Determinants of Consumers' Cash Expenditures in Poland in 2012: Evidence from Survey Data

Jan Lang Dominika Pawłowska

Paper prepared under the supervision of mgr. Aneta Dzik

University of Warsaw
Faculty of Economic Sciences

Warsaw 2013

[this page is intentionally left blank]

Contents

Part I: Introduction

Chapter 1: Introduction

Recent decades have brought about an immense development of payment methods alternative to cash. In Poland, the number of transactions involving payment cards increased from 54 million in 2000 to 844 million in 2010.[1] Naturally, the growing popularity of those instruments is a trend that has profound macroeconomic consequences.

First of all, that tendency contributes to the decreasing costs of issuance and operating of cash money. Electronic payments do not involve the material aspect of currency, which requires reproduction, transport, counting and storage. According to a paper published by the National Bank of Poland, those costs may amount up to 1% of the Gross Domestic Product in case of Poland.[2]

Meanwhile, studies have shown that the spread of electronic payments fosters economic growth by stimulating consumer spending.[3] Non-cash transactions allow for a decrease of costs of certain public services (e.g. decreasing the costs of paying pension benefits by transferring money directly to the bank accounts of the recipients).

The development of electronic payments has also been shown to reduce the scale of the shadow economy.[4] In fact, diminishing the black market and curbing money laundering is one of the main reasons behind the promotion of non-cash transactions by the public sector. It is estimated that the size of the shadow economy in Poland in 2009 was approximately 26% of the GDP. Therefore, even a partial reduction of this phenomenon could considerably increase fiscal revenues.[5]

Finally, the volume of cash and non-cash transactions is of particular interest to Poland due to perspective of entering the Eurozone; widespread usage of alternative means of payment would result in lower costs of producing banknotes and coins.

Thus, the costs and benefits associated with various types of transactions indicate that the consumers' and firms' choice of payment method is a crucial matter for the Polish economy. The goal of this paper is to identify and measure the impact of some of the determinants of cash usage among consumers.

The logic behind the choice of cash instead of other methods of settling payments is as following: the alternative forms may be aggregated in a single group as "non-cash" methods; therefore, knowing an individual's level of cash expenditures and overall expenditures will also allow one to quantify the level of usage of those alternative forms. However, the group of "non-cash" methods is quite diverse, with numerous instruments such as debit cards, online banking, or credit cards available to consumers. Hence, identifying and determining the value of transactions performed using all of the aforementioned means could be more problematic.

In literature, one may distinguish 2 groups of factors: those related to the transaction itself, and those associated with the consumer. This paper examines the influence of individual's characteristics, including demographic traits, on the intensity of usage of cash.

[1] Tochmański, Adam. *Problematyka opłaty interchange na rynku bezgotówkowych płatności kartowych w Polsce.* Narodowy Bank Polski

[2] *Analiza funkcjonowania opłaty interchange w transakcjach bezgotówkowych na rynku polskim.* Narodowy Bank Polski. p. 15

[3] *The Virtuous Circle: Electronic Payments and Economic Growth*, Visa Global Insight. p.10

[4] *Analiza funkcjonowania opłaty interchange w transakcjach bezgotówkowych na rynku polskim.* p.15

[5] Ibid.

Chapter 2: Hypotheses

Perhaps one of the most important factors we consider to have effect on intensity of usage of cash is the age of an individual. Elderly individuals are more likely to use cash due the fact that they are accustomed to this method of payment and might be reluctant to shift their habits. On the other hand, intuitively, younger individuals usually display a higher degree of interest in newer technologies, with which the more innovative payment instruments are associated.

Therefore, it is expected that age will have a positive impact on cash expenditures. It remains to be verified whether younger individuals in fact use cash more frequently in light of limited access to such forms of payments as credit cards; in this case, the relationship would rather be U-shaped than linear, and the variable would be squared. This will be verified by statistical means further on.

Another element that may play a significant role is the individual's income. It is expected that individuals with higher income exhibit a higher level of awareness with respect to various possibilities of storing their current savings (which they plan to spend), and will more frequently resort to electronic payments in order to maximize the interest earned on their income. Furthermore, for such individuals, low-value transactions, which oftentimes may be settled only in cash, will contribute a smaller portion to their expenditures.

Apart from income, the principal source of income is also taken into account. Entrepreneurs are expected to use cash the least due to strong familiarity with other payment instruments obtained in course of work. Individuals who receive a constant salary (mostly employees) are also expected to use cash less, as receiving regular payments may be more convenient by having them directly transferred to a bank account; hence, such individuals are deemed more likely to use payment cards. Cash is expected to be more popular among the remaining groups (retirees, recipients of social benefits, part-time workers, and students financially dependent on their relatives).

An important issue that also needs to be considered is the discrepancy in availability of different payment methods (payment card acceptance) between smaller and larger cities. Such an inconsistency, if true, may imply that inhabitants of smaller cities and towns prefer cash settlements.

The education level attained by an individual might also have influence on his preferences concerning payment methods; the reasons this may, however, lie beyond mere education. One may suspect that among individuals with primary or lower secondary education, who receive income, there is a relatively larger group of unregistered employees. Hence, cash spending among those groups could be higher.

The remaining two characteristics that were taken into account are marital status and gender. One may suspect that married individuals will show a stronger tendency towards the usage of non-cash instruments, as married couples may be more inclined to use a common account as means of synchronizing their personal finance.

As for gender, women are often responsible for their household's day-to-day expenditures, which are more frequently covered by cash. This could be the case especially for females with lower levels of education - therefore, interactions between gender and education will have to be taken into account.

