Effectiveness of stratified sampling to model payment card acceptance and usage[1]

Christopher S. Henry[2]

Tamás Ilyés[3]

Abstract

In our paper we analysed payment card acceptance and payment card use decisions in retail situations comparing logit models estimated on the comprehensive Online Cashier Register database and random stratified subsamples. We compare county, industry and store size strata with the true population estimates to simulate the usual stratification criteria for merchant surveys. In our model we control for several other factors relevant in payment instrument adoption and use but we primarily focus on the performance of these three types of stratification to estimate the exact county, industry and size effects for the entire population. In our comparison we create random stratified subsamples of 1 percent of the merchant database and random stratified subsamples of 0.01 percent of the transaction database. Based on our analysis in card acceptance models the store size stratification provides good estimates, however the same stratification cannot be used to effectively estimate variable effects in a card usage model.

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[1] Draft version, do not cite without author permission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] Corresponding author: Survey methodologist, Currency Department, Bank of Canada; PhD student, Université Clermont Auvergne. Email: chenry@bankofcanada.ca

[3] Email: tamas.ilyes23@gmail.com

1. INTRODUCTION

The objective of our study was to explore the aspects considered in the card acceptance decisions of retail merchants and card use decisions by customers in Hungary based on a comprehensive administrative database of the retail sector and compare the results to random stratified subsamples. 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 the 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 and payment card use. Thanks to the large sample size, we were able to reliably identify even narrow segments and negligible effects and compare the results to the usual stratified subsample approach of the surveys.

The Hungarian payment system 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 diary 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 main results of these surveys on the unique database of online cash registers, and analyse the effectiveness of random stratified surveys in the payment economics field. In our analysis by controlling for several factors, we primary focus on estimating the exact effect of three types of explanatory variables: county effects, industry effects and store size effects.

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 a 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 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. METHODOLOGY

3.1. Data source

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.

In the first 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.

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

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. In our analysis we 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.

The basis of our analysis for analysing card usage 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.

3.2. Variables used in the model

3.2.1. Card acceptance

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.

Value categories

Based on the payment literature, 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 we included dummy variables for the three types of stores – 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 digit of the identifier.

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.

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.

3.2.2. Card usage

Dependent variable

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. 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 the number of banknotes and coins handled in the ideal case as a dummy variable, up to a value of 10. These variables capture the ease of cash payment, which presumably correlates with payment time and as such, it can be considered 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 part of the study. As regards store attributes, most models include the log and square of annualised turnover and the aggregate form of the activity.

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.

3.3. Random stratified sampling

The objective of our analysis is to compare the models estimated on the full database to random subsamples of the administrative database. We created three types of random stratified samples for both the card acceptance and the card use model. Our aim was to determine which kind of stratified sampling provides the best estimates. For this reason we stratified the database based on counties, industries and store size. In both the acceptance and in the usage model we created stratas for every month and the respective subcategory. For the card acceptance model we took 10 random samples of 1 percent of the data from each strata of the store database, for the card use model the sampling ratio is only 0.01 % from the transactional dataset. To best simulate the methods of merchant surveys the transaction sampling should be based on randomly selecting stores from the database. Since there is significant size heterogeneity of stores, the random samples would provide vastly different number of transactions for the samples. To make the subsamples comparable for this small number of samples we do not follow this approach.

The card usage model estimated on the entire database is highly resource intensive and takes a considerable amount of time; by using stratified sampling we can create models in a more efficient way. To minimize the complexity of the model we omit the breakpoint for the transaction size – use relation at 32 000 HUF and only model transactions which are below this threshold. For computational purposes in the card usage model Frisch-Waugh-Lovell (FWL) Theorem is applied and we use Iteratively Reweighted Least Squares (IRLS) algorithm, which is equivalent to full maximum likelihood (Lowell 2008).

