552)	2,612 (2,650)	-0,936 (-1,694)
5-10 thousand HUF category share squared	-2,990 (-5,006)	-1,999 (-3,829)	-1,177 (-1,377)	-2,190 (-4,406)	-1,595 (-0,841)	-1,177 (-1,377)

10-20 thousand HUF category share squared	-2,832 (-5,229)	-0,939 (-1,895)	-2,977 (-3,589)	-3,008 (-6,254)	-4,169 (-1,926)	-2,977 (-3,589)
Detailed receipt	0,280 (18,528)	0,130 (7,548)	0,117 (4,630)	0,148 (11,102)	0,361 (10,890)	0,117 (4,630)
Network total revenue	0,419 (3,055)	6,723 (0,029)	-4,064 (-15,878)	0,443 (3,304)	4,265 (10,899)	-4,064 (-15,878)
Network total revenue squared	0,009 (2,520)	0,835 (0,124)	0,093 (13,764)	-0,004 (-1,119)	-0,104 (-11,097)	0,093 (13,764)
Network store count	-0,022 (-35,580)	0,000 (0,000)	0,010 (4,946)	-0,003 (-5,609)	0,060 (21,312)	0,010 (4,946)
Network store count squared	0,000 (21,915)	0,000 (0,000)	0,000 (-3,702)	0,000 (1,150)	0,000 (-27,435)	0,000 (-3,702)
Closed on Monday	-0,564 (-14,507)	-0,387 (-10,667)	-0,231 (-4,388)	-0,345 (-11,711)	-0,746 (-7,777)	-0,231 (-4,388)
Closed on Tuesday	0,040 (0,851)	-0,121 (-2,504)	-0,450 (-6,967)	-0,198 (-5,295)	0,308 (2,788)	-0,450 (-6,967)
Open on Sunday	-0,327 (-15,334)	-0,435 (-18,690)	-0,370 (-11,715)	-0,477 (-27,423)	-0,442 (-7,836)	-0,370 (-11,715)
County: Mozgóbolt	-0,575 (-5,117)	-0,818 (-5,680)	-0,159 (-1,107)	-0,590 (-6,942)	-0,430 (-1,577)	-0,159 (-1,107)
County: Bács-Kiskun	0,150 (2,080)	0,145 (2,458)	0,260 (3,055)	0,024 (0,498)	0,073 (0,456)	0,260 (3,055)
County: Baranya	-0,064 (-0,676)	-0,208 (-3,120)	0,105 (0,923)	-0,071 (-1,158)	0,148 (0,681)	0,105 (0,923)
County: Békés	-0,269 (-3,043)	-0,357 (-5,741)	-0,083 (-0,792)	-0,232 (-4,057)	-0,267 (-1,424)	-0,083 (-0,792)
County: Borsod-Abaúj-Zemplén	-0,537 (-5,534)	-0,426 (-6,053)	-0,019 (-0,174)	-0,348 (-5,538)	-0,324 (-1,514)	-0,019 (-0,174)
County: Budapest	-0,314 (-3,555)	-0,169 (-2,750)	0,074 (0,728)	-0,171 (-3,011)	-0,338 (-1,733)	0,074 (0,728)
County: Csongrád	-0,204 (-2,259)	-0,171 (-2,636)	0,219 (1,987)	-0,065 (-1,094)	-0,173 (-0,880)	0,219 (1,987)
County: Fejér	-0,088 (-0,962)	-0,094 (-1,405)	0,013 (0,108)	-0,011 (-0,176)	-0,232 (-1,160)	0,013 (0,108)
County: Győr-Moson-Sopron	-0,066 (-0,743)	-0,263 (-4,211)	-0,076 (-0,733)	-0,176 (-3,064)	-0,192 (-1,067)	-0,076 (-0,733)
County: Hajdú-Bihar	-0,170 (-1,902)	-0,188 (-3,002)	0,003 (0,029)	-0,112 (-1,951)	-0,213 (-1,101)	0,003 (0,029)
County: Heves	0,094 (0,928)	-0,012 (-0,168)	0,041 (0,351)	0,036 (0,548)	0,318 (1,376)	0,041 (0,351)
County: Jász-Nagykun-Szolnok	0,148 (1,521)	-0,110 (-1,561)	0,022 (0,174)	0,169 (2,558)	-0,385 (-1,844)	0,022 (0,174)