Chapter 3: Overview of the existing literature

As mentioned in the introduction, there are two major sets of factor affecting consumers' choice of payment that are described in the existing literature on the topic: the characteristics of individual transactions and the traits of consumers themselves. Nevetheless, the two oftentimes are found to be intertwined.

A discussion paper by Carlos Arango, Dylan Hogg and Alyssa Lee *Why Is Cash (Still) So Entrenched? Insights from the Bank of Canada's 2009 Methods-of-Payment Survey* published by the Bank of Canada. For the purposes of the research, data was collected in two types of survey. One of them was a questionnaire regarding respondents' personal finances, their incentives and habits in the use of different payments methods; 6,800 participants completed the form. Another group (3,500 individuals) was asked to keep a 3-day diary where data describing the transactions which they have conducted was included.

The research has shown that generally, elderly individuals use cash more frequently, while there is a negative correlation with income. The authors explain the latter by the fact that more comfortable payment instruments, such as credit cards, are available to those with better financial standing. Furthermore, the article emphasizes the role of transaction size in this case: high-income consumers may tend to make more high-valued transactions, where payment with card is more convenient, while for lower-income consumers, smaller-value purchases may constitute a larger share of expenditures, they may shop more often at locations where only cash is accepted, and thus become more reliant on cash.

Interestingly, the paper shows cash is frequently chosen by consumers even at locations where debit and credit cards are accepted.

Concluding, the paper finds that it is primarily the ease of use, speed, lack of acceptance of alternative instruments and high availability of cash that determines its usage. This could explain why cash payments are concentrated in the lower-value transactions, with about 70% of the payment volume for transactions below $25 being handled using cash. Finally, the consumer perception of cash has been shown to play an important role, with some consumers recognizing it as a simple tool for controlling spending.

Another research was conducted by Michael Cohen and Marc Rysman based on data collected in the United States, the results of which are presented their paper *Payment choice with Consumer Panel Data* (2012). Data used in the research has been drawn from a dataset maintained by the A.C. Nielsen Company, and was created by households who were asked to scan their grocery shopping for 3 years. The number of households involved in this broad survey was 13,574, and a total of 1.34 million transactions were recorded. The database distinguishes between three methods of payment: cash, cards and checks.

Both demographic and shopping-trip factors were taken into account. The research has shown that wealthier households tend to rely more heavily on cards; similarly, a higher education level entails more intensive card usage. Male have been shown to spend on average 5.2 pp. more using cash, although the standard deviations of the results were relatively large. Nevertheless, it was the expenditure size that was found to be the major determinant behind the consumers' choice of payment method.

The same research also found that consumers were generally unwilling to switch between payment methods.

Yet another research was conducted by Koulayev at al. (*Explaining adoption and use of payment* *instruments by U.S. consumers*, 2012) using a dataset from the Survey of Consumer Payment Choice; in total, 997 observations were used. The survey distinguished between 8 methods of payments. The research has shown that men are more likely than women to use debit and credit (on the contrary to what the previously described research suggested), higher educated people more frequently resort to credit cards, while employed individuals are less likely to use the former method (possible due the lack of need of credit).

Part II: Data and Estimation

Chapter 4: Database description and variables

The database used for the purposes of the model consists of 226 observations obtained by means of a survey. A questionnaire that was used in the survey can be found attached in the Appendix 1 to this paper. Initially, 227 observations had been gathered. One observation, due to lack of substantial data, had to be removed from the dataset.

The software used for the estimation of the model is Stata/IC 12.0

The dependent variable in our model, y_i (named "cash" in Stata), is the share of expenditures made using cash in total expenditures of the i-th individual. It is can be therefore represented as the following fraction:

$$y_i = \frac{z_{1i}}{z_{2i}}$$

where z_{1i} is the i-th individual's cash spending ("cash_exp"), and z_{2i} ("expenditures") - his or her total expenditures. The dependent variable, defined in this manner, is continuous and lies in the interval $[0, 1]$.

The independent variables in the model are as following:

- x_1 ("age") is the age of the respondent. It is a continuous variable;
- x_2 ("income") is the average monthly income of the respondent, in Polish złoty. It is a continuous variable;
- x_3 ("sex") is the gender of the respondent. It is a discrete variable which takes on value 1 for females and 2 for males;
- x_4 ("mstatus") is the marital status of the respondent. It is a discrete variable which takes on value 1 for singles, 2 for married inviduals, 3 for divorced individuals and 4 for widows and widowers;
- x_5 ("city") is the size of the city where the respondent lives. It is a discrete variable that takes on value 1 for cities/towns with population below 50,000, 2 for those with population in the interval $[50\,000, 100\,000)$, 3 for the interval $[100\,000, 250\,000)$, 4 for the interval $[250\,000, 500\,000)$, and finally, 5 for cities with population above 500 000;
- x_6 ("edu") is the level education attained so far by the respondent; it is a discrete variable. Value 1 corresponds to elementary education, value 2 to lower secondary education, value 3 to secondary education and value 4 to tertiary education;
- x_7 ("source") is a discrete variable which indicates the principal source of income of the respondent. It was designed to take on value 1 ("A") for university students, who still rely financially on their parents, value 2 ("B") for individuals receiving salary (based on an employment contract), value 3 ("C") for individuals employed using other (generally, more flexible) forms of contracts, value 4 ("D") for those who run a business, value 5 ("E") for recipients of social benefits (such as unemployment benefits), value 6 ("F") for those who receive a non-retirement related pension (for example, a disablity pension), value 7 ("G") for retirees , or value 8 ("H") to account for other, less frequently occuring sources of income (self-employment, capital earnings etc).