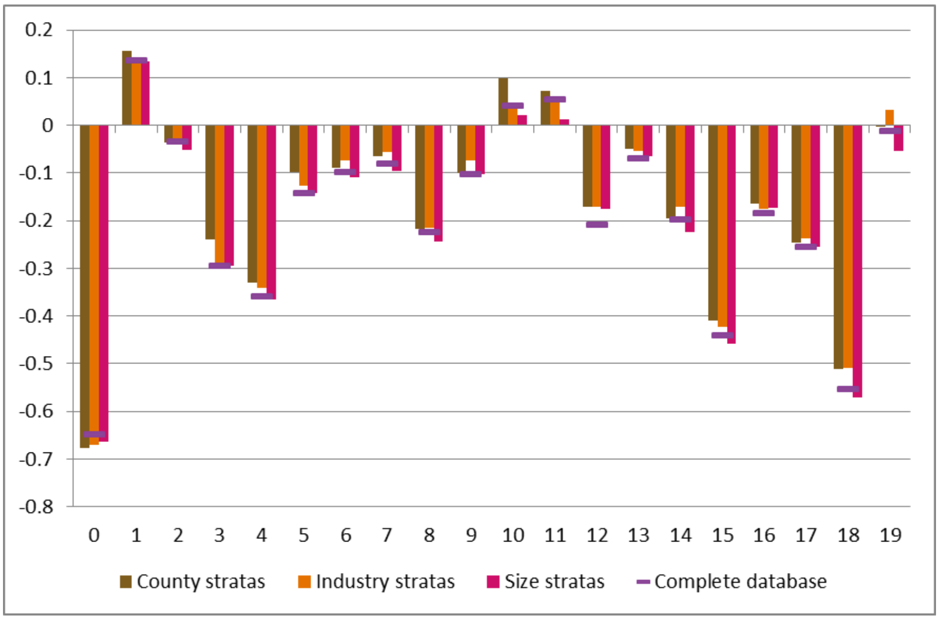
4. RESULTS

4.1. Card acceptance

Our analysis focus on three types of explanatory variables: county effects, industry effects and store size effects on the entire sector. Due to the high number of control variables we only discuss the estimates for these groups of variables. On general for the 88 parameter estimates on average the stratas based on the size of the store provide the best estimates – the average of the 10 subsample is in 52 cases inside of the 95 % confidence interval.

The model based on the entire database clearly shows that the most important factor – in line with the literature results – is the size of the store, which we characterise by the annual revenue. The distribution of store sizes follow a lognormal distribution which means that in the county and industry based stratas there is a low probability that they will be included in the sample. Without the biggest retailers, where county and industry effects are small compared to the size effects the estimates for these variables will on general be biased. The county and industry stratas overestimate these effects.

1. Figure: Average county dummy coefficient estimates in the card acceptance model



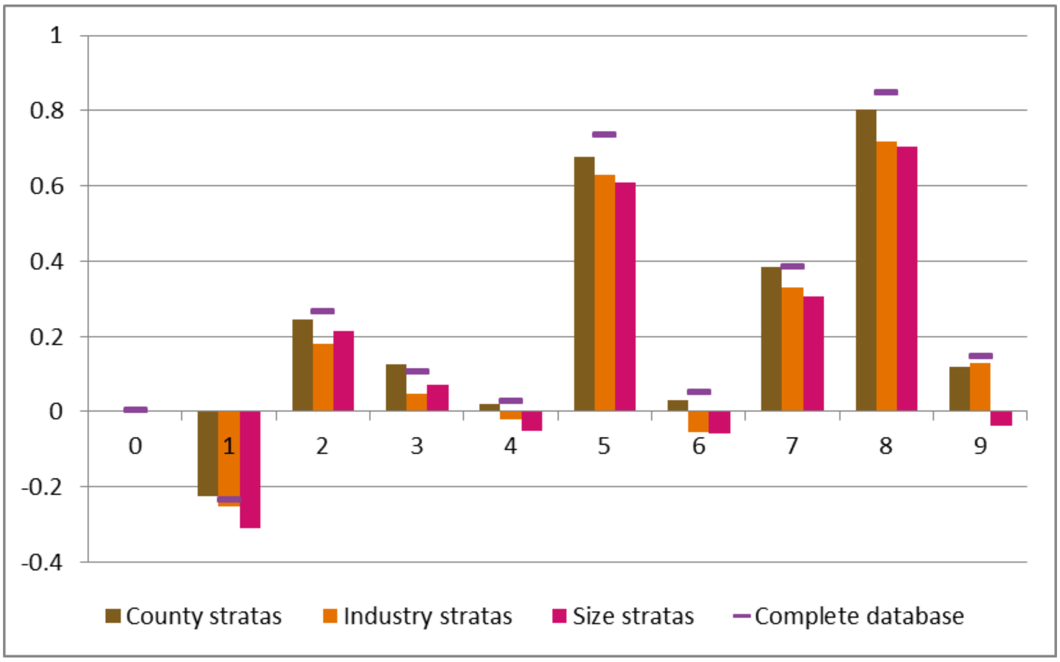
Since the functional relation between size and acceptance is non-linear and non-monotone, the stratas of the counties and industries do not provide good fits for these higher order polynomials and most of the size related variables – share of different size transactions, volume of transactions. In the case of industry effects the size based stratas underperform the industry based stratas. The industry distribution of the database is heavily concentrated on retail services and the other categories have only a very small share. The three categories of size effects – direct annual revenue, share of different size transactions and volume of transactions – are on general better estimated by size stratas.