County: Komárom-Esztergom	-0,106 (-0,908)	-0,215 (-2,475)	-0,180 (-1,279)	-0,113 (-1,445)	-0,343 (-1,218)	-0,180 (-1,279)
County: Nógrád	-0,034 (-0,434)	-0,059 (-1,066)	0,043 (0,448)	0,006 (0,115)	-0,130 (-0,761)	0,043 (0,448)
County: Pest	-0,195 (-1,927)	-0,180 (-2,561)	0,236 (2,068)	-0,097 (-1,499)	-0,564 (-2,710)	0,236 (2,068)
County: Somogy	-0,435 (-5,356)	-0,600 (-8,451)	-0,065 (-0,665)	-0,485 (-8,595)	-0,312 (-1,660)	-0,065 (-0,665)
County: Szabolcs-Szatmár-Bereg	-0,025 (-0,266)	-0,280 (-4,182)	-0,140 (-1,255)	-0,082 (-1,349)	-0,257 (-1,224)	-0,140 (-1,255)
County: Tolna	-0,108 (-1,038)	-0,441 (-5,666)	-0,133 (-1,012)	-0,264 (-3,821)	0,015 (0,061)	-0,133 (-1,012)
County: Vas	-0,622 (-5,919)	-0,659 (-8,622)	-0,170 (-1,416)	-0,614 (-9,093)	-0,356 (-1,702)	-0,170 (-1,416)
County: Veszprém	0,098 (1,026)	-0,090 (-1,356)	0,218 (1,967)	0,100 (1,630)	-0,248 (-1,225)	0,218 (1,967)
County: Zala						

Annex 6: Descriptives of main predictor variables

Name	Mean	StdDev	Min	Max
Logarithm of transaction value 1th degree orthogonal polinom below 32,000 HUF	-0,01525	0,004892	-0,03791	0,024333
Logarithm of transaction value 2nd degree orthogonal polinom below 32,000 HUF	0,004167	0,01095	-0,01388	0,090028
Logarithm of transaction value 3rd degree orthogonal polinom below 32,000 HUF	0,006774	0,009115	-0,162	0,034873
Logarithm of transaction value 4th degree orthogonal polinom below 32,000 HUF	-0,0047	0,008842	-0,01511	0,190489
Logarithm of transaction value 5th degree orthogonal polinom below 32,000 HUF	-0,00354	0,010616	-0,14551	0,039384
500 HUF dummy	0,003378	0,058022	0	1
1,000 HUF dummy	0,005943	0,076864	0	1
2,000 HUF dummy	0,00288	0,053593	0	1
5,000 HUF dummy	0,002477	0,049706	0	1
10,000 HUF dummy	0,001146	0,033832	0	1
20,000 HUF dummy	0,000202	0,014194	0	1
Store annual revenue logarithm	7,761561	0,725487	2,812913	9,942669

Number of sold items	4,151602	6,552706	0	539
Divisible by 10	0,381384	0,485727	0	1
Divisible by 100	0,079764	0,270927	0	1
Divisible by 1,000	0,018666	0,135343	0	1
IMR	0,338639	0,318712	0	3,503206