Chapter 5: Summary statistics and statistical analysis of data

5.1 Continuous variables

Summary statistics of the continuous variables in the model (the independent variable, income and age) are presented below. Mean, standard deviation, as well as minimum and maximum values of the variables are included. At this stage, 226 observations from the dataset were taken into account.

Tab. 1: Continuous variables

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
age	29.601	13.836	15	82
income	2457.593	2553.888	100	20000
cash	0.502	0.281	0.01	1
N		226		

Figure 1 at the right presents the histogram of age of the respondents. Numbers range from 19 to 82, which indicates that nearly all age groups to which alternative payments methods (i.e. credit/debit cards, online banking) are available have been included in the survey. The distribution is clearly asymmetrical, which is caused by the fact that most of the respondents were between 20 and 30 years old.

Variable "age" was tested for normality. Both the Jarque-Bera (χ^2 statistic equal to 51.10) and Shapiro-Wilk test and yielded p-values very close to 0, which meant the rejection of the null hypothesis that the variable is normally distributed even for very low significance levels (and most certainly so for $\alpha = 0.05$).

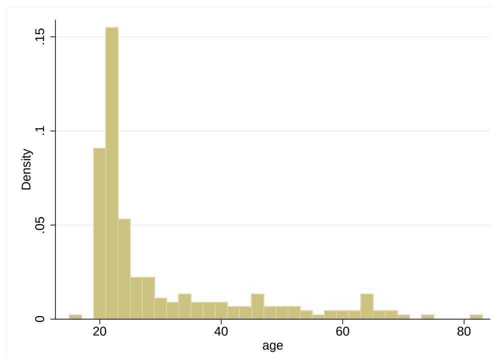


Figure 1: Age of the respondents

The histogram of the income of respondents is presented below on the left side, in Figure 2.

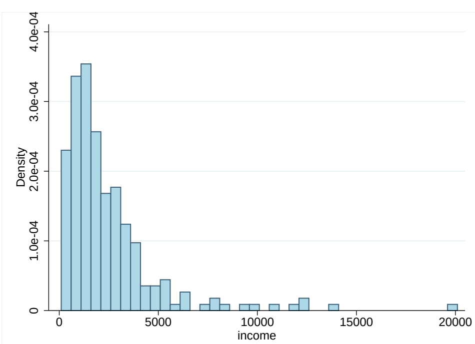


Figure 2: Income of the respondents

Incomes range from very low levels (100) to relatively high (20,000), which means that most income levels are represented. A majority of the observations are below 5,000; the median of income is 1800.

Variable "income" was also tested for normality with both the Jarque-Bera ($\chi^2 = 178.87$) and Shapiro-Wilk tests. The results were similar, with both tests yielding p-values very close to 0, which meant the rejection of the null hypothesis that the variable is normally distributed (for $\alpha = 0.05$).

Finally, Figure 3 on the next page depicts the histogram of the dependent variable itself - portion of individual's expenditures made using cash ("cash"), which, as previously explained, ranges from 0 to 1.

The distribution appears to be rather atypical - the shape of the histogram indicates that this might be a case of a multimodal distribution, possibly a bimodal one. What requires attention is certainly the very high number of results that are equal to 1. Nonetheless, this fact has a very natural explanation - these are the observations of individuals who spend solely using cash. In contrast, the very low number of respondents who spend *almost* entirely cash can probably be explained with the following reasoning - if an individual makes the effort to obtain a credit/debit card (or make available to him/herself any other alternative forms of payment), it is not for making only insignificant, minute purchases.

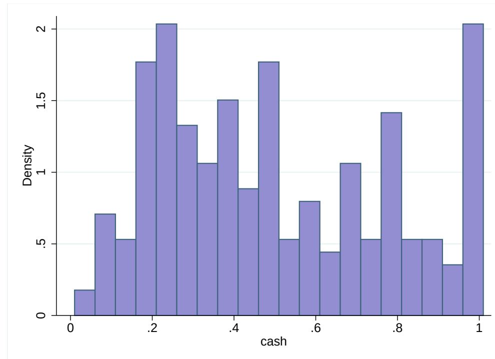


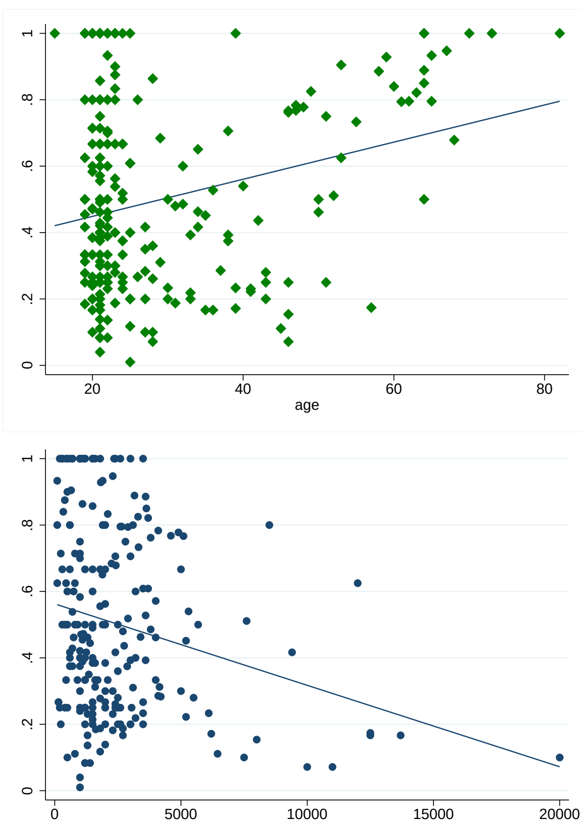
Figure 3: Cash expenditures

Although there is strong graphical evidence against the normality of its distribution, variable "cash" was tested for normality using the same tests as the previous two variables. The Shapiro-Wilk test yielded a p-value equal to 0.00001, which means that for $\alpha > 0.00001$ we reject that null hypothesis asserting normality of the distribution; Jarque-Bera test ($\chi^2 = 44.05$) produced an even lower p-value. Thus, assuming significance level equal to 0.05, we conclude that the variable is not normally distributed.