In conclusion we can state that the most efficient stratification is a random stratified sampling based on different store size categories and not on geographical or industry classification. The main causes are the importance of annual revenue over county and industry effects in card acceptance decisions and the complex functional relation between the two. By not basing stratification on store sizes as well, the model overestimates the other effects.

4.2. Card usage

In the case of modelling payment card use the above approach does not lead to good estimates. The model estimated on the full dataset clearly shows that the single most important factor is the transaction value and its higher order orthogonal polynomials. We simulate random stratified subsamples based on counties, industries and store sizes and not on transaction size. Because of this there is no single best method from these three types of stratification.

2. Figure: Average industry dummy coefficient estimates in the card use model



Because of the extremely small standard errors calculated from nearly 5 billion transactions all subsample estimation are on average outside of the 95 % confidence intervals. Based on the average estimations, the county stratas provide the closest estimations for most variables. The reason for this is the much greater county effect in card use decisions compared to card acceptance. However as opposed to what we observed in the card acceptance model, there is no difference in biases between the three types of stratification. All three models on average overestimate the county and transaction size effects, and underestimate the industry effects.

The main reason that the above stratifications provide poor results is the lognormality of transaction size distribution – which is similar to the lognormal distribution of store sizes. By not basing the stratification on this characteristic we do not make ensure that enough high value transaction is in the subsample. This bias is present even by limiting the sample to transaction below 32 thousand HUF (~100 EUR). Without this filtering, the absence of these extreme value transactions together with the non-monotonic, non-linear relationship between transaction size and use for the entire database would probably create even less efficient estimations.

In conclusion we can state that the usual stratification of merchant survey – geographical location, industry, size – is not applicable to card use models because they omit the most important factor in card use decision, the value of the transaction. From these three types of stratification, the county stratas provide the best estimates but there is a systematic bias for all three types of stratification.

5. CONCLUSION

In our paper we analysed payment card acceptance and payment card use decisions in retail situations comparing logit models estimated on the comprehensive Online Cashier Register database and random stratified subsamples. We compare county, industry and store size stratas with the true population estimates to simulate the usual stratification criteria for merchant surveys.

In our model we control for several other factors relevant in payment instrument adoption and use but we primarily focus on the performance of these three types of stratification to estimate the exact county, industry and size effects for the entire population. In our comparison we create 10 random stratified subsamples of 1 percent of the merchant database and 10 random stratified subsamples of 0.01 percent of the transaction database.

In the card acceptance model the stratification based on the store sizes provide the best estimates. In card acceptance decisions the most important factor is the size of the store, and other size related variables. However the store sizes follow a lognormal distribution. With small samples there is a high probability that the subsample does not have enough big stores and the average effects of size will be underestimated, the county and industry effects overestimated. In the card usage model we evaluate the same subsample types and show that county stratification provides the best results. However due to the same problem as in the acceptance model – lognormal distribution, non-monotone relation – and not creating stratas on transaction sizes all of the three analysed subsampling provide systematic biases.

In conclusion it can be stated that in card acceptance models the store size stratification provides good estimates, however the same stratification cannot be used to effectively estimate variable effects in a card usage model.

6. 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.

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

Lovell, M., (2008): *A Simple Proof of the FWL (Frisch,Waugh,Lovell) Theorem*. Journal of Economic Education.

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

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.