Annex 7: Coefficients of size variables

Variable name	Coefficien t	Std. error	Z score	P-value
(Intercept)	-12,9751	3,529241	- 3,67645	0,00023 7
Logarithm of transaction value 1th degree orthogonal polinom below 32,000 HUF	-1159,24	16,35476	- 70,8807	2,22E- 16
Logarithm of transaction value 2nd degree orthogonal polinom below 32,000 HUF	-963,898	12,54676	- 76,8244	2,22E- 16
Logarithm of transaction value 3rd degree orthogonal polinom below 32,000 HUF	-502,233	6,957532	- 72,1855	2,22E- 16
Logarithm of transaction value 4th degree orthogonal polinom below 32,000 HUF	-182,665	2,888923	- 63,2296	2,22E- 16
Logarithm of transaction value 5th degree orthogonal polinom below 32,000 HUF	-54,7167	0,784324	- 69,7629	2,22E- 16
Logarithm of transaction value 1th degree orthogonal polinom above 32,000 HUF	-103,896	55,677	-1,866	0,062
Logarithm of transaction value 2nd degree orthogonal polinom above 32,000 HUF	74,206	69,124	1,074	0,283
Logarithm of transaction value 3rd degree orthogonal polinom above 32,000 HUF	-74,496	65,200	-1,143	0,253
Logarithm of transaction value 4th degree orthogonal polinom above 32,000 HUF	67,343	36,941	1,823	0,068
Logarithm of transaction value 5th degree orthogonal polinom above 32,000 HUF	0,607	11,566	0,052	0,958
Store annual revenue logarithm	-0,96748	0,004511	-214,45	2,22E- 16
Store annual revenue logarithm squared	0,073481	0,000279	263,673 5	2,22E- 16
Number of sold items	0,004847	6,35E-05	76,3126	2,22E-

			3	16
Number of sold items squared	-4,89E-05	1,18E-06	- 41,6041	2,22E- 16

Annex 8: Coefficients of ease of payment variables

Variable name	Coefficient	Std. error	Z score	P-value
500 HUF dummy	0.083985	0.001197	70.16386	2.22E-16
1,000 HUF dummy	0.031859	0.000986	32.31007	2.22E-16
2,000 HUF dummy	-0.04878	0.001083	-45.0532	2.22E-16
5,000 HUF dummy	0.23815	0.000921	258.6929	2.22E-16
10,000 HUF dummy	0.087521	0.001137	76.96734	2.22E-16
20,000 HUF dummy	-0.22813	0.002527	-90.2625	2.22E-16
Divisible by 10	-0.07303	0.000118	-616.503	2.22E-16
Divisible by 100	-0.0404	0.000222	-182.358	2.22E-16
Divisible by 1,000	0.011508	0.000498	23.10849	2.22E-16
Number of needed bills = 1	-0.33724	0.055612	-6.06412	1.33E-09
Number of needed bills = 2	-0.49573	0.055402	-8.94775	2.22E-16
Number of needed bills = 3	-0.3078	0.055298	-5.56625	2.60E-08
Number of needed bills = 4	-0.20914	0.055264	-3.78444	0.000154
Number of needed bills = 5	-0.16107	0.055244	-2.91571	0.003549
Number of needed bills = 6	-0.12609	0.055266	-2.28143	0.022523
Number of needed bills = 7	-0.08339	0.055287	-1.50825	0.131489
Number of needed bills = 8	-0.08	0.055487	-1.44173	0.14938
Number of needed bills = 9	-0.11582	0.059535	-1.94543	0.051724
Number of needed bills = 10	-0.04702	0.066544	-0.70666	0.479775
Number of needed bills > 10	0	0	0	0

Annex 9: Coefficients of county dummy variables

Variable name	Coefficient	Std. error	Z score	P-value
COUNTIES=Mobile store	0.619743	0.008885	69.75306	2.22E-16
COUNTIES=Budapest	0.633496	0.006383	99.24407	2.22E-16
COUNTIES=Baranya	0.100812	0.008154	12.36274	2.22E-16
COUNTIES=Bacs-Kiskun	-0.21442	0.007985	-26.8519	2.22E-16
COUNTIES=Bekes	-0.16681	0.008893	-18.7571	2.22E-16
COUNTIES=Borsod-Abauj-	0.0514	0.007533	6.823669	8.87E-12