As a final point, the correlation between continiuous independent variables and the dependent variable was analyzed. Due to the conclusions of the normality tests conducted above, Spearman's rank correlation coefficient, ρ_S, was used as a measure.

Figure 4, on the upper-left side, presents the scatter plot and the regression line of the explained variable and age of respondents. The value of the correlation coefficient was $\rho_S = 0.1117$; it was not significant for significance level $\alpha = 0.05$, although it was for $\alpha = 0.1$. The positive value of the coefficient adds credibility to the claim that the share of cash expenditures in total expenditures increases with age.

Figure 5, below Figure 4, contains the scatter plot and the regression line of the explained variable and individual's income. In this case, $\rho_S = -0.18$ and the outcome was significant for $\alpha = 0.05$. This result (both the sign and magnitude of the coefficient) indicates that there could be a moderately strong, inverse relationship between the share of cash expenditures and one's income.



Figures 4, 5: Scatter plots of age and income on y

5.2 Explained variable in subgroups

Summary statistics of discrete variables (in separate tables) are available in Appendix 2. This section provides an analysis of the dependent variable within subgroups introduced by the discrete variables in the model.

Sex

The table below provides a summary of the dependent variable separately for respondents of different genders.

Tab. 2: Sex

Sex	Obs	Mean	Std. Dev.	Min.	Max.
Female	115	0.511	0.291	0.04	1
Male	111	0.494	0.272	0.01	1

The dependent variable was tested for normality in each subgroup using the Shapiro-Wilk test.[6] In case of females, p-value was equal to 0.034, while for males it was equal to 0.034. Hence, assuming $\alpha = 0.05$, the null hypothesis asserting normality was rejected for both. As result, in order to compare the distribution of y in subgroups, the non-parametric Mann-Whitney-Wilcoxon (MWW) test was chosen. The test yielded p-value equal to 0.7677, which, at significance level of 0.05, does not allow for the rejection of null hypothesis claiming that the distributions in the subgroups are the same. Therefore, one might expect that the variable "sex" may occur to be insiginificant in the regression.

Marital status

The table below provides a summary of the dependent variable separately for respondents with different marital statuses:

Tab. 3: Marital status

Mstatus	Obs	Mean	Std. Dev.	Min.	Max.
Single	161	0.4915227	0.277239	0.01	1
Married	49	0.5169387	0.2951419	0.0714286	1
Divorced	8	0.4030566	0.1846729	0.0714286	0.625
Widow/Widower	8	0.7306761	0.2862474	0.173913	1

Similarly as before, the explained variable was tested for normality in the above subgroups (with the same test). The obtained p-values were, respectively, 0.00010, 0.00109, 0.48492 and 0.15838. With $\alpha = 0.05$, we reject the null hypothesis for the first two cases; however, we cannot reject the normality of distribution of y among divorced individuals and widows/widowers. Nevertheless, one has to bear in mind the small sample size of the two former. In view of those results, the non-parametric Kruskal-Wallis one-way analysis of variance was implemented to test whether the samples originate from the same distribution. The value of χ^2 statistic was equal to 5.081, and p-value was equal to 0.1660; the value of χ^2 statistic with ties was equal to 5.090, while p-value to 0.1654. In both of those cases, assuming the same significance level as previously, the test is not significant and therefore, there is no evidence of difference bewteen all of the samples. Hence, on might expect that the variable will also be insignificant in the regression.

[6] Jarque-Bera test will not be used in this section due to relatively small sample sizes.

City

A summary of y for respondents from cities of different sizes is presented in Table 4 below:

Tab. 4: City

City	Obs	Mean	Std. Dev.	Min.	Max.
Below 50,000	40	0.5918057	0.2717248	0.01	1
50,000 to 100,000	21	0.5718575	0.3173735	0.1363636	1
100,000 to 250,000	23	0.5749815	0.3380754	0.1538462	1
250,000 to 500,000	18	0.5418463	0.2282505	0.173913	0.7954546
Above 500,000	124	0.4425482	0.2630024	0.0714286	1

The same procedure as in the previous set of subgroups was repeated. The results of Shapiro-Wilk tests were, respectively, p-values of 0.61400, 0.26129, 0.24700, 0.02636 and 0.00004. Thus, for $\alpha = 0.05$, hypothesis stating normality cannot be rejected for cities/town with population below 50 000, $[50\,000, 100\,000)$ and $[100\,000, 250\,000)$ and is rejected for the other two. Again, the Kruskal-Wallis test was utilized to check whether the samples originate from the same distribution. The value of χ^2 statistic with/without ties was 12.484/12.463, and the appropriate p-values were 0.0141/0.0142. Hence, assuming the same α as previously, the test is significant, leading to the conclusion that there is a statistically significant difference between the samples.

It is important to remember that the Kruskal-Wallis test does not identify how many samples are different; the result indicates that there is at least one sample which is significantly different from the rest.