7.1. Card acceptance model average coefficient estimates

	Full dataset	County stratas	Industry stratas	Size stratas
(Intercept)	-1.276	-2.216	-0.862	-1.647
Average number of items	-0.135	-0.140	-0.135	-0.138
Average value of transaction	0.000	0.000	0.000	0.000
Closed on Monday	-0.254	-0.261	-0.241	-0.238
Closed on Tuesday	0.018	0.021	0.003	0.001
Open on Sunday	-0.392	-0.380	-0.410	-0.403
Number of items	0.000	0.000	0.000	0.000
County code = 0	-0.649	-0.678	-0.671	-0.663
County code = 01	0.136	0.157	0.142	0.135
County code = 02	-0.035	-0.035	-0.034	-0.052
County code = 03	-0.295	-0.240	-0.292	-0.295
County code = 04	-0.360	-0.329	-0.341	-0.366
County code = 05	-0.143	-0.097	-0.126	-0.142
County code = 06	-0.100	-0.089	-0.074	-0.108
County code = 07	-0.082	-0.065	-0.055	-0.096
County code = 08	-0.225	-0.217	-0.216	-0.243
County code = 09	-0.103	-0.100	-0.073	-0.101
County code = 10	0.040	0.099	0.046	0.022
County code = 11	0.053	0.072	0.048	0.013
County code = 12	-0.209	-0.171	-0.171	-0.176
County code = 13	-0.070	-0.050	-0.054	-0.064
County code = 14	-0.198	-0.195	-0.172	-0.224
County code = 15	-0.441	-0.410	-0.423	-0.457
County code = 16	-0.185	-0.164	-0.176	-0.173
County code = 17	-0.256	-0.246	-0.237	-0.255
County code = 18	-0.554	-0.511	-0.509	-0.570
County code = 19	-0.012	0.000	0.032	-0.054
County code = 20	0.000	0.000	0.000	0.000
County code = 21	0.000	0.000	0.000	0.000
Network store count	-0.008	-0.008	-0.008	-0.008
Network store count squared	0.000	0.000	0.000	0.000
Network sum value	-0.275	-0.226	-0.332	-0.230
Network sum value squared	0.015	0.014	0.016	0.014
SHARE_10K	0.079	0.660	0.445	0.111

SHARE_10K2	-0.186	-0.458	-0.440	-0.238
SHARE_1K	1.422	1.978	1.574	1.445
SHARE_1K2	-0.968	-1.003	-0.965	-1.056
SHARE_20K	2.608	3.255	2.569	2.612
SHARE_20K2	-2.465	-2.738	-2.222	-2.533
SHARE_5K	0.401	0.722	0.599	0.311
SHARE_5K2	0.125	0.249	0.087	0.167
Industry code = 0	-0.609	-0.595	-0.642	-0.634
Industry code = 1	-0.337	-0.299	-0.370	-0.386
Industry code = 2	-0.303	-0.343	-0.349	-0.280
Industry code = 3	-0.066	-0.068	-0.076	-0.071
Industry code = 4	0.004	0.026	-0.031	-0.038
Industry code = 5	0.128	0.151	0.094	0.101
Industry code = 6	0.231	0.215	0.226	0.189
Industry code = 7	0.260	0.265	0.239	0.220
Industry code = 8	0.427	0.441	0.430	0.410
Industry code = 9	0.000	0.000	0.000	0.000
TIME_DUMMY=S_2015_1	-0.027	-0.021	-0.058	-0.008
TIME_DUMMY=S_2015_10	-0.178	-0.137	-0.208	-0.190
TIME_DUMMY=S_2015_11	-0.173	-0.202	-0.208	-0.162
TIME_DUMMY=S_2015_12	-0.373	-0.393	-0.392	-0.367
TIME_DUMMY=S_2015_2	-0.158	-0.153	-0.194	-0.146
TIME_DUMMY=S_2015_3	-0.277	-0.285	-0.310	-0.294
TIME_DUMMY=S_2015_4	-0.248	-0.273	-0.291	-0.248
TIME_DUMMY=S_2015_5	-0.259	-0.229	-0.285	-0.258
TIME_DUMMY=S_2015_6	-0.235	-0.217	-0.294	-0.249
TIME_DUMMY=S_2015_7	-0.225	-0.242	-0.242	-0.215
TIME_DUMMY=S_2015_8	-0.208	-0.186	-0.223	-0.198
TIME_DUMMY=S_2015_9	-0.147	-0.144	-0.179	-0.136
TIME_DUMMY=S_2016_1	-0.043	-0.078	-0.097	-0.043
TIME_DUMMY=S_2016_10	0.141	0.105	0.139	0.147
TIME_DUMMY=S_2016_11	0.028	0.020	0.009	0.045
TIME_DUMMY=S_2016_12	0.000	0.000	0.000	0.000
TIME_DUMMY=S_2016_2	-0.097	-0.117	-0.110	-0.102
TIME_DUMMY=S_2016_3	-0.072	-0.052	-0.099	-0.037
TIME_DUMMY=S_2016_4	0.096	0.089	0.062	0.125
TIME_DUMMY=S_2016_5	-0.017	-0.044	-0.059	0.000
TIME_DUMMY=S_2016_6	0.021	-0.014	-0.031	0.018
TIME_DUMMY=S_2016_7	0.103	0.095	0.088	0.117