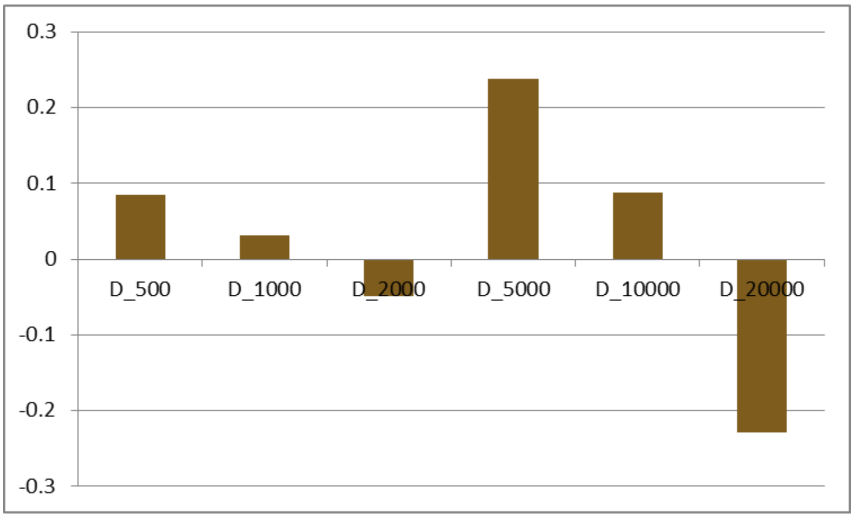
COUNTIES=Csongrad	0.104529	0.007847	13.32015	2.22E-16
COUNTIES=Fejer	0.217463	0.007923	27.44749	2.22E-16
COUNTIES=Gyor-Moson-So	0.113876	0.007541	15.10087	2.22E-16
COUNTIES=Hajdu-Bihar	-0.02824	0.007763	-3.63801	0.000275
COUNTIES=Heves	-0.10035	0.008742	-11.4796	2.22E-16
COUNTIES=Komarom-Eszte	0.173112	0.008417	20.56737	2.22E-16
COUNTIES=Nograd	-0.29459	0.011326	-26.0107	2.22E-16
COUNTIES=Pest	0.131806	0.006768	19.47494	2.22E-16
COUNTIES=Somogy	-0.06541	0.008781	-7.44961	2.22E-16
COUNTIES=Szabolcs-Szat	-0.30474	0.008551	-35.6368	2.22E-16
COUNTIES=Jasz-Nagykun-	-0.12801	0.008522	-15.0213	2.22E-16
COUNTIES=Tolna	-0.04976	0.009807	-5.07417	3.89E-07
COUNTIES=Vas	-0.09204	0.009155	-10.0541	2.22E-16
COUNTIES=Veszprem	0.205426	0.008072	25.45053	2.22E-16
COUNTIES=Zala	0	0	0	0

Annex 10: Coefficients of time variables

Variable name	Coefficient	Std. error	Z score	P-value
Year = 2015	-0.19756	0.00199	-99.2572	2.22E-16
Year = 2016	0	0	0	0
Month = 1	-0.01436	0.004987	-2.87863	0.003994
Month = 2	-0.06231	0.004894	-12.733	2.22E-16
Month = 3	-0.08823	0.004735	-18.6348	2.22E-16
Month = 4	-0.11332	0.004697	-24.1244	2.22E-16
Month = 5	-0.10911	0.004695	-23.2375	2.22E-16
Month = 6	-0.10705	0.004654	-23.0034	2.22E-16
Month = 7	-0.11341	0.004588	-24.7222	2.22E-16
Month =8	0.002754	0.004525	0.608563	0.542814
Month = 9	-0.08507	0.004645	-18.314	2.22E-16
Month = 10	-0.05284	0.004614	-11.4515	2.22E-16
Month = 11	-0.00413	0.004651	-0.8871	0.375026
Month =12	0	0	0	0
Day of month = 1	-0.02565	0.009705	-2.64317	0.008213
Day of month = 2	0.03447	0.009295	3.708447	0.000209
Day of month = 3	0.065578	0.009304	7.048275	2.22E-16