Education

Table 5 contains the summary of y for each education level attained by the respondent:

Tab. 5: Education level

Edu	Obs	Mean	Std. Dev.	Min.	Max.
Elementary	2	1	0	1	1
Lower secondary	1	1	-	1	1
Secondary	122	0.5289178	0.2777531	0.04	1
Tertiary	101	0.45551541	0.2726022	0.01	1

Observations with elementary and lower secondary education were not tested due to severely small sample size. For secondary and tertiary levels, Shapiro-Wilk test yielded p-values of 0.00754 and 0.00137, thereby leading to the rejection of the null hypothesis concerning the normality of data in both cases. The MWW test was subsequently conducted; the resulting p-value was equal to 0.0524, which means that at the significance level of 0.1, the null hypothesis claiming that the samples originate from the same distribution cannot be rejected. This, combined with the minuscule sizes of the other two subgroups, indicates that the variable may be insignificant in the regression.

Source of income

Table 6 presents the summary of explained variable for groups with the different principal source of income (an full explanation of categories is available in chapter 4):

Tab. 6: Income source

Source	Obs	Mean	Std. Dev.	Min.	Max.
(A) Non-working student	91	0.5089099	0.2852225	0.04	1
(B) Employment contract	61	0.4844254	0.2485638	0.1	1
(C) Employment (other)	33	0.4183311	0.2484076	0.0833333	1
(D) Business	13	0.3747441	0.2895448	0.0714286	1
(E) Social benefits	1	0.9047619	0.9047619	0.9047619	0.9047619
(F) Pension (non-ret.)	10	0.6119825	0.2708415	0.3125	1
(G) Pension (ret.)	10	0.8743617	0.1689397	0.5	1
(H) Other	7	0.461352	0.3521836	0.01	1

Categories (A)-(D), (F) and (G) were tested for normality, the Shapiro-Wilk test for (A), (B), (C) and (G) yielded p-values below 0.05; for (D) and (F) those were 0.10906 and 0.34386 respectively (hence, we cannot reject the null hypothesis concerning normality for the two latter). The χ^2 statistic with/without ties in Kruskal-Wallis test of the mentioned categories was equal to 22.676/22.638, yielding p-values of 0.0004 in both cases. Hence, assuming the same α as previously, we reach a conclusion that there is a significant difference between the samples.

5.3 Interactions

Initially, the product of variables “sex” and “education” was included in the model to account for a possible interaction bewteen them. Nonetheless, initial regression estimates showed that the the resulting variable is insignificant, hence it was subsequently dropped.

Chapter 6: Selecting a functional form

Box-Cox transformation was used to determine whether and which variables in the model should be logarithmically transformed. First, a test was conducted for the dependent variable. The obtained estimate of the parameter in Box-Cox transformation was equal to $\hat{\theta} = 0.5426404$, which is closer to 1 and thus indicates that the dependent variable should not be transformed. It is noteworthy though that the null hypotheses $H_0 : \theta = -1$, $H_0 : \theta = 0$ and $H_0 : \theta = 1$ are all rejected, with p-values virtually equal to 0.

Variable "income" was tested as well. In this case though, Box Cox transformations produced an estimate of $\hat{\lambda} = 0.259779$, which is quite close to 0 and indicates tha the variable should be logarithmically transformed. Furthermore, the hypothesis that $\lambda = 0$ cannot be rejected for any significance level below 0.281.

Finally, variable "age" was tested. Box-Cox transformation yielded an estimate of $\hat{\lambda} = 2.208335$, which indicates that the variable should be raised to the power of 2. Assuming that the coefficient in the regression would be positive, this result could suggest that middle-aged individuals - proportionately to their expenditures - use cash the least. A graphical analysis of the variable indicates that this claim could be sound. Furthermore, such a relationship would have a reasonable explanation - cash usage could be less popular among young individuals due to limitations in access to certain alternative payment instruments (such as credit cards), while the elderly could simply be unwilling to accustom to newer methods of handling payments.

Nevertheless, regression results showed that a squared variable had failed to provide a better fit. In fact, both the R^2 and adjusted R^2 were slightly lower for the regression with the squared variable (0.2803 and 0.2101, respectively, versus 0.2809 and 0.2108). Additionally, a squared variable does not offer such a clear interpretation as is the case in simple, linear relationship. Therefore, it was decided that the variable "age" will remain untransformed, though bearing in mind that perhaps with a broader dataset, the numerical results of the regression would have spoken in favor of incorporating the squared variable instead.

To conclude, the model that will be estimated in the next chapter will have the following form:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 \ln x_{2i} + \beta_3 x_3 + \sum_{j=2}^{4} x_{4j,i} + \sum_{j=2}^{5} x_{5j,i} + \sum_{j=2}^{4} x_{6j,i} + \sum_{j=2}^{8} x_{7j,i} + \varepsilon_i$$

where the j indices of discrete variables result from their decoding into binary variables (by categories), β_0 is a constant, ε_i is the error term, and $i = 1, 2, ..., 226$. Full explanation of variables used can be found in chapter 4.