TIME_DUMMY=S_2016_8	0.029	0.000	-0.026	0.044
TIME_DUMMY=S_2016_9	0.103	0.069	0.088	0.096
Network decison store dummy	0.000	0.000	0.000	0.000
Individual store dummy	0.201	0.207	0.179	0.213
Network independent store dummy	0.532	0.520	0.537	0.525
Annual revenue 1th order orthogonal polynomial	59.296	59.525	59.503	63.279
Annual revenue 2nd order orthogonal polynomial	-81.847	-80.740	-84.464	-78.947
Annual revenue 3rd order orthogonal polynomial	-24.489	-24.085	-24.322	-21.949

7.2. Card usage model average coefficient estimates

	County stratas	Industry stratas	Size stratas	Full database
Inverse Mills ratio	-0.713	-0.702	-0.714	-0.624
Logarith of store annual revenue	0.152	0.150	0.147	0.174
County = Bacs-Kiskun	-0.417	-0.336	-0.408	-0.475
County = Baranya	-0.094	-0.017	-0.077	-0.140
County = Bekes	-0.327	-0.263	-0.337	-0.388
County = Borsod-Abauj-	-0.136	-0.049	-0.145	-0.193
County = Budapest	0.350	0.435	0.360	0.297
County = Csongrad	-0.088	-0.020	-0.109	-0.122
County = Fejer	0.021	0.114	0.018	-0.040
County = Gyor-Moson-So	-0.118	-0.038	-0.116	-0.185
County = Hajdu-Bihar	-0.206	-0.125	-0.216	-0.253
County = Heves	-0.315	-0.235	-0.301	-0.372
County = Jasz-Nagykun-	-0.294	-0.202	-0.281	-0.334
County = Komarom-Eszte	-0.027	0.049	-0.011	-0.094
County = Mozgobolt	0.000	0.000	0.000	0.000
County = Nograd	-0.469	-0.404	-0.460	-0.509
County = Pest	-0.071	0.007	-0.053	-0.129
County = Somogy	-0.256	-0.179	-0.247	-0.332
County = Szabolcs-Szat	-0.484	-0.395	-0.452	-0.536
County = Tolna	-0.194	-0.110	-0.187	-0.261
County = Vas	-0.249	-0.201	-0.277	-0.334
County = Veszprem	-0.032	0.041	-0.017	-0.065
County = Zala	-0.209	-0.129	-0.214	-0.270

Number of bills = 1	0.000	0.000	0.000	0.000
Number of bills = 2	0.266	0.285	0.267	0.263
Number of bills = 3	0.406	0.409	0.400	0.407
Number of bills = 4	0.462	0.471	0.462	0.465
Number of bills = 5	0.512	0.518	0.505	0.504
Number of bills = 6	0.551	0.566	0.542	0.541
Number of bills = 7	0.617	0.626	0.597	0.572
Number of bills = 8	0.677	0.630	0.772	0.607
Industry code = 0	0.000	0.000	0.000	0.000
Industry code = 1	-0.225	-0.253	-0.309	-0.238
Industry code = 2	0.245	0.182	0.214	0.262
Industry code = 3	0.124	0.048	0.071	0.105
Industry code = 4	0.019	-0.021	-0.051	0.026
Industry code = 5	0.676	0.629	0.608	0.733
Industry code = 6	0.030	-0.055	-0.057	0.048
Industry code = 7	0.385	0.329	0.307	0.383
Industry code = 8	0.805	0.717	0.705	0.846
Industry code = 9	0.120	0.128	-0.039	0.143
Number of items	0.005	0.005	0.005	0.005
Transaction value 1th order orthogonal polynomial	-157.517	-167.243	-176.342	-184.446
Transaction value 2nd order orthogonal polynomial	-189.216	-194.939	-200.644	-206.922
Transaction value 3rd order orthogonal polynomial	-68.811	-71.832	-72.792	-76.017