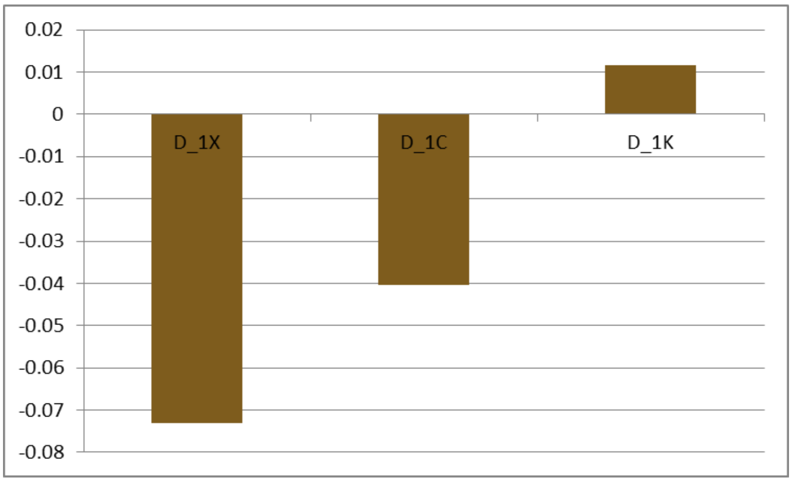
Day of month = 4	0.034715	0.009268	3.745599	0.00018
Day of month = 5	0.071647	0.009224	7.767427	2.22E-16
Day of month = 6	0.041842	0.00935	4.475236	7.63E-06
Day of month = 7	0.054695	0.009268	5.901514	3.60E-09
Day of month = 8	0.06504	0.009236	7.042033	2.22E-16
Day of month = 9	0.057955	0.009209	6.293285	3.11E-10
Day of month = 10	0.04876	0.009279	5.254893	1.48E-07
Day of month = 11	0.067962	0.009242	7.353546	2.22E-16
Day of month = 12	0.026515	0.009207	2.879831	0.003979
Day of month = 13	0.057902	0.009269	6.246765	4.19E-10
Day of month = 14	0.052594	0.009275	5.670318	1.43E-08
Day of month = 15	0.351288	0.009126	38.49433	2.22E-16
Day of month = 16	0.035944	0.009305	3.862953	0.000112
Day of month = 17	0.024302	0.00936	2.596313	0.009423
Day of month = 18	0.030756	0.009302	3.306379	0.000945
Day of month = 19	0.019697	0.009271	2.124444	0.033633
Day of month = 20	0.025451	0.009459	2.690822	0.007128
Day of month = 21	0.019454	0.009294	2.093322	0.03632
Day of month = 22	0.020687	0.009293	2.226031	0.026012
Day of month = 23	0.012608	0.00933	1.351289	0.176603
Day of month = 24	0.002917	0.009466	0.308176	0.757948
Day of month = 25	0.006799	0.009547	0.712156	0.476368
Day of month = 26	-0.00478	0.009504	-0.50249	0.615324
Day of month = 27	-0.00988	0.009469	-1.0436	0.296671
Day of month = 28	-0.00042	0.009453	-0.04456	0.96446
Day of month = 29	0.001281	0.009396	0.136346	0.891548
Day of month = 30	-0.00192	0.009403	-0.20403	0.838331
Day of month = 31	0	0	0	0
Weekday=1	-0.12778	0.004523	-28.2528	2.22E-16
Weekday=2	-0.04192	0.003539	-11.8446	2.22E-16
Weekday=3	-0.03457	0.003525	-9.80792	2.22E-16
Weekday=4	-0.02858	0.003491	-8.18594	2.22E-16
Weekday=5	-0.02278	0.003438	-6.62497	3.47E-11
Weekday=6	-0.02322	0.003361	-6.90753	4.93E-12
Weekday=7	0	0	0	0