Chapter 7: Regression

The results of regression ran for the model described in the previous chapters are presented below:

Source	SS	df	MS
Model	4.9449644	18	0.274720244
Residual	12.8716145	207	0.062181713
Total	17.8165789	225	0.079184795

Number of obs	=	226
F(18, 207)	=	4.42
Prob > F	=	0.0000
R-squared	=	0.2775
Adj R-squared	=	0.2147
Root MSE	=	0.24936

Cash	Coef.	Std. Dev.	t	P>|t|	[0.95 Conf.	Interval]
age	0.0089831	0.0026645	3.37	0.001	0.00373	0.0142361
ln-income	-0.0870481	0.027443	-3.17	0.002	-0.1411517	-0.0329445
-Isex-2	0.0177335	0.0352064	0.50	0.615	-0.0516755	0.0871426
-Iedu-2	0.2191012	0.3239475	0.68	0.500	-0.4195582	0.8577605
-Iedu-3	-0.1615485	0.1988015	-0.81	0.417	-0.5534838	0.2303868
-Iedu-4	-0.2049444	0.2010968	-1.02	0.309	-0.6014048	0.1915161
-Imstatus-2	-0.1391402	0.0667549	-2.08	0.038	-0.2707469	-0.0075336
-Imstatus-3	-0.1282464	0.105493	-1.22	0.226	-0.3362367	0.079744
-Imstatus-4	-0.1486742	0.1233559	-1.21	0.229	-0.391883	0.0945347
-Icity-2	-0.0159737	0.0730014	-0.22	0.827	-0.1599037	0.1279562
-Icity-3	-0.0735402	0.0717522	-1.02	0.307	-0.2150071	0.0679266
-Icity-4	0.0070966	0.076565	0.09	0.926	-0.1438592	0.1580524
-Icity-5	-0.0984876	0.0474727	-2.07	0.039	-0.1920849	-0.0048903
-Isource-2	0.0441228	0.0644096	0.69	0.494	-0.0828673	0.1711128
-Isource-3	-0.0150693	0.055461	-0.27	0.786	-0.1244165	0.0942778
-Isource-4	0.0045516	0.1016657	0.04	0.964	-0.1958928	0.2049961
-Isource-5	0.1733757	0.2707905	0.64	0.523	-0.3605159	0.7072672
-Isource-6	0.0354362	0.0892049	0.40	0.692	-0.1404404	0.2113129
-Isource-7	0.2046131	0.1302887	1.57	0.118	-0.0522646	0.4614908
-Isource-8	-0.0245669	0.0991168	-0.25	0.804	-0.2199858	0 .1708521
-cons	1.167945	0.276029	4.23	0.000	0.6237251	1.712164

Although all of the variables in the model are jointly significant (p-value of the F-test is almost equal to 0), there are individual variables which are insignificant. P-values of t-test for significance of a variable which are above 0.1 have been marked red. All categories of variables “sex”, “edcuation” and “source” have p-values above an acceptable level; hence, those variables were dropped.

City and marital status of an individual both have a category for which the p-value is below 0.05 and therefore, the binary variables into which those discrete variable had been decoded were tested for joint significance. The p-values produced by the test were, respectively, 0.1636 and 0.2346 - unfortunately, too high reject the null hypothesis asserting their joint insignificance (for an acceptable significance level). Similarly, the same test performed simultaneously on “city” and “mstatus” yielded a comparable p-value: 0.1629. Therefore, those variable were dropped too.

The results of the final regression, without the variables that have been dropped, are presented below:

Source	**SS**	**df**	**MS**
Model	3.78331535	2	1.89165767
Residual	14.0332636	223	0.062929433
Total	17.8165789	225	0.079184795

Number of obs	=	226
F(18, 207)	=	30.06
Prob > F	=	0.0000
R-squared	=	0.2123
Adj R-squared	=	0.2053
Root MSE	=	0.25086

Cash	**Coef.**	**Std. Dev.**	**t**	**P>|t|**	**[0.95 Conf.**	**Interval]**
age	0.0092564	0.0013445	6.88	0.000	0.0066069	0.0119059
ln-income	-0.1243127	0.0199819	-6.22	0.000	-0.1636901	-0.0849352
-cons	1.149482	0.1364667	8.42	0.000	0.8805527	1.418411

As a result of the removal of the mentioned variables, adjusted R^2 slightly decreased (by 0.055). Standard errors for the remaining variables have also decreased. At this stage, all variables in model are significant, both jointly and individually.

Chapter 8: Potential problems with the database

8.1 Multicollinearity

In order to determine the severity of multicollinearity in the model, the variance inflation factor (VIF) was utilized. The indices it provided for variables age and the natural logarithm of income were both equal to 1.24. Since this value is relatively low (considerably below 10), it can be concluded that multicollinearity is not an issue in this model.

8.2 Unusual observations

The following section provides an overview and explanation of unusual obervations in the database which was used for estimating the model.

For the purpose of analyzing the database in context of validity of the observations, their leverages and normalized residuals were calculated; the results are presented on a scatter plot on the right.

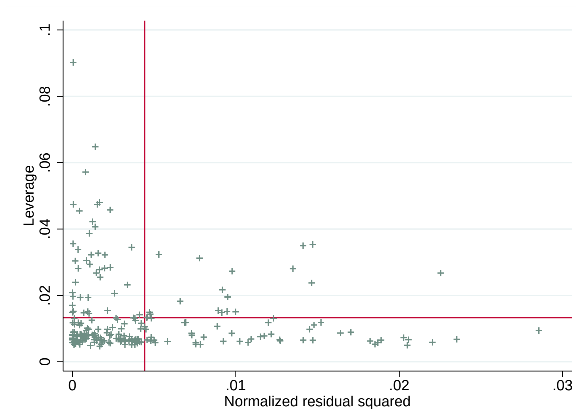


Figure 6: Levarage and normalized residuals squared

As one may notice, none of the observations occurs in the far upper-right corner of the scatter plot (i.e. there are no observations with exteremely high values of both leverage and normed residuals squared); this indicates that possibly none of the atypical observations (those with high values of leverage) will substantially influence the results of the regression.

In order to support the conclusions of graphical analysis, highly influential observations were detected numerically as well. Those for which the value of leverage is higher than $2K/N = 0.01769912$ and, at the same time, the absolute value of normalized residuals is higher than 2, were listed. Only one observation, summarized below, satisfied the two conditions:

Tab. 7: Atypical observations

Obs.	cash	age	ln-income	leverage	norm. resid.	Cook dst
224	0.8	23	9.047821	0.0267459	2.272424	0.0473031

The Cook distance for observation no. 224 is equal to 0.0473031, which is considerably higher than $4/N = 0.01769912$. This observation consists of a 23-year-old individual, with an income of 8,500 PLN, who uses cash to cover, on average, 80% of his monthly expenditures. Although his income may seems unusually high, it not impossible, and therefore the observation cannot be removed from the dataset on grounds of being unrealistic. Since it does violate the underlying assumptions of the model, it was deemed valid and remained in the dataset.

Chapter 9: Diagnostic tests

This chapter provides a discussion on the results of diagnostic tests conducted for the purpose of checking the correctness of the functional form of the model, as well as for determining whether some of the key assumptions behind the properties of the OLS estimator are satisfied.

The RESET Test

The Ramsey Regression Equation Specification Error Test (RESET) was used to determine whether the functional form of the model is correct. Given the null hypothesis that the model has no omitted variables, the value of the resulting F-statistic was $F(3, 220) = 0.68$, yielding a rather high p-value of 0.5638. Therefore, at a significance level of $\alpha = 0.05$ or even $\alpha = 0.1$, the test indicates that the functional form of the model is indeed acceptable.

The Breusch-Pagan Test

The Breusch-Pagan test was implemented to check whether the error terms in the model exhibit constant variance. With the null hypothesis asserting homoscedasticity, the χ^2 statistic assumed the value of 0.36. The test yielded a p-value of 0.5503, which means that the null hypothesis cannot be rejected for a $\alpha = 0.05$ or even $\alpha = 0.1$.

The White Test

One must remember that homoscedasticity is an important assumption, based on which the OLS estimator is proven to the best linear unbiased estimator. In case of existence of heteroscedastic disturbances, it may lose its efficiency.

In order to cross-check the results of the previous test, the White test was used. Given the null hypothesis stating homoscedasticity, against the alternative hypothesis of heteroscedasticity, the test yielded a p-value of 0.0748, with the value of χ^2 statistic equal to 10.01. Thus, at $\alpha = 0.05$, the White test confirmed the result of the previous test.

Analysis of Residuals

The final part of this chapter is the analysis of residuals. Figure 7 on the next page depicts the histogram of residuals (versus normal distribution), a box plot, a quantile graph and a probability graph.

The shape of the distribution of residuals clearly differs from that of normal distribution, which is visible on the histogram. The quantile graph illustrates the behavior of tails, which varies noticeably between the two; the probability graph shows that there are significant deviations from the values originating from the normal distribution.

In order to confirm the inferences drawn from the graphical analysis, the Jarque-Bera test was conducted to test the residuals for normality. The resultant p-value of 0.0148 indicates that for $\alpha = 0.05$, the null hypothesis asserting normality must be rejected.

Nonetheless, for sufficiently large samples, this outcome is not severe in consequences due to the fact that distributions of statistics in such samples are close to standard distributions, even if the normality assumption does not hold.

One of the effects of lack of normality is, though, that the prediction intervals would be incorrect.

It is noteworthy that the Jarque-Bera test used in the above analysis oftentimes rejects the null hypothesis for large samples, having detected even minor deviations from the normal distribution.

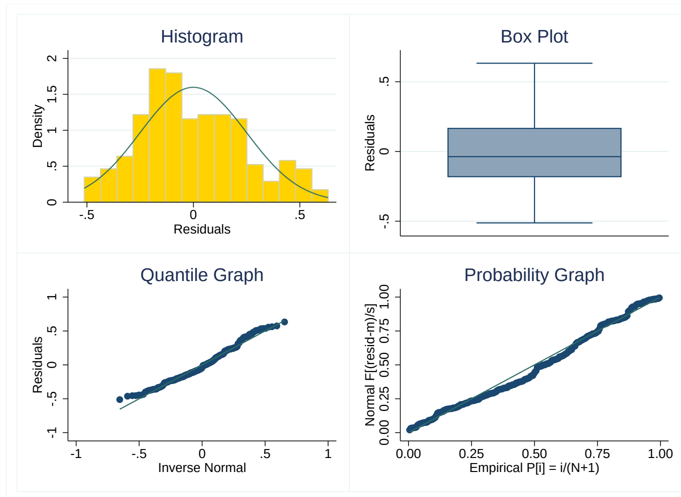


Figure 7: Graphical Analysis of Residuals

Part III: Interpretation and Conclusions

Chapter 10: Interpretation of parameters

For age, the value of the estimated parameter was $b_1 = 0.0092564$. The interpretation in this case is rather trivial: on average, an individual who is one year older will spend 0.0092564 percentage points more using cash.

In case of the variable income, this situation is not as straightforward due to the fact that it had been log-transformed, while the explained variable, y, remained in an unaltered form. Nevertheless, the obtained value of the estimated parameter of β_2, $b_2 = -0.1243127$ does indeed have an interpretation: irrespective of the base value of income, a fixed percentage change of income will result in the same shift of the explained variable. For example, a 1% increase in come will cause on average a 0.001236952494 pp. decrease in cash expenditures, a 5% increase in income will cause on average a 0.006065237041 pp. decrease in cash expenditures, and a 10% increase in income will result in a 0.01184826579 pp. decrease of cash expenditures.

Whenever refering to the results of the regression, one must nonetheless bear in mind the constraints placed on the dependent variable in the model: the value of y cannot be lower neither than 0 nor higher than 1.