Annex 11: Coefficients of the dummy variables of banknote denominations



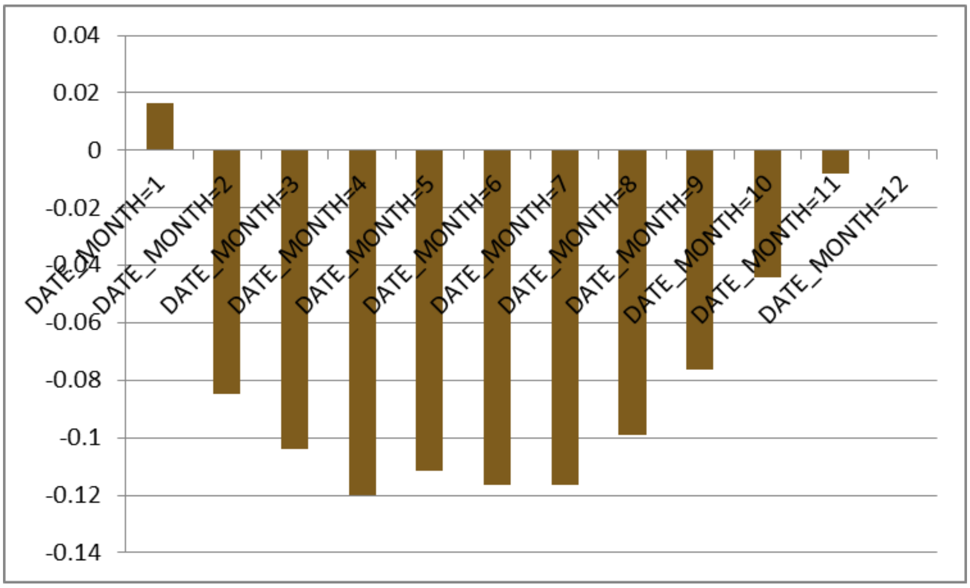
(Vertical axis: denomiation effect on the scale of the predictor variables, horizontal axis: transaction value)

Annex 12: Coefficients of the divisibility variables



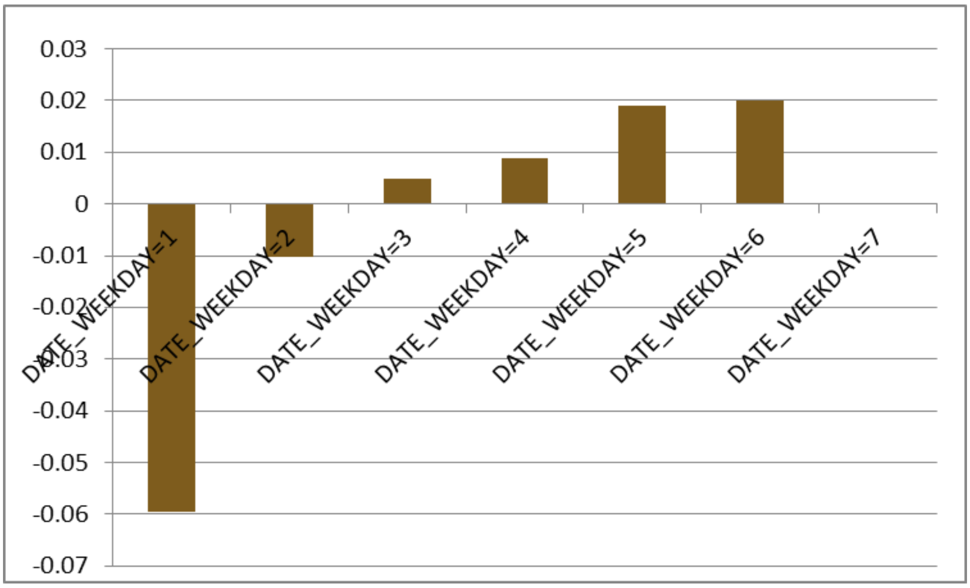
(Vertical axis: divisibility effect on the scale of the predictor variables, horizontal axis: divisible by 10/100/1,000)

Annex 13: Monthly seasonal coefficients



(Vertical axis: monthly fixed effects on the scale of the predictor variables, horizontal axis: month of the year)

Annex 14: Coefficients of the weekly time variables



(Vertical axis: intra week effects on the scale of the predictor variables, horizontal axis: weeday)