As for the measure of fit, $R^2 = 0.2123$, which means that 21.23% of variation of the dependent variable is explained by the model. The adjusted R^2 is not significantly lower, with $\bar{R}^2 = 0.2053$.

Chapter 11: Conclusions

As expected, age occurred to have positive impact on the intensity of cash usage. What remains to be verified is whether the relationship is indeed linear, or due to increased popularity of this instrument among the young, is in fact U-shaped. Perhaps a broader dataset could produce a more decisive answer.

Income, in accordance with the anticipated results, turned out to have a negative impact on the dependent variable.

The two results may encourage one to seek further, more fundamental reasons behind payment trends. One of the explanations provided by Arango, Hogg and Lee is that alternative methods of payment are dominant for high-valued transactions; since high-valued transactions comprise a greater portion of expenditures of wealthy individuals, the implication is that higher income entails smaller dependence on cash. Indeed, Cohen and Rysman support the claim that the expenditure size plays a crucial role in the choice of payment method among consumers.

Similarly, one may be willing to examine whether age is related to specific perceptions of various payment instruments.

Other variables cannot be interepreted in context of the regression, because they were recognized as insignificant. However, this does not imply that there is no relationship between some of them and the explained variable. For example, the mean values of dependent variable in the individual city categories suggest that there is indeed a positive relationship. Again, perhaps with a broader dataset, the regression results would have supported such a hypothesis.

Bibliography

1. *Analiza funkcjonowania opłaty interchange w transakcjach bezgotówkowych na rynku polskim.* Narodowy Bank Polski Departament Systemu Płatniczego. Warsaw, January 2012.

   <http://www.nbp.pl/systemplatniczy/obrot_bezgotowkowy/interchange.pdf>

2. Carlos Arango, Dylan Hogg and Alyssa Lee. *Why Is Cash (Still) So Entrenched? Insights from the Bank of Canada's 2009 Methods-of-Payment Survey.* February 2012.

   <http://www.bankofcanada.ca/wp-content/uploads/2012/02/dp2012-02.pdf>

3. Cohen, Michael and Marc Rysman. *Payment choice with Consumer Panel Data.* Boston University. November 5, 2012.

   <http://people.bu.edu/mrysman/research/grocerypayment.pdf>

4. Koulayev, Sergei et al. *Explaining Adoption and Use of Payment Instruments by U.S. Consumers.* February 26, 2012.

   <http://www.econ.uzh.ch/agenda/seminars/amss/Rysman.pdf>

5. Mycielski, Jerzy. *Ekonometria.* Wydanie 3, poprawione. 2010.

6. *The Virtuous Circle: Electronic Payments and Economic Growth.* Global Insight, Inc. June 2003.

   <http://www.visacemea.com/av/pdf/eg_virtuouscircle.pdf>

7. Tochmański, Adam. *Analiza funkcjonowania opłaty interchange w transakcjach bezgotówkowych na rynku polskim.* February 2012.

   <http://www.nbp.pl/aktualnosci/wiadomosci_2012/zespol_interchange.pdf>

Appendix 1: Survey

---

1. Please indicate your gender (underline the correct response): FEMALE/MALE
2. Please indicate your marital status:
   (a) Single
   (b) Married
   (c) Divorced
   (d) Widowed
3. What is your age?
4. Please indicate the size of the population of the town/city in which you live:
   (a) Below 50,000
   (b) 50,000 - 100,000
   (c) 100,000 - 250,000
   (d) 250,000 - 500,000
   (e) Above 500,000
5. How high is your average monthly income (in Polish złoty)? (**Note:** if your income has been constant for at least two months, please give the most accurate value)
6. How high are your average monthly expenditures (in Polish złoty)?
7. On average, how much cash do you spend on a monthly basis (in Polish złoty)?
8. Which level of education have you completed?
   (a) Elementary
   (b) Lower secondary
   (c) Secondary
   (d) Tertiary

9. What is your main source of income?

    (a) I do not work, I am a student
    (b) Salary (based on an employment contract)
    (c) Salary (other than b)
    (d) Business/enterprise
    (e) Social benefits
    (f) Pension (non-retirement)
    (g) Pension (retirement)
    (h) Other.

Appendix 2: Discrete variables

Summary statistics of discrete variable included in the model are presented below.

Gender of the respondents:

Tab. 8: **Sex**

sex	Freq.	Percent
female	115	50.88
male	111	49.12
Total	226	100.00

Marital status of the respondents:

Tab. 9: **Marital Status**

mstatus	Freq.	Percent
Single	161	71.24
Married	49	21.68
Divorced	8	3.54
Widow/Widower	8	3.54
Total	226	100.00

Population of the city/town where the respondent lives:

Tab. 10: **City**

city	Freq.	Percent
Below 50k	40	17.70
50k - 100k	21	9.29
100k - 250k	23	10.18
250k - 500k	18	7.96
Above 500k	124	54.87
Total	226	100.00

Education level attained by the respondent:

Tab. 11: **Education**

edu	**Freq.**	**Percent**
Elementary	2	0.88
Lower Secondary	1	0.44
Secondary	122	53.98
Tertiary	101	44.69
Total	226	100.00

Principal source of income of the respondent (for explanation of categories used see Chapter 4):

Tab. 12: **Source of income**

source	**Freq.**	**Percent**
(A)	91	40.27
(B)	61	29.99
(C)	33	14.60
(D)	13	5.75
(E)	1	0.44
(F)	10	4.42
(G)	10	4.42
(H)	7	3.10
Total	226	100.00